

Official Title: Participatory System Dynamics vs Audit and Feedback: A Cluster Randomized Trial of Mechanisms of Implementation Change to Expand Reach of Evidence-based Addiction and Mental Health Care

NCT number: Not yet assigned

Document Date: 11/30/2019

Specific Aims

Providers, patients, policy makers and scientists each have a stake in ensuring all patients with opioid use disorder (OUD), alcohol use disorder (AUD), depression and PTSD, receive timely, evidence-based care. In Veterans Health Administration (VA) these common, costly conditions comprise the majority of outpatient addiction/mental health needs.¹⁻³ But, highly effective treatments, such as evidence-based psychotherapy (EBPsy) and evidence-based pharmacotherapy (EBPharm) reach only 3-28% of patients.⁴⁻⁶ National evidence-based practice (EBP) dissemination programs,⁷⁻¹¹ policies,¹²⁻¹⁴ incentivized quality measures,^{15,16} and EBP-focused electronic health records¹⁷ have been insufficient for greater EBP reach,¹⁸⁻²⁰ key to preventing chronic impairment, relapse, overdose,²¹⁻²³ and suicide²⁴⁻²⁶ among Veterans.

In our R21, we piloted participatory system dynamics (PSD) due to the ineffectiveness of status quo Audit and Feedback (AF) for achieving desired EBP reach. Like most health systems,²⁷⁻²⁹ VA enlists a multicomponent strategy of team huddles and AF to coordinate and improve EBPsy/EBPharm care. But, AF outcomes vary widely,³⁰ with the least success for complex tasks, such as multidisciplinary EBP continuity and coordination.³¹ In fact, AF effect sizes in research studies have not increased for 14 years.³² AF and PSD are rich comparators, due to similarities (data review) and differences (system simulations).^{33,34} PSD research shows that fixing complex, multidisciplinary, team EBP delivery requires addressing system causes.³⁵⁻⁴⁰

PSD theory of change is that participatory learning from modeling^{41,42} enables new capacities for managing causal system dynamics or mechanisms of change.^{43,44} PSD learning to develop systems thinking overcomes cognitive limitations^{45,46} to prospectively restructure delay and feedback dynamics that cause EBP reach over time.⁴⁷ We iterated with stakeholders to co-create 4 universal models of 'limited EBP reach' (Care Coordination, CC; Medication Management, MM/EBPharm; Psychotherapy, Psy/EBPsy; Aggregate, Agg), understood by everyone to explain reach in any team as a function of local resources.^{48,49} We extracted existing health system data using standard definitions for diagnoses, appointments, visits, and EBPs to calculate model parameter values for local teams.^{15,50} Models define mechanisms identified by stakeholders' using systems theory and calculus (structural validity), validated against historical data (structural-behavioral validity).⁵¹ PSD learning is via real-time visual emergence of local causes of EBP reach in virtual experiments.⁵²⁻⁵⁴ Simulations are a safe way to build systems thinking in to EBP decisions.^{44,55,56}

Using PSD and AF with outpatient mental health teams, we will achieve the following specific aims:

We propose a two-arm, 24-site (12 clinics/arm) cluster randomized trial to test for superior EBP reach for depression, PTSD, OUD, and AUD (4 EBPsy; 7 EBPharm, see *Plan*) in PSD clinics over AF (aim 1). We test whether PSD works via systems thinking as predicted by PSD theory of change (aim 2). We test whether EBP causal mechanisms in MM/EBPharm and Psy/EBPsy models generalize in AF and PSD clinics (aim 3). Eligible clinics will be from regional VA health care systems (analyses control for clinics nested within VA⁵⁷) below the overall VA national quality median for EBPsy/EBPharm. Computer-assisted stratified block randomization will balance arms (6 clinics/wave over four waves) at baseline using VA data.

Aim 1: Test superiority of PSD over AF for increasing EBP initiation and course. The proportion of patients (1a) initiating EBPsy/EBPharm, and (1b) completing an adequate EBPsy/EBPharm course, will significantly increase in PSD clinics as compared to AF clinics, in omnibus tests across EBPs for (1a) initiation, and (1b) course, using clinic 12 month pre/post period EBP reach averages.

Aim 2: Test PSD theory of change that increased EBP reach is via systems thinking.^{47,58,59} The effect of PSD/AF on 12 month period EBP reach will be explained by 6 month team systems thinking (STS), in ratio of mediator probability weight (RMPW) mediation analyses using the R 'MultisiteMediation' package.⁶⁰⁻⁶²

Aim 3: Test the generality of mechanisms of change in EBPsy/EBPharm PSD models. Structural-behavioral validation tests^{51,63,64} of causal dynamics formulated in our Psy/EBPsy and MM/EBPharm models will generalize to explain EBP reach as a function of local data across PSD and AF clinics (in *Plan*).

