Published in final edited form as:

Addiction. 2013 June ; 108(6): 1145–1157. doi:10.1111/add.12117.

Which Elements of Improvement Collaboratives Are Most Effective? A Cluster-Randomized Trial

D. H. Gustafson¹ [research professor, director], A. R. Quanbeck¹ [associate researcher], J. M. Robinson² [statistician, director], J. H. Ford II¹ [assistant scientist], A. Pulvermacher¹ [project director], M. T. French³ [professor of health economics], K. J. McConnell⁴ [associate professor of emergency medicine], P. B. Batalden⁵ [medical doctor, professor of pediatrics], K. A. Hoffman⁴ [senior research associate], and D. McCarty⁴ [professor of public health and preventive medicine]

¹Center for Health Enhancement Systems Studies, University of Wisconsin–Madison, 1513 University Avenue, Madison, WI 53706, USA

²Center for Health Systems Research and Analysis, University of Wisconsin–Madison, 610 Walnut Street, Madison, WI 53726 USA

³University of Miami, 5202 University Drive, P. O. Box 248162, Coral Gables, FL 33124, USA

⁴Oregon Health and Science University, 3181 SW Sam Jackson Park Rd., Portland, OR 97239, USA

⁵Geisel School of Medicine at Dartmouth, 1449 Hythe St., St. Paul, MN 55108, USA

Abstract

Aims—Improvement collaboratives consisting of various components are used throughout healthcare to improve quality, but no study has identified which components work best. This study tested the effectiveness of different components in addiction treatment services, hypothesizing that a combination of all components would be most effective.

Design—An unblinded cluster-randomized trial assigned clinics to one of four groups: interest circle calls (group teleconferences), clinic-level coaching, learning sessions (large face-to-face meetings), and a combination of all three. Interest circle calls functioned as a minimal intervention comparison group.

Setting—Outpatient addiction treatment clinics in the U.S.

Participants—201 clinics in 5 states.

Measurements—Clinic data managers submitted data on three primary outcomes: waiting time (mean days between first contact and first treatment), retention (percent of patients retained from first to fourth treatment session), and annual number of new patients. State and group costs were collected for a cost-effectiveness analysis.

*Corresponding author: D. H. Gustafson, Center for Health Enhancement Systems Studies, Industrial and Systems Engineering Department, College of Engineering, University of Wisconsin–Madison, 1513 University Ave., Madison, WI 53706, USA. Tel.: 608-263-4882; fax: 608-890-1438. dhgustaf@wisc.edu.

Trial registration ClinicalTrials NCT0093414.

Declaration of interest: The study was funded by the U.S. National Institute on Drug Abuse (R01 DA020832). The funder had no role in the study design, data collection, analysis, interpretation of data, the content of the study, or deciding to submit the article for publication. D McC serves as the Principal Investigator on Research Service Agreements between Oregon Health & Science University and Purdue Pharma and Alkermes, Inc.

Findings—Waiting time declined significantly for 3 groups: coaching (an average of −4.6 days/clinic, $P=0.001$), learning sessions (−3.5 days/clinic, $P=0.012$), and the combination (−4.7 days/clinic, $P=0.001$). The coaching and combination groups significantly increased the number of new patients (19.5%, $P=0.028$; 8.9%, $P=0.029$; respectively). Interest circle calls showed no significant effects on outcomes. None of the groups significantly improved retention. The estimated cost/clinic was \$2,878 for coaching versus \$7,930 for the combination. Coaching and the combination of collaborative components were about equally effective in achieving study aims, but coaching was substantially more cost effective.

Conclusions—When trying to improve the effectiveness of addiction treatment services, clinic-level coaching appears to help improve waiting time and number of new patients while other components of improvement collaboratives (interest circles calls and learning sessions) do not seem to add further value.

Introduction

Quality improvement (QI) collaboratives are used in North America, Australia, England, and European countries to improve healthcare quality, and yet little is known about which components of collaboratives work best.¹ This study reports the primary outcomes from a trial designed to identify the most effective elements in improvement collaboratives.

The study applied QI to U.S. addiction treatment clinics. The American addiction treatment system resembles health systems in many countries because 80% of it is publicly funded.² Internationally, policy makers demand more effective treatments for alcohol and drug use disorders, and it has been suggested that process improvement can change what providers do and how treatments work.³

This study addressed problems of access and adherence to treatment by focusing on process targets (e.g., reducing waiting time). The link between process targets and patient outcomes is complex and depends in part on which patient outcomes are studied. A link was found between process goals and increased patient show rate,⁴ reduced medical expenses,⁵ and reduced arrests and incarcerations.⁶ Harris et al did not find an association between increased continuity of care and reduced problematic substance use;⁷ they did find a statistically significant but clinically modest association between improved engagement and problematic substance use.⁸ In a review of the literature, Humphreys and McLellan found that although process improvements can change how treatment programs work, the link to better patient outcomes is weak, in part because outcomes are so heavily influenced by events in and the environment of patients' lives.³

The primary unit of analysis is the clinic (not the patient) because clinic leaders set organizational policies and processes that affect patient care. The study's main goal was identifying which components of improvement collaboratives are most effective in helping clinics reduce waiting time to enter treatment, enhance treatment retention, and increase the number of new patients.

Methods

Study Design and Participants

NIATx 200 was a cluster-randomized trial conducted according to the CONSORT criteria in 201 addiction treatment clinics in five U.S. states. NIATx (the Network for the Improvement of Addiction Treatment), a research center at the University of Wisconsin–Madison, promotes process improvements in addiction treatment. Details about the methods have been previously published.⁹ Eligible clinics were publicly funded outpatient and intensive

outpatient clinics with at least 60 patients annually and no previous NIATx experience. All patients seen within an enrolled clinic were included in the analysis.

Randomization

Clinics were randomized into one of four groups that used different components of improvement collaboratives:^{1,10,11} (1) interest circle calls, (2) coaching, (3) learning sessions, and (4) the combination of all three components. Clinics, randomized within states, were stratified by size (measured in number of patients per year) and management score.¹² University of Wisconsin researchers enrolled clinics, generated the allocation sequence, and assigned clinics to groups using a computerized random number generator. Clinic staff and researchers were unblinded.

