Tracking the randomized rollout of a Veterans Affairs opioid risk management tool: A multi-method implementation evaluation using the Consolidated Framework for Implementation Research (CFIR)

Sharon A. McCarthy1,2, Matthew Chinman1,2,3, Shari S. Rogal2,4,5, Gloria Klima2, Leslie R. M. Hausmann2,4, Maria K. Mor2, Mala Shah2, Jennifer A. Hale2, Hongwei Zhang2, Adam J. Gordon6,7 and Walid F. Gellad2,4

Abstract

Background
The Veterans Health Administration (VHA) developed the Stratification Tool for Opioid Risk Mitigation (STORM) dashboard to assist in identifying Veterans at risk for adverse opioid overdose or suicide-related events. In 2018, a policy was implemented requiring VHA facilities to complete case reviews of Veterans identified by STORM as very high risk for adverse events. Nationally, facilities were randomized in STORM implementation to four arms based on required oversight and by the timing of an increase in the number of required case reviews. To help evaluate this policy intervention, we aimed to (1) identify barriers and facilitators to implementing case reviews; (2) assess variation across the four arms; and (3) evaluate associations between facility characteristics and implementation barriers and facilitators.

Method
Using the Consolidated Framework for Implementation Research (CFIR), we developed a semi-structured interview guide to examine barriers to and facilitators of implementing the STORM policy. A total of 78 staff from 39 purposefully selected facilities were invited to participate in telephone interviews. Interview transcripts were coded and then organized into memos, which were rated using the −2 to +2 CFIR rating system. Descriptive statistics were used to evaluate the mean ratings on each CFIR construct, the associations between ratings and study arm, and three facility characteristics (size, rurality, and academic detailing) associated with CFIR ratings. We used the mean CFIR rating for each site to determine which constructs differed between the sites with highest and lowest overall CFIR scores, and these constructs were described in detail.

1Veterans Integrated Service Network 4 Mental Illness Research, Education and Clinical Center, VA Pittsburgh Healthcare System, Pittsburgh, PA, USA
2Center for Health Equity Research and Promotion, VA Pittsburgh Healthcare System, Pittsburgh, PA, USA
3RAND Corporation, Pittsburgh, PA, USA
4Department of Medicine, University of Pittsburgh, Pittsburgh, PA, USA
5Department of Surgery, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA
6Informatics, Decision-Enhancement, and Analytic Sciences Center, VA Salt Lake City Health Care System, Salt Lake City, UT, USA
7Program for Addiction Research, Clinical Care, Knowledge, and Advocacy (PARCKA), Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, UT, USA

Corresponding author:
Sharon A. McCarthy, Center for Health Equity Research and Promotion, VA Pittsburgh Healthcare System, University Drive C, Pittsburgh, PA, USA.
Email: mccarthy@andrew.cmu.edu

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (https://us.sagepub.com/en-us/nam/open-access-at-sage).
Results
Two important CFIR constructs emerged as barriers to implementation: Access to knowledge and information and Evaluating and reflecting. Little time to complete the CASE reviews was a pervasive barrier. Sites with higher overall CFIR scores showed three important facilitators: Leadership engagement, Engaging, and Implementation climate. CFIR ratings were not significantly different between the four study arms, nor associated with facility characteristics.

Plain Language Summary: The Veterans Health Administration (VHA) created a tool called the Stratification Tool for Opioid Risk Mitigation dashboard. This dashboard shows Veterans at risk for opioid overdose or suicide-related events. In 2018, a national policy required all VHA facilities to complete case reviews for Veterans who were at high risk for these events. To evaluate this policy implementation, 78 staff from 39 facilities were interviewed. The Consolidated Framework for Implementation Research (CFIR) implementation framework was used to create the interview. Interview transcripts were coded and organized into site memos. The site memos were rated using CFIR’s −2 to +2 rating system. Ratings did not differ for four study arms related to oversight and timing. Ratings were not associated with facility characteristics. Leadership, engagement and implementation climate were the strongest facilitators for implementation. Lack of time, knowledge, and feedback were important barriers.

Keywords
implementation evaluation, implementation, substance abuse prevention, quantitative methods, mixed methods

Background
Many health systems, including the Veterans Health Administration (VHA), have taken proactive, multifaceted approaches to mitigate the harms of opioids in the United States. As part of these efforts, VHA developed the Stratification Tool for Opioid Risk Mitigation (STORM), a suite of provider-facing electronic reports which use a predictive model to identify patients at risk for opioid or suicide-related events (Oliva et al., 2017). Updated nightly, the STORM reports provide an estimated risk level for all Veteran patients, as well as additional information to assist clinicians to apply appropriate risk mitigation strategies tailored to individual patient risk factors and needs.

In March 2018, VHA released a national policy notice requiring that all facilities begin conducting “case reviews” of Veterans determined to be at highest risk based on the STORM model. Completing these case reviews entailed using a data tool, such as the STORM dashboard, to evaluate risk and determine the utility of providing risk mitigation strategies (e.g., referral to a pain specialist, providing naloxone). Completed case reviews and actions taken by the clinician(s) were documented using a standardized note in the VA electronic medical record.

The implementation of this policy notice was randomized to meet continuous improvement goals (Chinman et al., 2019; Oliva et al., 2017). Initially, all sites were required to review the top 1% of those at risk. Using a stepped-wedge design, VHA facilities were randomized to require an expanded number of case reviews (moving from top 1% of those at risk to top 5%) at either 9 or 15 months after the notice was released. In addition, VHA facilities were randomly assigned to receive a version of the policy notice requiring additional oversight if the site did not achieve a case review completion rate of 97% after 6 months. The rate was calculated as the number of completed case reviews of those Veterans identified by the STORM model as high risk (Chinman et al., 2019). This created four study arms, varying by oversight/no oversight and length of time before increasing the number of case reviews. This design allowed for examination of the effect of requiring case reviews for the top 1% versus 5% of high-risk Veterans and the impact of requiring additional oversight for facilities that failed to meet target goals for case review completion.

The first phase of an evaluation of this policy notice identified specific strategies associated with improved case review completion and found that being randomized to receive additional oversight did not impact the number or type of implementation strategies used to complete case reviews (Rogal et al., 2020). In the second phase of evaluation, described here, we analyzed 78 qualitative interviews representing 39 sites to examine barriers and facilitators to implementing the case review policy. This phase of evaluation aimed to (1) identify the barriers and facilitators for implementing case reviews; (2) assess variation in barriers and facilitators across the four study arms, and (3) evaluate the associations between Consolidated Framework for Implementation Research (CFIR) ratings and three facility characteristics: facility size, rural or urban location, and level of support provided through academic detailing.

The implementation of change in large-scale systems is especially complicated and requires a high level of training, multi-disciplinary team building, and the ability to identify barriers to change as they occur. (Mann & Lohrmann, 2019). VHA is one of few healthcare systems with the infrastructure to support an evaluation of these barriers and facilitators, with sites given substantial discretion as to how to set up their local implementation efforts. Thus, variability would be expected for both barriers and facilitators of the
implementation across sites. This evaluation, with both qualitative insight and multi-site data can inform future large-scale policy-driven implementation efforts.

Methods

This study was reviewed and approved by the VA Pittsburgh Health Care System’s Institutional Review Board. All participants provided informed consent prior to participating in the phone interview.

Sample Design and Participant Recruitment

Within each of the four study arms, we purposively sampled 10 sites, using performance on a national VA metric called the Opioid Therapy Guideline Adherence Metrics (OTG), which measures the extent a facility is following VHA national best practices for opioid treatment. Our aim with this sampling strategy was to interview sites with varying levels of baseline performance on opioid risk mitigation strategies. A statistician not associated with the interview process ranked sites in each arm based on their performance on the OTG metrics, and we selected 5 sites with the highest OTG scores and 5 with the lowest from each of the four study arms. If a site could not be reached or declined to participate, we moved to the next site on the list for that arm. Case completion rate was not available at the time of site selection, and was not considered in site selection.