Exploratory aim. We contextualize aims 1-3 using *provider surveys* of the clinic-level learning organization survey (LOS-27),⁶⁵ team-level decision-making (TDMQ)⁶⁶ and burnout (PACT),⁶⁷⁻⁷⁰ and PSD/AF feasibility, acceptability and appropriateness (FIM, AIM, IAM).⁷¹ We enlist *qualitative/observational coding* of PSD/AF fidelity, systems thinking⁷² during team decision-making, and online PSD/AF use and sustainment.^{34,73}

Innovation – Our design tests a well-established systems science strategy (PSD) for its superiority, theory of change, and causal mechanisms, against a highly prevalent, but variably effective implementation strategy (AF) in the largest integrated health care system in the US.⁷⁴ Achievement of specific aims will advance scientific knowledge regarding the causes of limited EBPsy/EBPharm reach. PSD innovates in empowering frontline teams of providers to understand and prospectively manage local causal dynamics to better meet the high-priority needs of addiction and mental health patient populations.

RESEARCH PLAN

SIGNIFICANCE. Current strategies are insufficient for improving reach of evidence-based practices (EBPs). Ten years of VA nationwide programs to disseminate, train, track and incentivize evidence-based psychotherapy (EBPsy) for PTSD and depression patients, and evidence-based pharmacotherapy (EBPharm) for patients with depression, opioid use disorders (OUDs) and alcohol use disorders (AUDs) has been insufficient for reaching adequate proportions of these patient populations (see *Table 1*). Opiate and alcohol misuse, depression and PTSD, are the primary reasons Veterans seek outpatient VA addiction and mental health care.^{1,3} Yet, at the median (see *Table 1*), less than 60% of PTSD patients, and less than 40% of depression patients, start psychotherapy of *any kind*. Only 28% of depression patients starting EBPharm receive a therapeutic course, and 71% of OUD patients do not initiate EBPharm. Among depression and PTSD patients who start psychotherapy, only 30-44% are retained for at least 3 visits. Only 3-5% of depression and PTSD patients start EBPsy (see *Table 1*).

Table 1. VA SAIL Quality Measures & RO1 Study EBP Reach (Q1 2017)

Population Coverage - Denominator (<i>diagnostic cohorts</i>)	Study	EBP	SAIL NAME	VA 50th %ile
% OUD <i>diagnosed</i> patients receiving opioid agonist or antagonist	A1, 2, 3/recruit	EBPharm	SUD16	28.4
% AUD <i>diagnosed</i> patients receiving medication assisted therapy*	A1, 2, 3/recruit	EBPharm	ALC-TOP	9.8
% Depression <i>diagnosed</i> patients with depression psychotherapy visit	recruit		Psy32	38.3
% PTSD <i>diagnosed</i> patients with psychotherapy visit for PTSD	recruit		Psy38	55.8
Continuity of Care - Denominator (<i>diagnosis + active treatment</i>)				
% Patients on <i>new antidepressant medication</i> (84 days continuous)	A1, 2, 3/recruit	EBPharm	MDD43h	73.0
% Patients on <i>new antidepressant medication</i> (180 days continuous)	A1, 2, 3/recruit	EBPharm	MDD47h	57.2
% Depression <i>treatment</i> patients - 3 psychotherapy visits in 6 weeks	recruit		Psy33	30.4
% PTSD <i>treatment</i> patients - 3 psychotherapy visits in 6 weeks	recruit		Psy39	44.0
EBPsy Reach Targets – Denominator (<i>diagnosis/EBP template</i>)				Reach %
PTSD - EBP Template for PE or CPT Session 1 (Initiate)	A1, 2, 3	EBPsy	PTSD 56	5.3
PTSD - EBP Template for PE or CPT Completion (Course)	A1, 2, 3	EBPsy	PTSD 56	-
Depression - EBP Template for CBT-D, ACT-D, IPT-D Session 1 (Initiate)	A1, 2, 3	EBPsy	n/a	2.6
Depression - EBP Template for CBT-D, ACT-D, IPT-D Completion (Course)	A1, 2, 3	EBPsy	n/a	-

Source. SAIL, Corporate Data Warehouse (CDW). EBP = Evidence Based Practice Reach Outcome. A1 = Aim 1. EBPharm = Evidence-based Pharmacotherapy. EBPsy = Evidence-based Psychotherapy. *Psychotropic Drug Safety Initiative (PDSI) Medication Assisted Therapy for AUD = Acamprosate, Disulfiram, Naltrexone, Topiramate. PE = Prolonged Exposure. CPT = Cognitive Processing Therapy. CBT = Cognitive Behavioral Therapy. ACT = Acceptance and Commitment Therapy. IPT = Interpersonal Process Therapy.