Intervention

The 18-month intervention, delivered at the clinic level, was divided into three 6-month periods, each with one aim and a set of web-based materials about NIATx 200 recommended practices, instructions on implementing changes, QI tools, measures, and case studies (Fig. 1).^{13,14} The goal was providing the same content to all participants, varied by the support provided. The components tested in this study come from Institute for Healthcare Improvement¹⁰ and NIATx¹⁴ models.

Interest circle calls were monthly teleconferences in which staff from different clinics learned from experts and discussed progress with one another. The calls were led by QI experts trained in the NIATx model.¹⁴ Interest circle calls provide a simple and inexpensive way for clinics to collaborate, but quality may vary by facilitator. In addition, the calls may conflict with some participants' schedules, limiting participation. This was the lowest cost, lowest intensity study condition, and it functioned as a minimal intervention comparison group.

Coaching assigned one of the QI experts to help clinic leaders and change teams make improvements. Coaching involved one initial site visit, monthly phone conferences, and e-mail correspondence. During the site visit, the coach met with clinic leaders and change team members to plan the clinic's first change project. Using follow-up calls and e-mail, the coach reviewed the assigned project aim with clinic leaders and suggested practices and QI tools from the website. Coaches encouraged clinics to work on the assigned aim for each 6-month period and wrote reports summarizing each clinic contact. Coaches tailor process improvement advice to leaders and change teams. The match between coach and organization may or may not be good, and the quality of coaching can vary.

Learning sessions occurred in each state during months 1, 6, and 12 of the 18-month intervention period. The same three coaches led all sessions. Learning sessions convened change teams from different clinics in face-to-face conferences to learn from coaches and one another. Participating state agencies hosted learning sessions. The agendas for the learning sessions guided the content delivered in subsequent months in the other three groups. Learning sessions are intended to promote peer learning.^{10,11} They can also provide inspiration and social support.¹⁵ Learning sessions are expensive and require that most participants travel.

The *combination group* had the same type and number of interest circle calls, coaching activities, and learning sessions as groups 1, 2, and 3.

Implementation

In the U.S., an agency of state government in each of the 50 states coordinates services and manages federal funds for addiction treatment. The research team worked with five of these state agencies to conduct the study. The agencies recruited clinics to the study, sometimes encouraging specific clinics to apply (e.g., clinics with a good record of providing data). Compared to all eligible clinics, enrolled clinics were larger by approximately 100 annual admissions, served a smaller proportion of African Americans, and were more often not-for-profit. The five states formed two cohorts. Cohort 1 had clinics in Michigan, New York, and Washington; recruitment lasted from March through September 2007 and the interventions from October 2007 through March 2009. Cohort 2 had clinics from Massachusetts and Oregon; dates shifted 4 months later.

All clinics began with baseline data collection and a “walkthrough,” in which staff members assumed the role of a patient to personally experience the intake process and identify areas needing improvement.¹⁶ At the beginning of each 6-month intervention, the web-based curriculum (www.niatx200.net) was launched for the aim of that period. Staff members were encouraged to use recommended practices (e.g., establish walk-in hours) and their assigned component (e.g., coaching) to address each aim. A sustainability period lasting up to 9 months followed each 6-month intervention to assess maintenance of change.

Outcome Measures and Data Collection

Primary outcome measures were mean days between first contact and first treatment (waiting time), retention rate from first to fourth treatment session, and percent increase in the annual number of new patients. Each outcome was aggregated to the clinic level for analysis.

Data Collection—Each state designed systems for collecting waiting time¹⁷ and retention data and designated a data manager who collected, de-identified, and sent patient data to the research team. A designated data coordinator at each clinic submitted dates of first contact, assessment, and first four treatment sessions for each patient admitted to care. States hosted training sessions for clinic data coordinators and provided technical support throughout the study. The number of new patients was collected in annual surveys.

Economic Analysis—The goal of the economic analysis was to estimate costs of each group for governmental authorities who might organize improvement collaboratives. Costs to the clinics of participating in the study—such as staff time spent on implementing changes—were not collected. The cost data collection instrument was based on the Drug Abuse Treatment Cost Analysis Program.¹⁸ The instrument collected the cost of personnel (state employees, NIATx employees, coaches, and consultants), data management, buildings and facilities, lodging, travel, phone calls, and miscellaneous costs. Costs were categorized as group specific (such as hotel costs for the learning sessions group) or non-group-specific, which included state-incurred costs for outreach, data management and infrastructure, encouraging participation, and administration.

Cost data were collected 3 times during the study period and aggregated to create a total cost estimate. To compare the cost effectiveness of the groups, the total cost of each group was divided by the total change in each group in each outcome. We calculated the annualized reduction in waiting days for each group by multiplying the average improvement in waiting days per clinic per patient by the number of clinics analyzed in the group and the average number of patients per year per clinic. A similar approach was used to calculate the annualized increase in new patients for each group and the annualized increase in the number of retained patients. We also calculated cost effectiveness by clinic.

Statistical Analysis

In determining sample size, we wanted to detect a 10% reduction in waiting time. Power calculations were done for various sample sizes considering clinic recruitment levels anticipated in each state. We assumed a clinic attrition rate of 20% and a baseline average waiting time of 30 days. A sample of 200 clinics provided 80% power to detect a difference of 10.6% in waiting time, 7.5% in retention, and 14.2% in annual number of new patients.

Mixed-effect regression models were fit to primary outcomes, including terms to isolate state and group effects. Organization-level random effects were included to model the correlation among clinics within the same organization, and clinic-level random effects were included to model correlations between repeated observations from the same clinic over time. A vector of monthly waiting time and retention averages for each clinic served as the unit of analysis. Monthly averages based on fewer than 5 patient records were removed. To prevent larger clinics from dominating the results, equal weight was given to each clinic. The number of new patients was aggregated to an annual level to minimize seasonality. Increases in the number of new patients were summarized by changes in the natural logarithm of new patients from baseline data collected and combined for 2006 and 2007 (before the first intervention period) and surveys taken and combined for 2008 and 2009 (after the interventions began). Groups were compared pair-wise to detect statistically significant differences between groups. Clinic-level covariates accounted for in the analysis included clinic size, management score, and state affiliation.