We contacted the individual who served as the STORM point of contact (POC) at the selected sites by email, with phone and direct-messaging follow ups, and invited them to participate in one 45-min interview. A total of 63 sites were contacted to obtain the target sample of 40 sites. During the interview, the POC was asked to suggest one or two other individuals who could comment on the policy implementation at their site, and these individuals were then invited to participate. A total of 82 participants at 40 sites were interviewed, with 9 sites having 1 interview, 21 sites with 2 interviews, and 9 sites with 3 interviews. Two interviews could not be used due to poor audio quality and one site was dropped because they had not started implementation, resulting in a final total of 78 interviews from 39 sites for analysis.

Interview Development

The interview guide was developed using CFIR, a meta-theoretical framework for evaluating the factors that may influence implementation efforts (Damschroder & Lowery, 2013; Damschroder et al., 2009b, 2017). This framework was chosen because it is especially flexible and can be tailored to specific content. The CFIR has been used in other VA evaluations, including multi-site evaluations. (Bokhour et al., 2018; Damschroder & Lowery, 2013). As such, it was especially appropriate for this complex implementation effort (Gale et al., 2019).

CFIR consists of 32 constructs organized into 5 domains: implementation process, characteristics of the clinical intervention being implemented (in this case, the case reviews), characteristics of individuals implementing the clinical intervention, characteristics of the implementing organization (called inner setting), and the impact of forces outside of the organization (called outer setting). CFIR has an existing interview guide that includes questions about all 32 constructs which can be tailored for individual evaluations.

Sixteen of the most relevant constructs within four CFIR domains were selected for the interview guide in this study (Table 3). Data from the individual characteristics domain was not collected because we believed that factors at larger ecological levels were more critical, and asking about all domains would make the interview burdensome. Two research assistants experienced in qualitative interviews were trained to administer the interview, and the interview was piloted with clinicians from two facilities outside the study sample. All interviews were conducted by phone, recorded, transcribed, and validated by trained research staff. Interviews lasted approximately 45 min and were conducted from March 2019 to August 2019, approximately 1 year after the policy was released. Participants provided demographic information, including gender, time in VHA, time in their current role, type of training and clinical practice, and involvement with completing case reviews.

Interview Coding and Memo Development

A qualitative content analysis approach was followed (Forman & Damschroder, 2007), using the CFIR constructs as the structure. The coding was approached both deductively using CFIR constructs and inductively as codes emerged from the data. Following transcription and transcript verification, a codebook for the interviews was developed by an experienced qualitative coder (GK). The codebook was reviewed by the research team for completeness, and the coder coded the first 35 interviews using Nvivo 12 software. A second coder (MS) was trained to apply the codes in the codebook to the transcripts and double-coded five interviews. Three authors experienced in qualitative research (SM, GK, and MS) reviewed the double-coded interviews and resolved any differences by consensus, adapting the codebook and reviewing previous interviews to maintain consistency. The main coder continued to code interviews, with the secondary coder double coding every fifth interview to prevent drift in coding. All discrepancies were addressed by the qualitative team and significant questions were raised to the entire research team for resolution (Campbell et al., 2013; Damschroder et al., 2009b, 2017).
Individual interviews from a single site were combined to create a “memo”—text from all interviews at a site organized by CFIR construct—an analysis technique used to provide a more complete view of each site. In the memo, each individual CFIR construct is rated on a scale of −2, −1, 0, +1, and +2, with negative valence indicating the construct represented a barrier to implementation, and positive valence indicating the construct was a facilitator (https://cfirguide.org). Six team members were trained in the rating process following CFIR guidelines, and three teams of two reviewed site memos and rated the constructs using standard CFIR guidance and exemplars of each category. Rating pairs met to achieve consensus after independently rating constructs for a memo. The entire group of raters also met regularly to discuss any uncertainty in rating and to maintain consistency. As a further check for consistency, an experienced CFIR rater (SM) rated two memos from each team and a comparison showed a high degree of consistency and adherence to CFIR rating guidance. Over time, teams were varied such that coding was done by different partners, to assure continued consistency and prevent team drift in rating. In addition to the memo rating, the first author and one other author (GK) reviewed all memos and interviews qualitatively, as well as the coded data results (codes within the constructs), and the research team met regularly to discuss recurring themes, both within the constructs and across constructs.

Study Arms and Facility Characteristics

Study arm was operationalized using two dummy variables: one for the type of policy memo received (standard vs. increased oversight), and one for the timing of the increase in case review requests (early vs. late). Facility-level variables included the number of academic detailing visits, a measure of rurality, and a measure of facility complexity (an algorithm that considers patient risk, number and breadth of available specialists, intensive care unit availability, and teaching and research activities) (Chinman et al., 2015). Academic detailing is a defined support bundle provided by pharmacists within VHA who help clinicians improve prescribing using training, problem-solving, and data feedback, and was used to operationalize training and support for implementing case reviews.

Results

The 39 facilities targeted for interviews were similar in terms of facility characteristics to non-interview sites (Supplementary Table 1). The 78 interview participants from these 39 sites were majority women (58%) and covered a range of professional disciplines. Most participants had 1–5 years of experience in their current roles and were leaders in the STORM implementation efforts at their facility (Table 1). Six facilities were rural and 33 were urban, although many urban facilities have outpatient clinics in more rural locations. Geographically, 8 sites were Northeast, 9 were Midwest, 7 were West, and 15 were from the South.

CFIR ratings and associations between CFIR ratings and study arm, OTG, and facility characteristics: Means of the CFIR construct ratings across the 39 facilities and randomization arm are listed in Table 2. A positive score indicates the construct is facilitative for implementation, and a negative score indicates the construct is more of a barrier to implementation.

The CFIR ratings for each of the 16 CFIR constructs, the total CFIR scores, and the proportion of positive scores were not significantly different between the four randomization arms (Table 2). For example, overall, the mean CFIR rating in the Engaging construct was 0.6 across all 39 facilities, with a score of 0.6, 0.9, 0.5, and 0.4 across facilities in Arms 1
Table 1. Interview participant characteristics.

| Participant characteristics | N = 78 |
|-----------------------------|--------|
| Gender, n %                 |        |
| Female                      | 45 (58%) |
| Male                        | 33 (42%) |
| Training                    |        |
| Physician                   | 24 (31%) |
| Pharmacist                  | 28 (36%) |
| Psychologist or social worker | 15 (19%) |
| NP or nurse                 | 11 (14%) |
| Duration of VA role         |        |
| <1 year                     | 10 (13%) |
| 1–5 years                   | 48 (62%) |
| 6–10 years                  | 8 (10%)  |
| >10 years                   | 12 (15%) |
| Role in STORM               |        |
| Lead                        | 50 (64%) |
| Team member                 | 24 (31%) |
| No active role              | 4 (5%)   |

Note. STORM = Stratification Tool for Opioid Risk Mitigation.