Background of the ‘Limited EBP Reach’ Problem: Ineffective Audit and Feedback (AF) Strategies. VA/DOD clinical practice guidelines, national EBP training and consultation programs, EBP note templates in electronic health records, and VA quality measures known as Strategic Analytics for Improvement and Learning (SAIL),^{12,13,15,75,17} were all developed, due to the efficacy of EBPsy and EBPharm (see *Table 1*) for reducing PTSD^{76,77} and depression,^{78–82} alcohol^{83–86} and opioid misuse,^{87,88} and reducing risk of death by suicide or overdose.^{23,26} VA dissemination programs also demonstrate EBPsy/EBPharm effectiveness. Patients who received cognitive behavioral therapy (CPT) for depression experienced a 40% reduction in depression symptoms⁸ and over 60% of Veterans who received prolonged exposure (PE) experienced clinically significant improvements in PTSD.¹⁰ VA’s **multicomponent strategy** of 1) setting guidelines, 2) providing training/resources to meet guidelines, 3) tracking performance, and 4) providing feedback to identify gaps or progress, form the chief principles of AF benchmarking in any health system.^{15,16,29} But, AF systemic reviews show AF must also provide actionable insights to gain increases in EBP reach.^{27,31,32,89–91}

Reach. We define reach as the proportion of the outpatient population who initiate an EBP and complete a therapeutic course. Implementation gaps for depression, PTSD, AUD and OUD patients (*Table 1*), reflect the complexity of identifying optimal health system improvements. Too few VA patients receive EBPsy even in specialty programs.^{11,92} Improving reach involves *interdependent fixes* within and across programs (e.g., general vs. specialty), meeting a variety of common, and often comorbid patient needs (e.g., mental

health/addiction), by multidisciplinary teams with varying capacities/expertise (e.g., EBPharm prescribers vs. EBPsy providers). ***What generally causes limited EBP reach under these conditions?***

Implementation Science (IS) - Determining what works, why and under what conditions

Background of the IS Problem: Understanding causes of limited EBP reach is critical to our stakeholders and to our field. IS seeks to determine why and under what conditions a strategy increases EBP reach.^{93–95} Over 61 IS frameworks elaborate multiple domains,^{96–99} and doing so, underscore limited knowledge of their underlying dynamic, multi-causal premises.¹⁰⁰ The Participatory System Dynamics (PSD) theory of change is that participatory stakeholder co-learning from modeling enables new capacities for managing system dynamics or mechanisms of change.^{39,54} Informed by 60 years of PSD research, R01 specific aims define EBP reach as a system behavior, and test the PSD theory of change that grasp of system dynamics (mechanisms of change) is not just *for scientists*. Rather, systems thinking *by frontline staff*, is the basis for improving EBP reach. We briefly highlight limits of extant IS theories of change, and tests of mechanisms of change, focusing on what a PSD paradigm adds to IS, and its advantages over AF.

Theories of Change (TOC) are backward mapping methods for identifying the necessary determinants of change.¹⁰¹ TOC is used as a planning strategy to link process and outcome.¹⁰² These process frameworks are practical on face, but do not specify causal relationships that can be generalized.^{32,103,104} For example, comparative qualitative analysis (CQA) is an inductive method to develop a TOC.^{105,106} CQA retrospectively classifies sets of conditions co-observed with desired outcomes to descriptively infer causality, often identifying multiple pathways (permutations) to the same outcome (equifinality, see below).¹⁰⁵ Determinants are not scientific theories¹⁰³ and pose risks as causal paradigms; Lack of tests for temporal precedence or effect size,¹⁰⁶ propensity for false positives¹⁰⁷ and confirmation bias,¹⁰⁸ all indicate caution. IS frameworks presume several components, but multiple comparisons inflate type I error,¹⁰⁹ impeding theory tests, just as finding multiple pathways leaves stakeholders weak prospective guidance for making EBP-related change.

Mechanisms of Change (MOC) are targets that will explain findings and accelerate research progress. Many implementation strategies produce no effect.¹¹⁰ Review of multilevel mechanisms of implementation strategies in controlled trials, found nine tests of mediation, *all unsupported*.¹¹¹ Again, there are challenges; Confirmatory hypotheses require a strong theoretical and empirical base, as inclusion of multicollinear variables, and/or use of multiple tests both inflate error. Without a resource-intensive experimental design, use of mediating *associations* to infer causality is debated; ‘Mediators differ from *mechanisms* which invoke a higher level of specificity and describe the precise sequence of operations or underlying causal processes through which an effect occurs’ (Williams, 2016; p. 784). Finally, multilevel mediation is in the expert domain and inaccessible to most stakeholders, with limited utility for guiding local change. ***Is there an alternative?***