Participant flow

Figure 2 shows the flow of clinics through the trial. Clinics were included in the analysis regardless of their level of participation according to the intention-to-treat principle.

Results

Baseline data

Table 1 shows baseline characteristics of the 201 clinics. Baseline data collection extended one month into the first intervention period as clinics increased their data collection capacity. At baseline, across all clinics, mean waiting time was 36.8 days, mean retention through 4 sessions was 74.9%, and mean number of new patients was 563 per year. Although better baseline management scores were associated with lower waiting times¹² and baseline retention varied significantly by state, differences in these covariates were not associated with outcome changes.

Waiting Time

Table 2a summarizes waiting time changes. At the end of the 6-month intervention, the coaching and combination groups had statistically significant reductions in waiting time (4.9 days, $P=0.013$ and 6.2 days, $P=0.002$, respectively). Learning sessions had a modest waiting time reduction while interest circle calls had a slight increase, but these two groups' changes were not statistically significant. Pair-wise comparisons between groups reveal a significant difference in improvement between interest circle calls and the combination group ($P=0.024$).

Table 2b shows the reduction from baseline in average waiting days per patient averaged across each of the 14 months of the intervention and sustainability periods. For these months, three groups had statistically significant reductions--coaching (4.6 days), learning sessions (3.5 days), and the combination (4.7 days). Differences between coaching and interest circle calls and between the combination and interest circle calls were statistically significant ($P=0.028$ and $P=0.024$, respectively). Although the three groups ended the

evaluation period with similar levels of improvement, the combination group had the greatest improvement (followed by coaching, and then learning sessions) because of the patterns of improvement over time.

Retention

None of the groups showed significant improvement in retention for the 6-month intervention period (Table 3a) or the entire intervention and sustainability period (Table 3b), and there were no significant differences between groups.

Number of New Patients

The last QI aim was increasing the number of new patients. The intervention and sustainability periods were combined for this outcome because numbers were aggregated to a yearly level to reduce seasonality. Table 4 shows that the coaching and combination groups both had statistically significant increases of 19.5% ($P=0.028$) and 8.9% ($P=0.029$), while learning sessions and interest circle calls did not differ from baseline. (Despite a substantial difference in coefficients, the coaching and combination group P values are similar because the coaching group had a higher standard error.) Pair-wise comparisons indicate that the coaching and combination groups both had significantly greater increases than interest circle calls ($P=0.018$ and $P=0.029$, respectively).

Sensitivity Analyses

A sensitivity analysis was run to weight each clinic's contribution to outcomes differentially (based on patient counts) rather than weighting each clinic equally. This alternative model did not change the ordering of the groups' performance with respect to waiting time or number of new patients, suggesting that larger and smaller clinics performed similarly on these outcomes.

We used an alternative measure of retention in an exploratory analysis. Baseline retention rate from first to fourth treatment averaged nearly 75%, higher than reported elsewhere,^{7,8} possibly causing a ceiling effect. Furthermore, many patients can be lost between first request and treatment;¹⁹ measuring retention only from first treatment misses this early dropoff. The Massachusetts and New York data included records of all patients who requested treatment, regardless of whether they received any. For 57 clinics in these two states, it was possible to measure retention from first request to fourth treatment and capture the early dropoff. For these clinics, the baseline retention rate from first request to fourth treatment was 32%. Measuring retention this way removed the ceiling effect. Improvements in retention rate from first request (rather than first treatment) to fourth treatment were 4% for interest circle calls, 22% for coaching, 27% for learning sessions, and 26% for the combination group. These results parallel those for waiting time—interest circle calls showed little improvement while the coaching and combination groups did. No statistical test was performed because the analysis was exploratory.

Costs and Cost Effectiveness

Table 5a shows the non-group-specific costs incurred by states. Each state enrolled between 37 and 43 clinics and costs per state ranged from US\$85,475 to US\$394,729. These estimates suggest the range of costs for running a research-based collaborative for approximately 40 clinics. One state (New York) had very high data infrastructure costs, in part to create a system to collect data about first contact for all prospective patients, indicating that the costs of running a collaborative depend greatly on the data infrastructure already in place. Figure 6a also shows per-clinic costs. Data infrastructure costs are excluded

from the per-clinic costs because they are assumed to be one-time costs and largely independent of the number of clinics enrolled.

Table 5b shows group costs and cost-effectiveness ratios (CERs). The least expensive group was interest circle calls, followed by coaching, learning sessions, and the combination. Our approach in calculating CERs was to order the groups by cost and eliminate any dominated alternatives (i.e., more costly but less effective groups). Dominated alternatives are listed as “N/A,” and the CER is based on the next available alternative in terms of cost. For waiting time, coaching had a CER of \$0.56 per patient waiting day saved compared to interest circle calls; coaching dominated learning sessions; and the CER for the combination compared to coaching was \$37.30. For new patients, coaching had a CER of \$11.36 per new patient compared to interest circle calls and dominated both learning sessions and the combination. While cost-effectiveness ratios are reported for retention in Figure 6b, none of the groups had a statistically significant effect, making any interpretation of CERs tenuous.

Discussion

Principal findings

This randomized trial demonstrates that coaching and the combination of collaborative elements can produce statistically significant improvements in waiting time and the number of new patients compared to interest circle calls (a web site and monthly teleconferences). However, learning sessions (three face-to-face conferences at 6-month intervals) did not produce statistically significant outcome changes compared to interest circle calls. Coaching is the more cost-effective component. The combination costs almost 300% more than coaching by itself. For statistically significant differences between groups, effect sizes varied between 0.23 to 0.60, a range common for non-laboratory studies.²⁰ Organizational change is difficult because it depends on the interaction among individuals, the setting, the organizational climate, and the change itself.²¹ A study that examined 54 reviews of various interventions aimed at changing clinical practice found average effects of about 10% for main targets, similar to effects in this study.²²

Strengths and weaknesses of the study

This study addresses a research question that has not to our knowledge been asked before: Which component(s) of improvement collaboratives are most effective in improving healthcare quality? The study is unique in the literature on QI in healthcare both in its scope and the strength of its design.

The study has limitations. Assessment of clinic-level costs was outside the scope of this research, though these costs will likely have an important effect on whether clinics ultimately decide to participate in QI collaboratives. Assessment of patient-level health outcomes was not possible. Limits to generalizability include clinics in NIATx 200 being significantly larger than the average addiction treatment center in each state, and the possibility that organizations electing to participate may have been more receptive to QI than those that declined.