Table 2. Consolidated Framework for Implementation Research (CFIR) ratings by randomization arm for 39 facilities.

| CFIR construct                  | Average percentage of positive ratings |
|---------------------------------|----------------------------------------|
|                                 | All arms Mean (SD) | Arm 1 No oversight and early increase, Mean (SD) | Arm 2 No oversight and late increase, Mean (SD) | Arm 3 Oversight and early increase, Mean (SD) | Arm 4 Oversight and late increase, Mean (SD) |
| Engaging                        | 0.6 (1.0) | 0.6 (1.0) | 0.9 (0.9) | 0.5 (1.1) | 0.4 (1.0) | 0.71 |
| Planning                        | 0.5 (1.1) | 0.4 (0.8) | 0.9 (1.0) | 0.5 (1.2) | 0.3 (1.3) | 0.66 |
| Reflecting evaluating           | -0.4 (1.0) | -0.8 (1.1) | -0.5 (0.7) | -0.2 (1.0) | 0.1 (0.9) | 0.17 |
| Tension for change              | 0.7 (0.5) | 0.7 (0.7) | 0.8 (0.4) | 0.7 (0.5) | 0.7 (0.5) | 0.91 |
| Implementation climate          | -0.1 (1.1) | -0.2 (1.2) | 0.3 (1.1) | -0.1 (0.7) | -0.2 (1.2) | 0.71 |
| Compatibility                   | 0.0 (1.1) | -0.2 (1.2) | 0.1 (1.0) | 0.2 (1.0) | -0.1 (1.1) | 0.84 |
| Structural characteristics      | -0.2 (0.9) | -0.3 (0.9) | -0.1 (0.9) | 0.2 (0.9) | -0.6 (0.7) | 25  |
| Networks communications         | 0.5 (0.9) | 0.5 (1.1) | 0.4 (1.1) | 0.2 (0.8) | 0.8 (0.7) | 0.57 |
| Leadership engagement           | 0.3 (1.1) | 0.0 (1.4) | 0.3 (0.8) | 0.1 (1.1) | 0.7 (1.0) | 0.64 |
| Relative priority               | 0.7 (0.9) | 0.2 (1.2) | 0.8 (0.4) | 0.9 (0.6) | 1.0 (0.9) | 0.27 |
| Available resources             | 0.1 (1.0) | 0.0 (1.1) | -0.3 (0.8) | 0.3 (1.0) | 0.4 (1.1) | 0.56 |
| Access to knowledge Information | 0.1 (0.9) | -0.2 (0.8) | 0.0 (1.1) | 0.2 (1.0) | 0.4 (0.7) | 0.42 |
| Evidence                        | -0.4 (0.8) | -0.2 (0.9) | -0.4 (0.8) | -0.6 (0.8) | -0.5 (0.8) | 0.78 |
| Relative advantage              | 0.2 (0.7) | 0.0 (0.7) | 0.2 (0.7) | 0.1 (0.7) | 0.4 (0.9) | 0.66 |
| Patient needs Resources         | 0.3 (0.7) | 0.5 (0.5) | -0.2 (0.8) | 0.5 (0.5) | 0.6 (0.7) | 0.09 |
| Peer pressure                   | 0.1 (0.7) | 0.0 (0.8) | 0.1 (0.6) | 0.1 (0.9) | 0.3 (0.7) | 0.79 |
| Total CFIR score (mean of 16 construct ratings) | 0.2 (0.5) | 0.1 (0.6) | 0.2 (0.4) | 0.2 (0.4) | 0.3 (0.5) | 0.77 |

P-value

| Proportion of scores in 16 STORM ratings mean (SD) | 0.28 | 0.77 | 0.41 |

Note. STORM = Stratification Tool for Opioid Risk Mitigation.
## Table 3. Joint display of quantitative and qualitative data about STORM barriers and facilitators.

| CFIR domain/constructs rated as facilitator/rated as barrier | Typical facilitator quotes | Typical barrier quotes | Frequency of CFIR ratings* | Mean rating for bottom/ top quartiles (difference score) |
|--------------------------------------------------------------|-----------------------------|------------------------|---------------------------|---------------------------------------------------------|
| **Intervention characteristics**                             |                             |                        |                           |                                                         |
| Evidence strength and quality                                | Q: Are you aware of any evidence that shows that completing these case reviews will result in better outcomes for Veterans? | “We do have… reports that we’ve pulled… they’re showing a downward trend of any increases in MEDDS, which is an overall a good thing.” | Frequency of CFIR ratings (N=38) | Mean rating |
| (strong personal or research evidence to support STORM case review process/was not aware of evidence) | | “I’m not aware of any data that we’ve been presented with thus far that shows, at least that specific clinical outcomes are linked to implementation of the Notice, no.” | −2 1 0 1 2 | −0.7 |
| Relative advantage                                           | Q: How does the case review process required by the STORM notice compare to other similar existing processes or tools in your setting? | “…now that I’m familiar with it, I think that it is just very user friendly and it presents information in a way that you can make quick sense out of… I think… the user interface is great, is much better.” | Frequency of CFIR ratings (N=38) | Mean rating |
| (saw strong advantages to STORM/saw few advantages to STORM) | | “… everybody’s time is already stretched- so they may perform a review on someone that, due to the algorithm used to pull the data, may not lead to clinical intervention and so there’s some lost time involved in that process.” | 0 6 20 11 1 | 0 |
| Outer setting                                                 | Q: Are you aware of how your team is doing, compared to other medical centers…? | “I think we are, aware of that… You can look at some of the… reports… To see where we are, compared to everyone else.” | Frequency of CFIR ratings (N=39) | Mean rating |
| Peer pressure (aware of the performance of other facilities on STORM/not aware or did not care how other sites were performing) | | “No… when I go in and look at my metrics, you can see ours and then national… that’s the only other comparison that I’ve looked at.” | −2 −1 0 1 2 | −0.2 |
| Patient needs and resources (felt the STORM process worked to support patient needs/did not think STORM supported patient needs) | Q: “How well do you think the STORM Notice is helping to meet the needs of your Veterans? | “I think pretty well… it’s really good conversation. I think things are identified from a medical perspective… like, “oh, look at this guy’s… level of this or we should be thinking about that” …I think it’s beneficial. | Frequency of CFIR ratings (N=39) | Mean rating |
| | | Not an important barrier | −2 −1 0 1 2 | 0.2 |

(Continued)
### Table 3. (Continued)

| CFIR domain/constructs rated as facilitator/rated as barrier | Typical facilitator quotes | Typical barrier quotes |
|-------------------------------------------------------------|----------------------------|------------------------|
| **Inner setting**                                           | And it, it puts eyes on something that otherwise maybe people are not really paying much attention to” | |
| **Structural characteristics**                             | Q: How was the STORM notice implemented in CBOCs? | |
| (felt roll out was equally successful in both main facility and satellite facilities vs. saw special challenges at community-based care settings) | Not a facilitator | “Some of the patients have difficulty getting to the main campus... for Pain, Pain Management follow-ups and stuff. Or if they need methadone or methadone or suboxone. It’s difficult for them to get to the main campus which is where... we’re doing it.” |
| **Networks and communications**                            | Q: How do you communicate the results of the case reviews back to providers? | |
| (described positive and persistent interactions between different teams and departments/described challenges between teams and lack of ongoing communication) | “It’s communicated back with the Opioid Alert in (the medical record) and then we also send an email... so it’s a little more private, a little more personal. Then, if they want to discuss-to feel free to reach out to us. If it’s something that’s a critical situation...we’ll actually reach out to the provider by phone or instant message, something where we can get their attention a little bit quicker...” | “So, what we do is reach out to the provider so add that provider as a co-signer...So, when it comes to the community providers, it’s a little bit challenging... because I cannot add that provider as a co-signer...So now I need to contact the provider and it’s hell to get in touch with those providers. It’s very difficult.” |
| **Tension for change**                                      | Q: Why do you think the STORM notice was implemented? | |
| (saw a strong need to change the current situation/saw little need) | “...there’s a recognition that the population with chronic pain who are on opioids, are at risk, both for their mental and physical...” | Not an important barrier |