The retention results show that the interventions had no significant effect, and yet the exploratory study (which included people seeking but not getting treatment) produced results similar to those for waiting time and the number of new patients, suggesting that a more accurate indication of clinic performance would require collecting data on all calls for help, not just those resulting in treatment.

It may be that the study outcomes respond to different types of improvements. The process changes recommended in the NIATx 200 curriculum may be more suitable to administrative

outcomes such as waiting time and the number of new patients, while retention may require an approach focused more on clinicians and clinical practice. One could argue that coaching, which consisted in this study of one site visit and monthly phone calls, was too light or that the content and quality of the learning sessions and interest circle calls could have been different. Yet the design and content were based on established work on improvement collaboratives.^{10,14} Results could vary under other circumstances.

Explanation and implications

Why might coaching outperform the other groups? One reason may be that coaching tailors instruction.²³ A coach can encourage a clinic to stay with a topic longer, if needed; interest circle calls and learning sessions proceed at a more predetermined pace. Coaches can be a more persistent voice for improvement than less personal interventions. Coaches can also respond directly to change leaders' concerns and smooth transitions when staff turns over. Schein identifies three ways consultants can help organizations—being an expert resource, diagnosing problems, and consulting on process²⁴—and the coach reports submitted after each clinic contact indicate that NIATx 200 coaches did help in these ways. Although coaching is increasingly a component of QI initiatives in healthcare,^{25, 26} the literature about coaching is sparse.

Why would coaching be as effective as the combination of services? The combination may have offered more information than participants could use. Information from one source (e.g., interest circle calls) also might have competed with information from another (e.g., coaching), causing confusion. The data in this and other studies^{23,27} also suggest that although the combination group had the best waiting time results in the first six months, coaching (and to a lesser extent, learning sessions) “caught up” and even began to surpass the combination over the sustainability period. Combining services appears to produce the greatest initial results, while coaching delivers steady gains over time.

Future research

Effective coaching bears further research. Coaching in this study drew from the work of Deming, who encouraged focusing on processes and using an outside consultant—a “master”—to help organizations make improvements.²⁸ Others, such as Donald Schön²⁹ and Atul Gawande,³⁰ describe coaching as it relates to individual work performance. What are the characteristics of effective coaching for organizations and individuals? How can a good match be defined between coach and organization and coach and individual?

The study raises questions about a key element of improvement collaboratives, learning sessions.¹⁰ This may be good because learning sessions are expensive. The results of this study should encourage the use of QI and suggest ways to reduce the cost of doing so.

Acknowledgments

We thank the thousands of very dedicated, very busy staff members from treatment clinics in Massachusetts, Michigan, Oregon, New York, and Washington who took part in this study to improve the care of the patients they serve; Tom Hilton, Ph.D., and Redonna Chandler, Ph.D., from the National Institute on Drug Abuse for their deep interest in and commitment to the project; Susan Brandau, Michael Botticelli, Phil Chvojka, Shawn Clark, Michael Ellis, Dawn Lambert-Wacey, Eric Larson, Robin Roberts, Mark Steinberg, Karen Wheeler, Fritz Wrede, and Dagan Wright from the state authorities for substance abuse treatment, whose work was indispensable to the project; staff members at the Center for Health Enhancement Systems Studies at the University of Wisconsin–Madison: Nataliya Batina, Jim Kadunc, and John Oruongo for their help with the data analysis; Victor Cappocia, Ph.D., for his input on data analysis; Bobbie Johnson for writing and editing assistance; Skyler Neylon for input about data analysis; Victoria Sviridova for data entry and data collection; Kyle Grazier, Ph.D., School of Public Health and School of Medicine, University of Michigan, and David Zimmerman, Ph.D., Center for Health Systems Research and Analysis, University of Wisconsin–Madison, for their input on data analysis; Frank Davidoff, M.D., Institute for

Healthcare Improvement, for his thoughtful reviews of the manuscript; and the coaches and facilitators for their commitment and good work with the addiction treatment clinics.

References

1. Schouten LMT, Hulscher MEJL, Everdingen JJE, Huijsman R, Grol RPTM. Evidence for the impact of quality improvement collaboratives: systematic review. *Br Med J*. 2008; 336(7659):1491. [PubMed: 18577559]
2. Mark TL, Levit KR, Vandivort-Warren R, Buck BA, Coffey RM. Changes in US spending on mental health and substance abuse treatment, 1986–2005, and implications for policy. *Health Aff*. 2011; 30:284–292.
3. Humphreys K, McLellan AT. A policy-oriented review of strategies for improving the outcomes of services for substance use disorder patients. *Addiction*. 2011; 106:2058–2066. [PubMed: 21631620]
4. Stasiewick PR, Stalker R. A comparison of three interventions on pretreatment dropout rates in an outpatient substance abuse clinic. *Addict Behav*. 1999; 24(4):579–582. [PubMed: 10466854]
5. Wickizer TM, Krupski A, Stark KD, Mancuso D, Campbell K. The effect of substance abuse treatment on Medicaid expenditures among general assistance welfare clients in Washington State. *Milbank Q*. 2006; 84(3):555–576. [PubMed: 16953810]
6. Garnick DW, Horgan CM, Lee MT, Panas L, Ritter GA, Davis S, et al. Are Washington Circle performance measures associated with decreased criminal activity following treatment? *J Subst Abuse Treat*. 2007; 33(4):341–352. [PubMed: 17524596]
7. Harris AHS, Humphreys K, Bowe T, Kivlahan DR, Finney JW. Measuring the quality of substance use disorder treatment: evaluating the validity of the Department of Veterans Affairs continuity of care performance measure. *J Subst Abuse Treat*. 2009; 36(3):296–305.
8. Harris AHS, Humphreys K, Bowe T, Tiet Q, Finney JW. Does meeting the HEDIS substance abuse treatment engagement criterion predict patient outcomes? *J Behav Health Serv Res*. 2010; 37(1):25–39. [PubMed: 18770044]
9. Quanbeck AR, Gustafson DH, Ford JH II, Pulvermacher A, French MT, McConnell KJ, et al. Disseminating quality improvement: study protocol for a large cluster-randomized trial. *Implement Sci*. 2011; 6:44. [PubMed: 21524303]
10. Institute for Healthcare Improvement. The breakthrough series: IHI's collaborative model for achieving breakthrough improvement. Boston, MA: Institute for Healthcare Improvement; 2003.
11. Wilson T, Berwick DM, Cleary PD. What do collaborative improvement projects do? Experience from seven countries. *Jt Comm J Qual Saf*. 2003; 29(2):85–93. [PubMed: 12616923]
12. McConnell KJ, Hoffman KA, Quanbeck A, McCarty D. Management practices in substance abuse treatment programs. *J Subst Abuse Treat*. 2009; 37(1):79–89. [PubMed: 19195813]
13. Gustafson DH, Hundt AS. Findings of innovation research applied to quality management principles for health care. *Health Care Manage Rev*. 1995; 20(2):16–33. [PubMed: 7607882]
14. Gustafson, DH.; Johnson, KA.; Capoccia, V.; Cotter, F.; Ford, JH., 2nd; Holloway, D., et al. The NIATx model: process improvement in behavioral health. Madison, WI: University of Wisconsin, Center for Health Enhancement Systems Studies; 2011.
15. Nembhard IM. Learning and improving in quality improvement collaboratives: Which collaborative features do participants value most? *Health Serv Res*. 2009; 44(2):359–378. [PubMed: 19040423]
16. Ford JH 2nd, Green CA, Hoffman KA, Wisdom JP, Riley KJ, Bergmann L, et al. Process improvement needs in substance abuse treatment: admissions walk-through results. *J Subst Abuse Treat*. 2007; 33(4):379–389. [PubMed: 17499961]
17. Hoffman KA, Quanbeck A, Ford JH 2nd, Wrede F, Wright D, Lambert-Wacey D, et al. Improving substance abuse data systems to measure 'waiting time to treatment': Lessons learned from a quality improvement initiative. *Health Informatics J*. 2011; 17(4):256–265. [PubMed: 22193826]
18. French MT, Dunlap LJ, Zarkin GA, McGeary KA, McLellan AT. A structured instrument for estimating the economic cost of drug abuse treatment: The Drug Abuse Treatment Cost Analysis Program (DATCAP). *J Subst Abuse Treat*. 1997; 14(5):445–455. [PubMed: 9437614]

19. Hser Y, Maglione M, Polinsky ML, Anglin MD. Predicting drug treatment entry among treatment-seeking individuals. *J Subst Abuse Treat.* 1998; 15(3):213–220. [PubMed: 9633033]
20. Cohen, J. *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Earlbaum Associates; 1988.
21. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implement Sci.* 2009; 4:50. [PubMed: 19664226]
22. Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. *Lancet.* 2003; 362:1225–1230. [PubMed: 14568747]
23. McCarty D, Gustafson DH, Wisdom JP, Ford J, Choi D, Molfenter T, et al. The Network for the Improvement of Addiction Treatment (NIATx): enhancing access and retention. *Drug Alcohol Depend.* 2007; 88(2–3):138–145. [PubMed: 17129680]
24. Schein, EH. *Helping: how to offer, give, and receive help.* San Francisco, CA: Berrett-Koehler Publishers; 2009.
25. Horbar JD, Rogowski J, Plsek PE, Delmore P, Edwards WH, Hocker J, et al. Collaborative quality improvement for neonatal intensive care. *Pediatrics.* 2001; 107(1):14–22. [PubMed: 11134428]
26. Ovreteit J, Bate P, Cleary P, Cretin S, Gustafson D, McInnes K, et al. Quality collaboratives: lessons from research. *Qual Saf Healthcare.* 2002; 11(4):345–351.
27. Hoffman KA, Ford IJH 2nd, Choi D, Gustafson DH, McCarty D. Replication and sustainability of improved access and retention within the Network for the Improvement of Addiction Treatment. *Drug Alcohol Depend.* 2008; 98(1–2):63–69. [PubMed: 18565693]
28. Deming, WE. *Out of the crisis.* Cambridge, MA: Massachusetts Institute of Technology, Center for Advanced Engineering Studies; 1986.
29. Schön, DA. *Educating the Reflective Practitioner.* San Francisco, CA: Jossey-Bass; 1987.
30. Gawande A. Personal best. *New Yorker.* 2011; 87(30):44–53.

1a: Study Timeline

A. Launch included recruitment, modification of data systems, and randomization to 4 groups.
 B. All 4 groups had the same aim for each 6-month intervention.
 C. The sustainability period for each aim lasted up to 9 months.

1b: Recommended Practices and Tools

Aim	Months		
	1–6	7–12	13–18
Promising practices	Reduce waiting time	Increase retention	Increase number of new patients
	Increase availability of time slots: Add groups Add time slots Combine intake appointments Double-book time slots Increase the amount of counseling: Ask patients to complete paperwork Assign backup counselors Cross-train counselors Eliminate excessive paperwork Reassign tasks Transition existing patients Eliminate appointments: Establish walk-in hours Provide interim services Make appointments immediately: Make appointments at the front desk Make appointments during the first call for service Open access to all time slots Suspend financial arrangements	Make it easy to enter treatment: Connect with patients during first contact Establish clear 2-way expectations Help eliminate barriers to treatment Include family and friends Offer an inviting physical environment Remind patients about the next appointment Make it difficult to refuse or quit treatment: Collaborate with referral sources Follow up with no-shows Identify clients at risk for leaving and intervene Make it easy to stay in treatment: Assign peer buddies Build community among patients Have patients help create their treatment plans Have patients select groups Use contingency management	Make admissions easy for patients: Offer a tour guide Overlap levels of care Blend levels of care Include family and friends in discharge and admission planning Use motivational interviewing Use video conferencing Map out continuing treatment Orient patients to outpatient treatment Offer telephone support Reward attendance at first outpatient appointment Make referrals easy for referrers: Assign one contact person to each referral source Schedule outpatient appointments before patients leave Guide referrers to make appropriate referrals Tailor brochures for each referral source Hold joint staffing Streamline paperwork Increase referral sources
	Tools	Tools	Tools
	Conducting a walk-through Collecting data Using flowcharting Using the change project charter Conducting the nominal group process Using Plan-Do-Study-Act cycles	Producing a business case Interpreting data Using story boards	Using cost-benefit analysis Planning for sustainability

Figure 1.
 NIATx 200 Study Design

Figure 2. Flow diagram

^a “Lost to follow up” indicates clinics that did not send primary outcomes data to state data managers. Clinics that did send in primary outcomes data were included in the intent-to-treat analysis. Cleaning scripts were run on the data sent by the state data managers to the research team, and clean data were entered into a Microsoft SQL Server database. Loss to follow-up rates did not differ significantly between groups: interest circle calls (8.2%), coaching (6.0%), learning sessions (5.6%), and combination (0.0%).