### Frequency of CFIR ratings* Mean rating for bottom/ top quartiles (difference score) |

|                      | Frequency of CFIR ratings (N=38) | Mean rating |
|----------------------|-----------------------------------|-------------|
| Q: How was the STORM notice implemented in CBOCs? | -2 -1 0 1 2 | -0.4 -0.1 (0.3) |
| Q: How do you communicate the results of the case reviews back to providers? | -2 -1 0 1 2 | -0.5 1.1 (1.6) |
| Q: Why do you think the STORM notice was implemented? | -2 -1 0 1 2 | -0.5 1.1 (1.6) |
Table 3. (Continued)

| CFIR domain/constructs rated as facilitator/rated as barrier | Typical facilitator quotes | Typical barrier quotes | Frequency of CFIR ratings* | Mean rating for bottom/top quartiles (difference score) |
|-------------------------------------------------------------|----------------------------|------------------------|---------------------------|--------------------------------------------------------|
| Implementation climate (positive attitudes toward change and new STORM processes/ resistance to STORM within the organizational culture) | “...I have heard from leadership that they find it a valuable and kind of a crucial part of the care that we provide at our facility. ...Leadership in several departments have told me directly that this is important work and that they're willing to help us block these clinicians' time to make sure this works happens, because it's important.” | “I would say, not very receptive. ...they already have so many duties that they're trying to juggle in any given day. I haven't found that people are not willing to do it, but...I am doing a lot of the reviews myself...Just to make sure that they get done.” | Frequency of CFIR ratings (N=39) | Mean rating |
| Q: How receptive are people in your organization to implementing the STORM Notice? | 3 11 12 11 2 | -0.9 | 0.7 |
| | | | (1.6) |
| | “We were already meeting regularly. So, really it was just a matter of verifying that what we were already doing was meeting the intent...of the Initiative.” | “...one of the bigger issues...was because the facility would not allot us time to actually do it. ...we're doing it on our lunch break which, I don't even get a lunch break...So there are things that I feel like they should or could be doing that they're not doing to accommodate.” | Frequency of CFIR ratings (N=39) | Mean rating |
| Q: How has doing the case reviews been integrated into the current workflow? | 2 12 12 10 3 | -0.7 | 0.8 |
| | | | (1.5) |
| | “Well, we understood that it was very important. And...they told us that we needed to prioritize getting these reviews underway. So, we understood that we had to get it done and get it done quickly.” | “Well...a lot of primary care people don't take it seriously. But, a lot of them don't take pain seriously...” | Frequency of CFIR ratings (N=38) | Mean rating |
| Q: Compared to other high-priority initiatives going on at your medical center, how important was it to comply with the STORM notice? | 0 5 6 22 5 | 0 | 1.3 |
| | | | (1.3) |
Table 3. (Continued)

| CFIR domain/constructs rated as facilitator/rated as barrier | Typical facilitator quotes | Typical barrier quotes |
|-------------------------------------------------------------|-----------------------------|------------------------|
| Leadership engagement                                       | Q: What level of endorsement or support have you seen or heard from leadership? | Frequency of CFIR ratings (N=39) | Mean rating |
| (feel leadership is engaged and offer strong support for carrying out the STORM case reviews/did not see leadership support) | “we had… full support from leadership …which is perfect. I mean, when you bring it up to them, ‘Look, this is something that we must do, and you have to provide time for the providers to be there for the review.’ They’ve been fantastic. The providers have been able to have participated in every single one. And, the time when we first implemented the STORM reviews, again, leadership was absolutely pivotal in allowing us to do it in one day, so that we could become compliant immediately.” | 2 9 9 15 4 | .9 |
|                                                             | “Leadership didn’t really jump in there even though they were on quite a number of the emails. They didn’t jump in and say, ‘look guys, this is required. …it’s not something that we can just… say, ‘oh well’…I don’t think they stressed the importance of it.” | 2 9 9 15 4 | 1.2 |
| Available resources                                         | Q: What resources have you needed to implement and carry out the case reviews? (time, training, space, support from others) | Frequency of CFIR ratings (N=30) | Mean rating |
| (felt that needed resources were/were not available)        | “I think we need the time… the space is easy to find, it’s more so people need the time to do it… people need the time and they need the administrative support” | 0 10 10 7 3 | -0.5 |
|                                                             | “I don’t really remember any training.” | 0 10 10 7 3 | 0.8 |
|                                                             | “I feel like… I’m educated and I’m not…. I don’t understand where this came from and why, and what the point is…” | 0 10 10 7 3 | 1.3 |
| Access to knowledge and information                         | Q: “Did you feel the training prepared you to carry out the roles and responsibilities expected of you?” | Frequency of CFIR ratings (N=39) | Mean rating |
| (felt that adequate and appropriate training was provided /did not receive training or felt training or support were inadequate) | “Yes. I do… as far as like completing the note and doing the reviews, yes. I think it did an excellent job.” | 1 11 10 17 0 | -0.9 |
|                                                             | “I don’t really remember any training.” | 1 11 10 17 0 | 0.6 |
|                                                             | “I feel like… I’m educated and I’m not…. I don’t understand where this came from and why, and what the point is…” | 1 11 10 17 0 | 1.5 |

Process

| Planning (clear description of the planning process) | Q: Were you involved in developing a plan for implementing the STORM case reviews at your facility? Could you describe the plan? | Frequency of CFIR ratings (N=39) | Mean rating |
|-----------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------|-------------|
| CFIR domain/constructs rated as facilitator/rated as barrier | Typical facilitator quotes | Typical barrier quotes | Frequency of CFIR ratings* | Mean rating for bottom/ top quartiles (difference score) |
|------------------------------------------------------------|-----------------------------|------------------------|----------------------------|----------------------------------------------------------|
| with a quality plan / little formal planning, no final plan in place | “…at our facility level… we sat down with the primary leads on our pain committee…and a couple other heads of different programs… hashing out what makes sense, how to be proactive about this, how to have it ready… when the directive came down. And that led us to have a process roughly in place…” | “It was a bit haphazard. I would have expected that there would have been more, a little more direction about the expectations of completion of the STORM report. And then, …evaluation of actually the resources needed, but—.” | -2 1 0 2 | -2 1 0 2 |
| Engaging (strong, appropriate team or individuals were engaged in the rollout/ the team lacked the right individuals to be successful) | Q: Who else is involved with leading the STORM notice implementation at your site? | Frequency of CFIR ratings (N=39) | Mean rating |
| | | -2 1 0 2 | 1 6 4 24 4 | -0.5 1 3 |
| | “The team for the risk reviews is two psychiatrists, a clinical pharmacist, a registered nurse, myself-so we got four disciplines. And it gives us an opportunity to look at the case holistically from a lot of different perspectives… so I think it’s a better, well-rounded review.” | “I think some of the team members are not as engaged as I would think that they should be, to make sure that they’re done, cause sometimes they show up and then sometimes they don’t. And I feel like sometimes it’s like pulling teeth.” | |
| Reflecting and evaluating (were aware of feedback on STORM, felt feedback was appropriate/were not aware of feedback, or felt feedback was inaccurate or unhelpful) | Q: Tell me about the feedback reports that you receive from VA Central Office about your progress implementing the case reviews? | Frequency of CFIR ratings (N=39) | Mean rating |
| | | -2 1 0 2 | 5 12 15 6 1 | -1.1 0 |
| | “So, Our… Technical Person, He Puts Together All The Reports And Then Our Pharmacist Executive…, She Went Over It With Us And I Saw That We Were At 100 Compliance…” | “It Left Me Very Confused Because We Track The Number Of Patients…That We Reviewed, And It Was Wildly Different Than What Was Listed In The Report. And, I Remember Thinking, ‘That’s Really Weird.’” | |

Note. CFIR = Consolidated Framework for Implementation Research; STORM = Stratification Tool for Opioid Risk Mitigation.

*Each of the facilities received a rating for this construct, from −2 to +2. The numbers in each row reflect the number of facilities that received each rating.
While most constructs functioned as both barriers and facilitators across the sample, some had primarily negative or positive ratings. For example, Evidence strength and quality was generally a barrier, with positive ratings (+1) at only six sites. Overall, comments show little awareness of research or personal evidence to support the completion of case reviews. The Peer pressure construct had only 13 positive ratings (+1) (33%), and comments showed little evidence that this was either a barrier or facilitator, suggesting little interest in or awareness of how other medical centers nationally were doing on the STORM measures.

Patient needs and resources and Tension for change were generally perceived positively. Patient needs and resources was rated at nearly all sites as neutral (n = 19) or positive (n = 15), and was rated negatively at only four sites. Likewise, 29 of 39 sites (74%) described positive, though moderate (no sites were rated +2) Tension for change, indicating a belief in a clear need for attention to opioid prescribing.

Access to knowledge and information was generally perceived to be an implementation barrier; only 17 of 39 sites rated this positively, and many quotes described a lack of training or a lack of awareness of training (e.g., see Table 3). Similarly, Reflecting and evaluating was only rated moderately positively (+1) by 6 sites, and many comments described little or confusing feedback on their progress.

The three CFIR constructs with the largest difference scores between the top and bottom CFIR quartiles are Leadership engagement, Engaging, and Implementation climate. In addition, the Available resources construct was repeatedly identified by the qualitative team review as an important determinant of implementation and was therefore included in further analyses. These four constructs are described in more detail below.

Leadership Engagement

The Leadership engagement construct had the largest difference score of 2.1, reflecting a high difference in mean rating between the top and bottom quartile facilities (1.2 vs. −0.9). Lack of leadership engagement was a notable barrier for low quartile sites, shown in the following quote where explaining the case review process to leadership is both frustrating and time consuming. When asked “What level of endorsement or support have you seen or heard from leadership?”, the team leader commented: “I’ve heard none…we’ve had to advocate for what we needed… I’ve been questioned, like, ‘Why do so many people need to be in the room?’” Like ‘Why do you need more than one doctor to do that kind of work?’ And it just says to me, ‘Oh, my God, you don’t appreciate the complexity of this work.’” Asked the same question about level of support, one participant in a low quartile site simply said: “I haven’t heard any.”

In contrast, facilities with overall high mean CFIR scores described supportive leadership as facilitating their implementation efforts. One participant described having time protected to conduct case reviews and said that their supervisors were, “very supportive of this as well. …if I needed to go to them and say, ‘hey, I need more time blocked to do these STORM reviews’…they would do that.”

Engaging

The Engaging construct had a difference score of 1.8, with a CFIR mean rating from the bottom quartile of −0.5 and a CFIR rating from the 10 sites in the top quartile of 1.3. Although Engaging was rated moderately positively across all sites (overall mean rating = .6), the ability to form and work effectively with a team to perform case reviews clearly differed between sites with high versus low mean CFIR ratings.

The Engaging construct was operationalized by asking whether appropriate individuals were involved in the rollout and how available and engaged they were in establishing the process. Having a team in place was a strong facilitator for sites, where the addition of the STORM policy was often rolled into the existing process with minimal effort. Sites stated having a preexisting opioid prescribing safety program or team made it easy to implement the STORM policy. A staff member at one of the top quartile sites noted: “So…it had already been in the works… we had a lot of flow maps already written down. I think we had a charter in place. And, the STORM just became one more thing we were already…working on.”

For these sites, the process was incremental and involved changes to existing processes rather than the creation of new ones. One staff member said: “it was a little easier for us because we were already doing these things… so now we’re in good shape and …it was easy to just say…We have to change our note… “. Although pre-existing teams had to adjust to a new method of review, they had the existing structure and skill to respond to the notice, making the demands on these sites quite different from sites without existing teams.

In sites without existing teams, a single individual or very small group might attempt to do the case reviews, often with mixed results, depending on the number of case reviews they were asked to complete. Consistent with the Engaging construct, the sheer effort of creating the team, including finding the right members, learning the needed skills, and developing a process to complete the reviews was an implementation barrier for some sites. One individual in a low quartile site noted: “we felt pressured and we…also didn’t know how we were going to gather all that information and put it in the note, so that took a little bit of time. So, if we had all that worked out beforehand, it would have been a much easier, much simpler process. More effective.”
A related challenge was team development. The policy notice recommended an interdisciplinary team, and across interviews, 27 different roles were mentioned as participating in the case review process, with professional backgrounds varying from medical doctors and nurse practitioners to recreational therapists and podiatrists. The complexity of creating such an interdisciplinary team, in a large and sometimes siloed environment like the VA, was considerable. Participants described difficulty finding members to participate, challenges with identifying time to work on reviews, and communication struggles between disciplines. Forming new teams often created an implementation barrier, as evidenced by the frustration from this interviewee: “…we’ve tried to partner with the suicide prevention people and they’re overwhelmed too. They’re like … oh we can’t take responsibility for that. Don’t tag us onto the note. Don’t tell us… Yeah, so it’s like, well if not you, who?”

Another participant noted a lack of role clarity around who should take responsibility for establishing and maintaining the team, stating: “There is that push and pull, ‘This should be Mental Health’s baby. No, this should be Primary Care’s baby. No, this should be Pain Clinic’s baby.’ ...I think it should be an interdisciplinary team.”

Implementation Climate

The Implementation climate construct had a difference score of 1.6, with a mean CFIR rating from the bottom quartile of 0.9, and a CFIR rating from the top quartile of 0.7. This was the third highest difference score, tied with Networks and communications. However, the memos and research team discussion identified Implementation climate as an important indicator of the overall implementation at a site. The construct was assessed by asking “How receptive are people in your organization to implementing the STORM Notice?” In low quartile sites, participants described organizational resistance: “The people in our organization unfortunately are not that interested to do new things. Unfortunately.” Specific implementation climate barriers, as reported by participants from two low quartile sites, included provider perception of burden and the difficulty of managing pain: “People are pretty resistant to implementing STORM, as well as most OSI initiatives, even those required by state law...because everybody feels like it is more work”. Another stated: “look, a lot of people just hate dealing with pain. They don’t know how. They don’t like doing it. And, anything with opioids and Tramadol and now Lyrica makes it harder for them.”

In a high quartile site, with a more receptive climate, one participant noted: “I think at first, just like with any new notice, (they) sort of, groaned a little bit and... dragged their feet but I think we’re, we’re good now. I think our facility is really receptive.” Sites in the top quartile were more likely to have provided a positive context for implementing the STORM notice.

Available Resources

The Available resources construct showed a difference of 1.3 in ratings between sites in the top (mean rating = 0.8) versus bottom (mean rating = −0.5) quartiles of overall CFIR rating. Available resources is a composite of physical and time resources, which may have impacted the overall difference score. While most participants stated they had adequate space and equipment to carry out the reviews, many described a lack of staff and time for the work. Thus, while potentially not reflected in the overall difference ratings, insufficient time to carry out the reviews was the most pervasive barrier experienced across sites. This was reflected in direct questions about resources and indirectly mentioned frequently throughout the interviews. Staff with less flexibility in their schedule sometimes completed the reviews before their workday started or during their lunch break. Many simply stated that they “fit it in somehow” to their day. Although many described the importance of completing the work, in some cases this lack of time created resentment from staff.

Interviews in the low quartile sites demonstrated this concern, with one staff member stating: “…the problem is that we get zero dedicated time. And we’ve asked leadership several times, and it just falls on deaf ears so, no, we get zero dedicated time to do this.” Many were adding this to an already full workload, as described by this individual: “I’m also a provider...I have a full patient load...And there’s not really time to do all this.” In a few cases, clinic time slots were blocked so providers could complete case reviews, but more typically, and still infrequently, they were only given time to attend the team meeting to discuss the reviews. When providers had flexibility in their schedule to complete the reviews, they described this as a facilitator.