^b Most clinics that were lost to follow up had scheduling conflicts that precluded their participating in QI activities. Most clinic staff have only limited control over their schedules.

^c “Hours” are hours of participation per clinic, unweighted by the number of clinic staff members participating from each clinic. The interest circle call group, with 18 maximum possible hours of participation, was conceived as a type of control or comparison group.

Table 1

Baseline clinic characteristics by group: Number of clinics in group unless otherwise specified

Characteristic of clinic ^a	Interest circle calls N = 49	Coaching N = 50	Learning sessions N = 54	Combination N = 48
Cohort 1 ^b				
Michigan	11	11	10	10
New York	10	10	10	11
Washington	8	10	11	9
Cohort 2				
Massachusetts	11	11	12	9
Oregon	9	8	11	9
Type (#, %)				
Private for-profit	0	0	3 (5.6)	0
Private not-for-profit	37 (75.5)	41 (82.0)	43 (79.6)	41 (85.4)
Unit of state government	4 (8.2)	1 (2.0)	3 (5.6)	4 (8.3)
Unit of tribal government	1 (2.0)	2 (4.0)	2 (3.7)	2 (4.2)
Unit of other government	7 (14.3)	6 (12.0)	3 (5.6)	1 (2.1)
Primary setting (#, %)				
Hospital or health center (including primary setting)	4 (8.2)	6 (12.0)	3 (5.6)	5 (10.4)
Community mental health clinic	7 (14.3)	13 (26.0)	10 (18.5)	5 (10.4)
Free-standing alcohol or drug treatment clinic	26 (53.1)	25 (50.0)	32 (59.3)	26 (54.2)
Family or children's service agency	3 (6.1)	2 (4.0)	2 (3.7)	1 (2.1)
Social services agency	1 (2.0)	2 (4.0)	2 (3.7)	3 (6.3)
Corrections	3 (6.1)	0	1 (1.9)	1 (2.1)
Other or unreported	5 (10.2)	2 (4.0)	4 (7.4)	7 (14.6)
New patients per year (mean; standard deviation)	500 (485)	578 (684)	671 (1031)	488 (437)
Management score ^c (mean; standard deviation)	3.02 (0.54)	2.97 (0.64)	2.95 (0.73)	2.99 (0.57)
Non-white/non-Hispanic patients (%)	26.0	33.2	29.8	27.4

^aOrganizations could designate up to 4 clinics to participate in the study, although most (155 of 174) designated a single clinic.

^bThe states formed 2 cohorts because 3 states were ready to start sooner than the other 2. Cohort 1 started 4 months before Cohort 2. We monitored for possible cohort contamination by tracking who participated in each cohort and each group and identified no case of a participant from one state taking part in the services intended for another state. We know of only one instance of group contamination: In Oregon, the staff from one clinic assigned to the interest circle calls group attended a learning session held at 12 months.

^cThe management score resulted from a telephone interview of a senior leader at each clinic. The interview had questions about 14 management practices grouped into 4 areas: intake and retention (2 practices); quality monitoring and reviewing (5 practices), which includes performance tracking and review; (3) targets (3 practices), which includes the realism and transparency of program goals; and (4) employee incentives (4 practices), which includes promotion criteria, pay, and bonuses. The interviewer asked one question per practice and scored each answer between 1 and 5, with a higher score indicating better management. The mean management score was 2.98.

Table 2

Summary of Improvement in Waiting Time^a

2a: Improvement in Waiting Time During Waiting Time Intervention Period									
	Estimated Improvement (Day)	N	Standard Error	95% CI Lower Bound	95% CI Upper Bound	[†] Ratio	P Value	Effect Size ^b	% of Clinics That Improved (>10%)
Interest circle calls	-0.235	36	2.116	-4.467	3.997	-0.111	0.911	0.019	47%
Coaching	4.863	42	1.950	0.963	8.763	2.494	0.013	0.388	95%
Learning sessions	3.143	44	1.927	-0.712	6.997	1.630	0.103	0.251	91%
Combination	6.159	41	1.970	2.218	10.099	3.125	0.002	0.492	95%
Coaching vs. interest circle calls	5.098	N/A	2.812	-0.526	10.721	1.813	0.070	0.407	N/A
Learning sessions vs. interest circle calls	3.378	N/A	2.799	-2.220	8.976	1.207	0.228	0.270	N/A
Combination vs. interest circle calls	6.394	N/A	2.825	0.744	12.044	2.263	0.024	0.510	N/A
Combination vs. learning sessions	3.016	N/A	2.691	-2.365	8.397	1.121	0.262	0.241	N/A
Combination vs. coaching	1.296	N/A	2.705	-4.115	6.707	0.479	0.632	0.103	N/A
Coaching vs. learning sessions	1.720	N/A	2.680	-3.639	7.079	0.642	0.521	0.137	N/A

2b. Average Improvement in Waiting Time from Base ^c Through Sustainability Period ^d									
	Estimated Improvement (Days)	N	Standard Error	95% CI Lower Bound	95% CI Upper Bound	[†] Ratio	P Value	Effect Size ^b	% of Clinics That Improved (>10%)
Interest circle calls	0.253	36	1.514	-2.775	3.282	0.167	0.867	0.028	42%
Coaching	4.610	42	1.414	1.781	7.438	3.259	0.001	0.509	95%
Learning sessions	3.520	44	1.407	0.706	6.335	2.502	0.012	0.388	80%
Combination	4.723	41	1.421	1.882	7.564	3.325	0.001	0.521	90%
Coaching vs. interest circle calls	4.356	N/A	1.981	0.395	8.318	2.200	0.028	0.481	N/A
Learning sessions vs. interest circle calls	3.267	N/A	1.977	-0.687	7.221	1.653	0.098	0.360	N/A
Combination vs. interest circle calls	4.470	N/A	1.983	0.504	8.435	2.254	0.024	0.493	N/A

2b. Average Improvement in Waiting Time from Base ^c Through Sustainability Period ^d										
Combination vs. learning sessions	1.203	N/A	1.907	-2.611	5.016	0.631	0.528	0.133	N/A	N/A
Combination vs. coaching	0.113	N/A	1.912	-3.710	3.937	0.059	0.953	0.013	N/A	N/A
Coaching vs. learning sessions	1.089	N/A	1.907	-2.725	4.904	0.571	0.568	0.120	N/A	N/A

CI: Confidence interval.

^aEstimates of improvement show the average days of Improvement per month based on a best linear unbiased predictor estimate for each site.

^bEffect size is defined as the gross improvement in the outcome measure divided by s , the standard deviation of the response variable. Arriving at the effect size involved computing a pooled standard deviation of clinic improvement using the standard errors of the average improvement by group, squaring each of the four standard errors, and multiplying each by $n-1$ to obtain estimates of the group-specific variances of outcome improvement. A weighted average of these variances was taken to obtain the pooled variance of the outcome improvements. Taking the square root obtains the pooled standard deviation of the outcome improvement.

^cThe base period consists of data from three months—for Cohort 1, from September - November 2007, and for Cohort 2, from January - March 2008.

^dWhile the coaching, learning session, and combination groups had similar endpoints for waiting time improvement, Figure 3b reports greater average improvement for the coaching and combination groups over the entire period. Compared to the learning session group, the coaching and combination groups had better results sooner, and the patients who were helped sooner continued to benefit, meaning that a greater number of patients benefited in these two groups than in the learning session group. The pair-wise comparisons in Figure 3b may be conceptualized as difference-in-area calculations rather than differences in endpoint values between groups.

Table 3

Summary of Improvement in Retention^a

3a: Improvement in Retention During Retention Intervention Period										
	Estimated Improvement (Percentage Points)	N	Standard Error	95% CI Lower Bound	95% CI Upper Bound	t Ratio	P Value	Effect Size ^b	% of Clinics That Improved (>0%)	% of Clinics That Improved (>10%)
Interest circle calls	-0.003	35	0.025	-0.054	0.048	-0.111	0.912	0.019	49%	11%
Coaching	0.035	41	0.022	-0.010	0.080	1.561	0.118	0.236	68%	22%
Learning sessions	-0.003	42	0.024	-0.051	0.045	-0.128	0.899	0.021	45%	5%
Combination	-0.003	43	0.023	-0.050	0.043	-0.137	0.891	0.021	40%	9%
Coaching vs. interest circle calls	0.038	N/A	0.034	-0.030	0.106	1.116	0.264	0.255	N/A	N/A
Learning sessions vs. interest circle calls	0.000	N/A	0.035	-0.070	0.070	-0.007	0.995	0.002	N/A	N/A
Combination vs. interest circle calls	0.000	N/A	0.034	-0.069	0.069	-0.010	0.992	0.002	N/A	N/A
Combination vs. learning sessions	0.000	N/A	0.033	-0.067	0.067	-0.004	0.997	0.001	N/A	N/A
Combination vs. coaching	-0.038	N/A	0.032	-0.103	0.026	-1.184	0.236	0.257	N/A	N/A
Coaching vs. learning sessions	0.038	N/A	0.033	-0.028	0.104	1.160	0.246	0.256	N/A	N/A

3b: Average Improvement in Retention from Base^c Through Retention sustainability Period

	Estimated Improvement (Percentage Points)	N	Standard Error	95% CI Lower Bound	95% CI Upper Bound	t Ratio	P Value	Effect Size ^b	% of Clinics That Improved (>0%)	% of Clinics That Improved (>10%)
Interest circle calls	-0.019	35	0.021	-0.060	0.023	-0.896	0.371	0.168	29%	6%
Coaching	-0.020	41	0.016	-0.053	0.013	-1.222	0.222	0.180	29%	10%
Learning sessions	0.013	42	0.018	-0.022	0.048	0.718	0.473	0.114	60%	7%
Combination	-0.004	43	0.017	-0.037	0.029	-0.249	0.803	0.037	58%	5%
Coaching vs. interest circle calls	-0.001	N/A	0.026	-0.052	0.050	-0.050	0.960	0.011	N/A	N/A
Learning sessions vs. interest circle calls	0.031	N/A	0.026	-0.021	0.084	1.187	0.235	0.282	N/A	N/A
Combination vs. interest circle calls	0.015	N/A	0.026	-0.037	0.066	0.564	0.573	0.131	N/A	N/A

3b: Average Improvement in Retention from Base ^c Through Retention sustainability Period										
	Estimated Improvement (Percentage Points)	N	Standard Error	95% CI Lower Bound	95% CI Upper Bound	tRatio	P Value	Effect Size ^b	% of Clinics That Improved (>0%)	% of Clinics That Improved (>10%)
Combination vs. learning sessions	-0.017	N/A	0.023	-0.063	0.030	-0.721	0.471	0.151	N/A	N/A
Combination vs. coaching	0.016	N/A	0.022	-0.029	0.060	0.708	0.479	0.142	N/A	N/A
Coaching vs. learning sessions	-0.033	N/A	0.023	-0.079	0.013	-1.413	0.158	0.293	N/A	N/A

CI: Confidence interval.

^a Estimates of improvement show the average percentage points of improvement per month based on a best linear unbiased predictor estimate for each site. To account for potential right censoring of the retention outcome, the analysis period for each clinic was truncated if its reported activity level (requests, assessments, and treatments combined) dipped below its monthly median and stayed below the median in all subsequent months.

^b Effect size is defined as the gross improvement in the outcome measure divided by s , the standard deviation of the response variable. Arriving at the effect size involved computing a pooled standard deviation of clinic improvement using the standard errors of the average improvement by group, squaring each of the four standard errors, and multiplying each by $n-1$ to obtain estimates of the group-specific variances of outcome improvement. A weighted average of these variances was taken to obtain the pooled variance of the outcome improvements. Taking the square root obtains the pooled standard deviation of the outcome improvement.