Other CFIR constructs were barriers or facilitators in some cases, as described in Table 3, but were less pervasive or intense than those identified above. This is evident from the qualitative review as well as the percentage of positive ratings provided for each construct.

Discussion

In this evaluation of the randomized rollout of a policy requiring case reviews at VHA facilities, we identified key implementation barriers and facilitators, although the differing policies for oversight and pacing did not seem to influence what sites were needed for successful implementation. In addition, characteristics such as facility size, complexity, rurality, and implementation resources (i.e., academic detailing) were not associated either positively or negatively with barriers and facilitators. The predominant facilitators were strong and appropriate engagement, supportive leadership, and a positive implementation climate. Important barriers included lack of time to complete the case reviews and a perceived lack of evidence for the intervention. The mixed-methods approach and use of joint displays is a
novel approach that enabled the analysis of a large volume of qualitative data. The combination of ratings and coded data helped to develop a strong picture of the barriers and facilitators for this intervention.

The differing study arms appeared to have no effect on implementation barriers and facilitators. Having increased oversight and requiring sites to complete action planning did not seem to impact how facilities implemented the notice. This finding replicates earlier work indicating that oversight did not change the number or type of strategies used for the implementation (Rogal et al., 2020). The oversight could have been viewed as both positive and negative, since it included assistance from the national office, perhaps confounding the effect of this contingency. The other randomization condition—altering the timing of requiring more case reviews—also did not impact implementation barriers and facilitators, even though some facilities experienced a large increase in the number of required case reviews prior to their interviews. Because facilities were not told in advance about the increase, it is not surprising that this evaluation, which focused on the process for completing the case reviews, did not show differences between time frames. Because the interviewers were intentionally blinded to the condition of the interviewees, we did not ask specific questions about this increase, and it was seldom mentioned spontaneously.

We were surprised to find no association between implementation experience and the facility characteristics explored. It might be expected that facilities with higher levels of care (complexity) or more training (academic detailing) would develop different, perhaps better, approaches to this implementation. This difference has previously been demonstrated, including work on following clinical practice guidelines at the VA, where less complex, Western, more rural sites, had less implementation of practice guidelines (Buscaglia et al., 2015). It seems possible that the high level of interdisciplinary work required by this notice cut across those simple characteristics and favored a configuration of resources, leadership, climate, and engaging to maximize positive implementation. This need for synergy in the implementation of complex innovations is also described by Rapp et al. (2010).

We identified four CFIR constructs as facilitators: from the CFIR Process Domain, the Engaging construct, and from the Inner Setting Domain, Readiness for Implementation sub-domain: Leadership engagement, Implementation climate, and Available resources. Two constructs were important barriers: in the “Readiness for implementation” sub-domain, Access to knowledge and information and from the Process domain: Reflecting and evaluating. This constellation of barriers and facilitators suggests a lack of readiness for the implementation at many sites, which relates directly to Implementation climate. The concept of implementation climate was originated by Klein et al. (2001), and is directly related to employee perceptions of how the innovation will be supported by leadership and supported with resources and training. The complexity of the case review process further intensified the need for planning and preparation for this implementation. As Damschroder notes, the need for clearly detailed implementation planning only increases with the complexity of the intervention (Damschroder et al., 2009a, extra file 4).

Implementing the STORM case reviews required sites across the country with varying capacities to respond to a very specific and complex task: completing interdisciplinary case reviews and successfully logging completion of reviews using a note in the electronic medical record. The task required team creation and management, pain expertise, and technical skills in managing the case review note and understanding the STORM tool. Further, the task was charged to a group already stretched by clinical demands, and each site was asked to develop an individualized approach to completing the task, rather than being provided with specific tools and task assignments. While sites developed many different approaches to the task, some had greater resources and capacity to accomplish this, and some were simply exhausted and frustrated. Although recent work by Kim et al. (2020) points to the importance of heterogeneity in implementation efforts, the complexity of this innovation points to a need for greater task clarity and specificity. In addition, although the task was complex, the evaluation of the task was extremely simple, a correctly titled note in the medical record. Fidelity to this task was not evaluated, and some participants commented that there was likely to be great variation in the depth and quality of this note. Finally, as Greenhalgh et al. (2004) stress, it is important to recognize the potential impact of sociopolitical context in which this implementation was taking place, as this was a time of political stress and resource challenges for the VHA system, with calls for privatization and changes to allow outside providers to serve Veterans.

While this was a novel approach to evaluation with several notable findings, there were limitations. First, only a limited number of CFIR constructs can be included in an interview without becoming overwhelming, and it is possible that some important constructs were omitted. Second, 23 sites declined to participate in the qualitative interview, possibly leading to a selection bias. However, participating and non-participating facilities did not significantly differ on objective measures, and a 45-minute interview may simply have seemed too challenging for busy providers to accommodate. Third, the analytical strategy of comparing the sites with the top and bottom quartiles of overall CFIR scores may have masked important constructs that functioned in the moderately rated sites. However, we believe comparing the top and bottom quartile allows us to glean useful barriers and facilitators that clearly differentiated sites. Fourth, sites were characterized on the basis of only a few individuals, although every effort was made to ensure these individuals were those most knowledgeable about this implementation. Fifth, there was variability regarding the point in time that the interview was...
conducted relative to the implementation. This may have influenced the nature of barriers and facilitators reported. Finally, an objective implementation measure was not used in any analyses. Although the case review completion rate was the designated implementation measure for each site, mapping this measure on to the memo ratings proved unreliable due to the differing temporal relationship between the interviews and the completion rate, calculated quarterly. In addition, the measure did not quantify the actual number of case reviews completed but instead the proportion of reviews that were completed for a site over time. The outcomes of this policy initiative are being evaluated, and early evidence shows positive clinical outcomes for Veterans who are identified as at risk by the STORM dashboard. These Veterans are more likely to have a case review completed and to have risk mitigation strategies put in place (Strombotne et al., 2021).

In conclusion, this evaluation of a national implementation identified key barriers and facilitators across multiple implementation sites. Although we found no difference in implementation barriers and facilitators across randomization arms, the evaluation demonstrated the value of strong, supportive leadership and climate, realistic expectations about time, and engaging the right people in creating a positive experience for implementation. Further, a perceived lack of training and accurate, well-explained feedback were barriers to the process. In future large-scale, nationwide implementations it would be constructive to consider the overall readiness for implementation, including the development of strong leadership implementation (Bonham et al., 2014). In addition, more proactivity (Birken et al. 2013) and comprehensive training might help with the improved adoption of an intervention of this complexity, as shown in much previous research (Damschroder et al., 2009b; Greenhalgh et al., 2004; Phillips & Allred, 2006). Finally, the findings suggest the limits to using official policies promising low-level consequences as a strategy to overcome implementation barriers.

Conclusion

In evaluating the randomized rollout of a policy requiring case reviews at VHA facilities, we did not find any association between study arms and implementation barriers and facilitators. Facilitators were strong engagement of appropriate individuals, engaged and supportive leadership, and a positive implementation climate. Lack of time to complete the case reviews and a perceived lack of evidence for the intervention were barriers. The evaluation used a mixed-methods approach and joint displays, an innovative approach that enabled the analysis of a large volume of data.

Acknowledgments

We acknowledge that this work would not be possible without the cooperation and support of our partners in the Office of Mental Health and Suicide Prevention and in the HSR&D-funded Partnered Evidence-Based Policy Resource Center. The contents of this paper are solely from the authors and do not represent the views of the Department of Veterans Affairs or the United States Government.