^c The base period consists of data from three months—for Cohort 1, from September – November 2007, and for Cohort 2, from January – March 2008.

Table 4

Summary of Improvement in Annual Number of Patients Treated

	Estimated Improvement (Percentage Points)	N	Standard Error	95% CI Lower Bound	95% CI Upper Bound	T Ratio	P Value	Effect Size ^a	% of Clinics That Improved (>0%)	% of Clinics That Improved (>10%)
Interest circle calls	-0.036	41	0.040	-0.117	0.044	-0.898	0.369	0.094	56%	32%
Coaching	0.195	40	0.089	0.017	0.373	2.191	0.028	0.503	70%	45%
Learning sessions	-0.001	38	0.065	-0.132	0.129	-0.022	0.982	0.004	63%	37%
Combination	0.089	33	0.041	0.008	0.170	2.188	0.029	0.230	70%	52%
Coaching vs. interest circle calls	0.231	N/A	0.098	0.036	0.426	2.366	0.018	0.597	N/A	N/A
Learning sessions vs. interest circle calls	0.035	N/A	0.077	-0.119	0.188	0.452	0.651	0.090	N/A	N/A
Combination vs. interest circle calls	0.125	N/A	0.057	0.011	0.240	2.185	0.029	0.323	N/A	N/A
Combination vs. learning sessions	0.090	N/A	0.077	-0.064	0.244	1.173	0.241	0.233	N/A	N/A
Combination vs. coaching	-0.106	N/A	0.098	-0.302	0.090	-1.084	0.278	0.274	N/A	N/A
Coaching vs. learning sessions	0.196	N/A	0.110	-0.024	0.417	1.778	0.075	0.507	N/A	N/A

CI: Confidence interval.

^aEffect size is defined as the gross improvement in the outcome measure divided by s , the standard deviation of the response variable. Arriving at the effect size involved computing a pooled standard deviation of clinic improvement using the standard errors of the average improvement by group, squaring each of the four standard errors, and multiplying each by $n-1$ to obtain estimates of the group-specific variances of outcome improvement. A weighted average of these variances was taken to obtain the pooled variance of the outcome improvements. Taking the square root obtains the pooled standard deviation of the outcome improvement.

Table 5

Costs and Cost Effectiveness of NIATx Interventions

5a. Non-Group-Specific Costs ^a Incurred by US States (US\$) in Three 6-month interventions						
	MA	MI	NY	OR	WA	
Outreach ^b	21,918	10,045	20,271	32,180	16,515	
Data management ^c	28,303	42,026	62,737	53,489	43,991	
Data infrastructure ^d	0	0	189,327	46,305	0	
Encouraging participation ^e	78,272	2,409	71,331	35,495	27,569	
Admin/Misc ^f	22,895	30,995	51,064	22,788	43,536	
Total	151,388	85,475	394,729	190,257	131,611	
Clinics randomized	43	42	41	37	38	
Total cost/clinic	3,521	2,035	9,628	5,142	3,463	
Cost/clinic (excluding data infrastructure)	3,521	2,035	5,010	3,891	3,463	

5b. Group Cost ^d and Cost Effectiveness															
Waiting time					Retention					New patients					
Clinic level					Group level		Clinic level		Group level		Clinic level			Group level	
Group	Cost	No. of clinics analyzed ^b	Cost/clinic	Reduction in days/clinic/month	CERs ^c	Days saved (entire group) ^d	CERs	Change in rate/clinic/month	CERs	No. of additional retained patients (entire group) ^e	Average change in log of annual admissions/clinic	CERs	No. of new patients (entire group) ^f	CERs	
Interest circle calls	\$59,790	45	\$1,329	0.253	-	7,465	-	-0.019	-	-561	-0.036	-	-796	-	
Coaching	\$135,244	47	\$2,878	4.610	356	142,063	0.56	-0.020	N/A	-616	0.195	6,706	5,849	11.36	
Learning session	\$229,229	51	\$4,495	3.520	N/A	117,705	N/A	0.013	98,938	435	-0.001	N/A	-34	N/A	
Combination	\$380,652	48	\$7,930	4.723	44,708	148,642	37.30	-0.004	N/A	-126	0.089	N/A	2,180	N/A	

^aNon-group-specific costs are those costs incurred by states to participate in the study apart from the costs of conducting the four groups. These costs include one-time setup costs and continuing costs.

^b“Outreach” costs refer to the costs of advertising the study and recruiting clinics to participate. Outreach costs were incurred before the first intervention started.

^c“Data management” costs refer to the extra staff time (measured in full-time equivalents, or FTEs) associated with collecting new data for the study and managing the data so they would be available for analysis.

^d“Data infrastructure” costs refer to purchases of computer hardware and software incurred by some states in order to collect data for the study.

^e“Encouraging participation” costs refer to Incentives, communication, and visits with clinics enrolled in the study. These were primarily FTE costs for efforts to keep clinics from dropping out of the study.

^f“Admin/Misc” costs refer to oversight and general administration not included in the other four categories listed above.

CERs: cost-effectiveness ratios; N/A: a dominated alternative [i.e., more costly, less effective).

^aGroup costs are those specific to each group. For example, learning sessions include the cost of hotels where the sessions took place. Most group costs were incurred after randomization.

^bCorresponds to the number of clinics included in the intention-to-treat analysis [see Figure 2).

^cCERs are calculated by dividing the difference between the costs by the difference between the effects. For example, “355” shown below for the clinic-level CER for waiting time is the result of dividing the difference between 2,878 and 1,329 (the cost/clinic of the coaching and interest circle call groups) by the difference between 4.610 and 0.253 (the reduction in days/clinic/month in the coaching and interest circle call groups).

^dCalculated by multiplying the clinic-level effect x number of clinics analyzed x average clinic size x a scaling factor of 14/12. (Annualized figures are reported, but changes occurred over a 14-month period.)

^eCalculated by multiplying the clinic-level effect x number of clinics analyzed x average clinic size x a scaling factor of 14/12. (Annualized figures are reported, but changes occurred over a 14-month period.)

^fCalculated as the average change in annual admissions between pre and post periods, multiplied by number of clinics analyzed.