Authors’ contribution

SM, WG, GK, and MS designed and conducted the interviews. SM, MC, MS, GK, WG, and LH rated memos. HZ, MM, GK, SR, and SM analyzed the data. SM drafted the manuscript. JH provided exceptional administrative support and AG provided significant editing. All authors worked on study conception and design, interpretation of data, and critical review and approval of the manuscript.

Availability of Data and Materials

Please contact the corresponding author.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethics approval and consent to participate

The VA Pittsburgh Healthcare System approved this research study.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Health Services Research and Development, (grant number SDR 16-193).

ORCID iD

Sharon A. McCarthy https://orcid.org/0000-0002-7299-5129

Supplemental material

Supplemental material for this article is available online.

References

Birken, S. A., Lee, S. Y., Weiner, B. J., Chin, M. H., & Schaefer, C. T. (2013). Improving the effectiveness of health care innovation implementation: Middle managers as change agents. Medical Care Research and Review, 70(1), 29–45. https://doi.org/10.1177/1077558712457427

Bokhour, B. G., Fix, G. M., Mueller, N. M., Barker, A. M., Lavela, S. L., & Hill, J. N., Solomon J. L., Lukas, C. V. (2018). How can healthcare organizations implement patient-centered care? Examining a large-scale cultural transformation. BMC Health Services Research, 18(1), 168. https://doi.org/10.1186/s12913-018-2949-5

Bonham, C. A., Sommerfeld, D., Willging, C., & Aarons, G. A. (2014). Organizational factors influencing implementation of evidence-based practices for integrated treatment in behavioral health agencies. Psychiatry Journal, Epub 2014 Mar 3. 802983. https://doi.org/10.1155/2014/802983
Buscaglia, A. C., Paik, M. C., Lewis, E., & Trafton, J. A., & VA Opioid Metric Development Team (2015). Baseline variation in use of VA/DOD clinical practice guideline recommended opioid prescribing practices across VA health care systems. *The Clinical Journal of Pain, 31*(9), 803–812. https://doi.org/10.1097/AJP.000000000000016013

Campbell, J. L., Quincy, C., Osserman, J., & Pedersen, O. K. (2013). Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement. *Sociological Methods & Research, 42*(3), 294–320. https://doi.org/10.1177/0049124113500475

Chinman, M., Gellad, W. F., McCarthy, S., Gordon, A. J., Rogal, S., Mor, M. K., & Hausmann, L. (2019). Protocol for evaluating the nationwide implementation of the VA stratification tool for opioid risk management (STORM). *Implementation Science, 14*(1), 5. https://doi.org/10.1186/s13012-019-0852-z

Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., & Lowery, J. C. (2009a). Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. *Implementation Science, 4*, 50. https://doi.org/10.1186/1748-5908-4-50, extra file 4

Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., & Lowery, J. C. (2009b). Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. *Implementation Science, 4*, 50. https://doi.org/10.1186/1748-5908-4-50

Damschroder, L. J., & Lowery, J. C. (2013). Evaluation of a large-scale weight management program using the consolidated framework for implementation research (CFIR). *Implementation Science, 8*(51). https://doi.org/10.1186/1748-5908-8-51

Damschroder, L. J., Reardon, C. M., Sperber, N., Robinson, C. H., Fickel, J. J., & Oddone, E. Z. (2017). Implementation evaluation of the telephone lifestyle coaching (TLC) program: Organizational factors associated with successful implementation. *Translational Behavioral Medicine, 7*(2), 233–241. https://doi.org/10.1007/s13142-016-0424-6

Forman, J., & Damschroder, L. (2007). *Empirical methods for bioethics: A primer*. Emerald Group Publishing Limited.

Gale, R. C., Wu, J., Erhardt, T., Bunthavong, M., Reardon, C. M., Damschroder, L. J., & Midboe, A. M. (2019). Comparison of rapid vs in-depth qualitative analytic methods from a process evaluation of academic detailing in the veterans health administration. *Implementation Science, 14*(1), 11. https://doi.org/10.1186/s13012-019-0853-y

Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., & Kyriakidou, O. (2004). Diffusion of innovations in service organizations: Systematic review and recommendations. *The Milbank Quarterly, 82*(4), 581–629. https://doi.org/10.1111/j.0887-378x.2004.00325.x

Guetterman, T. C., Fetters, M. D., & Creswell, J. W. (2015). Integrating quantitative and qualitative results in health science mixed methods research through joint displays. *Annals of Family Medicine, 13*(6), 554–561. https://doi.org/10.1370/afm.1865

Kim, B., Sullivan, J. L., Ritchie, M. J., Connolly, S. L., Drummond, K. L., Miller, C. J., Greenan, M. A., & Bauer, M. S. (2020). Comparing variations in implementation processes and influences across multiple sites: What works, for whom, and how? *Psychiatry Research, 283*, 112520. https://doi.org/10.1016/j.psychres.2019.112520

Klein, K. J., Conn, A. B., & Sorra, J. S. (2001). Implementing computerized technology: An organizational analysis. *The Journal of Applied Psychology, 86*(5), 811–824. https://doi.org/10.1037/0021-9010.86.5.811

Mann, M. J., & Lohrmann, D. K. (2019). Addressing challenges to the reliable, large-scale implementation of effective school health education. *Health Promotion Practice, 20*(6), 834–844. https://doi.org/10.1177/1524839919870196

Oliva, E. M., Bowe, T., Tavakoli, S., Martins, S., Lewis, E. T., Paik, M., Wiechers, I., Henderson, P., Harvey, M., Avoundjian, T., Medhanie, A., & Trafton, J. A. (2017). Development and applications of the veterans health administration’s stratification tool for opioid risk mitigation (STORM) to improve opioid safety and prevent overdose and suicide. *Psychological Services, 14*(1), 34–49. https://doi.org/10.1037/ser0000099

Phillips, S. D., & Allred, C. A. (2006). Organizational management: What service providers are doing while researchers are disseminating interventions. *The Journal of Behavioral Health Services & Research, 33*(2), 156–175. https://doi.org/10.1007/s11414-006-9016-4

Rapp, C. A., Etzel-Wise, D., Marty, D., Coffman, M., Carlson, L., Asher, D., Callaghan, J., & Holter, M. (2010). Barriers to evidence-based practice implementation: Results of a qualitative study. *Community Mental Health Journal, 46*(2), 112–118. https://doi.org/10.1007/s10597-009-9238-z

Rogal, S. S., Chinman, M., Gellad, W. F., Mor, M. K., Zhang, H., McCarthy, S. A., Mauro, G. T., Hale, J. A., Lewis, E. T., Oliva, E. M., Trafton, J. A., Yakovichenko, V., Gordon, A. J., Hausmann, L. R. M. (2020, June 23) Tracking implementation strategies in the randomized rollout of a Veterans Affairs national opioid risk management initiative. *Implement Sci, 15*(1), 48. https://doi.org/10.1186/s13012-020-01005-y

Strombo, K., Legler, A., Minegishi, T., Trafton, J. A., Oliva, E. M., & Lewis, E. T. (2021, June 14–17). Reducing adverse events from opioid prescriptions in the Veterans Health Administration: a stepped wedge cluster randomized controlled trial [Paper presentation]. Annual Research Meeting of AcademyHealth. Virtual.
**Appendix 1**

See Table A1.

**Table A1.** Relationship between CFIR constructs and three randomized variables.

|                          | Oversight | No oversight | P-value | OTG high | OTG low | P-value | Early increase in case reviews | Late increase in case reviews | P-Value |
|--------------------------|-----------|--------------|---------|----------|---------|---------|-------------------------------|-------------------------------|---------|
| **Mean CFIR rating**     | 0.3 (0.4)| 0.1 (0.5)    | .44     | 0.2 (0.5)| 0.2 (0.5)| .96     | 0.3 (0.4)                    | 0.1 (0.5)                    | .46     |
| **Process constructs**   | 0.3 (0.8)| 0.3 (0.8)    | .91     | 0.3 (0.8)| 0.3 (0.8)| .91     | 0.4 (0.8)                    | 0.2 (0.8)                    | .44     |
| **Inner setting constructs** | 0.3 (0.5)| 0.2 (0.6)    | .41     | 0.2 (0.6)| 0.2 (0.5)| .93     | 0.3 (0.5)                    | 0.2 (0.6)                    | .44     |
| **Implementation climate** | 0.4 (0.5)| 0.3 (0.7)    | .75     | 0.3 (0.7)| 0.4 (0.5)| .85     | 0.4 (0.6)                    | 0.3 (0.7)                    | .48     |
| **Readiness**            | 0.3 (0.8)| −0.0 (0.9)   | .23     | 0.2 (0.9)| 0.1 (0.8)| .87     | 0.3 (0.7)                    | 0.0 (0.9)                    | .35     |
| **Outer setting constructs** | 0.4 (0.5)| 0.1 (0.5)    | .1      | 0.3 (0.5)| 0.2 (0.5)| .39     | 0.2 (0.5)                    | 0.3 (0.5)                    | .59     |
| **Intervention Constructs** | −0.1 (0.5)| −0.1 (0.6)  | .97     | −0.2 (0.5)| −0.1 (0.6)| .41     | −0.1 (0.6)                   | −0.2 (0.5)                   | .59     |

*Note. CFIR = Consolidated Framework for Implementation Research; OTG = Opioid Therapy Guideline Adherence Metrics.*

Scales were created using the means of the constructs as shown.

**Process:** Three variables: Engaging, planning, and reflecting.

**Inner setting:** Nine variables: Structural, networks, implementation, tension, compatibility, relative P, leadership, available resources, and access to knowledge.

Two subscales:

**Climate:** Four variables: Implementation, tension, compatibility, and relative P.

**Readiness:** Three variables: Leadership, available resources, and access to knowledge.

**Outer setting:** Two variables: Patient needs and peer pressure.

**Intervention:** Two variables: Relative advantage and evidence strength.
Appendix 2

See Table A2.

Table A2. Association between the 16 Consolidated Framework for Implementation Research (CFIR) construct scores and three facility characteristics.

| CFIR construct rating (N = 39) | Medical center highest complexity Level 1 \(^a\) | Other levels | Top quartile of academic detailing provided \(^b\) | Other quartiles Mean (SD) | P-value | Rural \(^c\) | Urban \(^c\) | P-value |
|-------------------------------|---------------------------------|--------------|---------------------------------|-----------------------|--------|--------|--------|--------|
| Engaging                      | 0.6 (1.0)                       | 0.8 (0.9)    | 0.3 (1.1)                       | 0.1399                | 0.6 (0.9) | 0.8394 | 0.8 (0.4) | 0.6 (1.0) | .7719 |
| Planning                      | 0.5 (1.1)                       | 0.6 (1.0)    | 0.4 (1.3)                       | .7897                 | 0.2 (1.1) | 0.253  | 0.8 (0.4) | 0.5 (1.1) | .6314 |
| Reflecting evaluating         | −0.4 (1.0)                      | −0.2 (1.0)   | −0.6 (1.0)                      | −0.1971               | −0.3 (1.1) | −0.9731 | −0.3 (1.0) | −0.4 (1.0) | .8542 |
| Tension for change            | 0.7 (0.5)                       | 0.6 (0.6)    | 0.9 (0.4)                       | 0.2247                | 0.7 (0.5) | 0.7831 | 0.8 (0.4) | 0.7 (0.5) | .5905 |
| Implementation climate        | −0.1 (1.1)                      | 0.1 (1.1)    | −0.4 (1.0)                      | 0.2123                | −0.1 (0.9) | −0.8411 | 0.0 (0.6) | −0.1 (1.1) | .9195 |
| Compatibility                 | 0.0 (1.1)                       | 0.0 (1.0)    | 0.0 (1.2)                       | 0.8790                | −0.4 (1.1) | 0.1756 | 0.3 (0.5) | −0.1 (1.1) | .3214 |
| Structural characteristics    | −0.2 (0.9)                      | −0.0 (0.9)   | −0.5 (0.9)                      | −0.1531               | 0.0 (0.9) | −0.4578 | −0.7 (0.8) | −0.1 (0.9) | .1370 |
| Networks communications       | 0.5 (0.9)                       | 0.6 (0.8)    | 0.3 (1.1)                       | 0.3483                | 0.4 (0.8) | 0.8204 | 0.7 (0.8) | 0.4 (0.9) | .7352 |
| Leadership engagement         | 0.3 (1.1)                       | 0.3 (1.0)    | 0.2 (1.3)                       | 0.9270                | −0.1 (1.1) | 0.2679 | 0.7 (1.4) | 0.2 (1.0) | .3195 |
| Relative priority             | 0.7 (0.9)                       | 0.7 (0.9)    | 0.7 (0.9)                       | 1.0000                | 0.4 (1.1) | 0.2902 | 1.2 (0.8) | 0.6 (0.9) | .1837 |
| Available resources           | 0.1 (1.0)                       | 0.2 (1.0)    | −0.1 (1.1)                      | 0.3817                | 0.3 (0.7) | 0.2658 | 0.4 (1.3) | 0.0 (0.9) | .5603 |
| Access to knowledge information| 0.1 (0.9)                      | 0.1 (0.8)    | 0.1 (1.1)                       | 0.6619                | 0.0 (0.9) | 0.6680 | 0.0 (0.9) | 0.1 (0.9) | .7241 |
| Evidence                      | −0.4 (0.8)                      | −0.5 (0.7)   | −0.2 (1.0)                      | −0.2740               | −0.8 (0.8) | −0.0708 | −0.3 (0.8) | −0.4 (0.8) | .8122 |
| Relative advantage            | 0.2 (0.7)                       | 0.2 (0.7)    | 0.1 (0.9)                       | 0.6774                | 0.3 (0.8) | 0.4439 | 0.7 (0.8) | 0.1 (0.7) | .1344 |
| Patient needs resources       | 0.3 (0.7)                       | 0.2 (0.8)    | 0.5 (0.5)                       | 0.2667                | 0.4 (0.8) | 0.8875 | 0.5 (0.5) | 0.3 (0.7) | .5348 |
| Peer pressure                 | 0.1 (0.7)                       | 0.2 (0.7)    | 0.0 (0.8)                       | 0.4383                | 0.1 (0.7) | 0.7152 | −0.2 (0.8) | 0.2 (0.7) | .2931 |
| Mean of 16 STORM ratings      | 0.2 (0.5)                       | 0.2 (0.4)    | 0.1 (0.5)                       | 0.6600                | 0.1 (0.6) | 0.7473 | 0.3 (0.3) | 0.2 (0.5) | .5327 |
| Proportion of positive scores in | 0.4 (0.2)                      | 0.4 (0.2)    | 0.4 (0.2)                       | 0.9764                | 0.4 (0.3) | 1.0000 | 0.5 (0.1) | 0.4 (0.2) | .8444 |

Note. STORM = Stratification Tool for Opioid Risk Mitigation.

\(^a\)Facility complexity is a long-standing VA variable about the nature of the services provided at VA facilities. The score is classified numerically from 1 to 3, with level 1 being the most complex. We compared CFIR construct scores between facilities with complexity level 1 versus other levels.

\(^b\)We obtained measures of academic detailing provided at VA facilities and grouped them based on quartiles. We then compared ratings between facilities at highest versus other quartiles.

\(^c\)Rurality was classified as yes/no, using the VA definition of rurality, which defines urban as being in census tracts with at least 30% of the population living in an urbanized area.