Preliminary Research – Two Causal PSD Theories for Improving EBP Reach – AF versus PSD (Aim 1)

Based on our R21 pilot, PSD is well suited to the need for improving EBP reach *and* need for IS progress in defining and testing TOC and MOC. The PSD paradigm sees the status quo ‘limited EBP reach’ as a function of the mental models that teams use to guide decisions, which are inadequate for redesigning EBP-related system dynamics.^{36–38,47} PSD shows how ‘today’s problems come from yesterday’s solutions.’¹¹² PSD simulation learning to make system causes transparent and analyzable, increases systems thinking enabling more effective ongoing change. PSD enlists two ‘classic’ causal theories: decision theory (Aim 2 - TOC) and systems theory (Aim 3 - MOC). Distinct from IS determinant frameworks and process models, ‘Classic theories originate from fields external to IS, e.g., psychology...organizational theory, which can be applied to provide understanding and/or explanation of aspects of implementation’ (Nilsen, 2015; p.8).^{32,103,104}

AF is central to VA EBP implementation^{15,50} and is one of the most common IS strategies used around the world.^{29,30} AF and PSD each require data review.^{34,89} But, PSD research challenges the sufficiency of data for selecting effective changes.^{45,113} AF reviews have found variability in effectiveness could be traced to the complexity of the EBP (with AF more effective for simple practices⁹¹) and the feedback audience (with AF for team practices, more effectively delivered to teams^{114,115}). AF must be frequent, in writing, and include a correct solution.^{27,29} PSD agrees with need for timely, actionable feedback.^{116,117} But, without systems thinking and modeling, PSD doubts the ability of policy makers to provide correct solutions for the myriad local

‘Despite being extensively studied, health care A&F interventions remain variably effective, with overall effect sizes that have not improved since 2003.’...‘The development of the scientific basis of A&F in healthcare appears to have stagnated; we are not developing more effective A&F interventions than we were 20 years ago.’

– Colquhoun et al., 2017

issues that trouble EBP delivery.^{118–120} PSD cautions, ‘The cure can be worse than the disease.’^{121,122} **How might AF be counterproductive?**

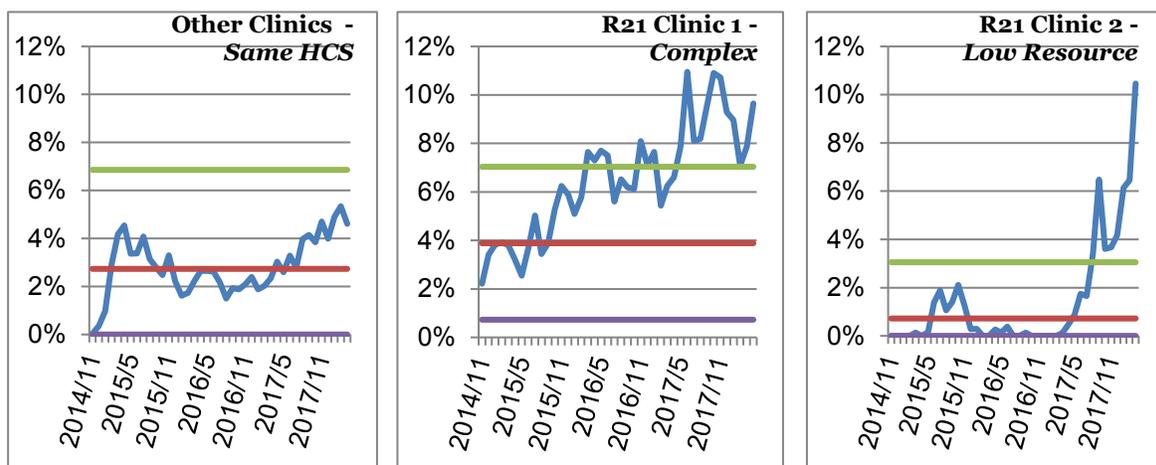
PSD uses the term policy resistance¹²² to describe the need for holism when making changes, represented in aphorisms such as ‘The harder you push, the harder the system pushes back,’^{112,123} and ‘There are no side effects in systems. There are only effects.’^{139,40,124} System resistance does *not refer to stakeholders’ attitudes; it refers to powerful dynamics of the system.*

Preliminary Pilot. For example, prior to our R21 pilot study, our pilot partners invited us to try PSD after a previous SMART goal effort failed. The clinic set the following Specific, Measurable, Timebound goal: ‘By April 2015, 40% of patients newly seen in outpatient mental health for depression, PTSD, or anxiety disorders will have two psychotherapy visits completed within 28 days from time of intake.’ Staff saw initial improvement in scheduling (up from 25% to 65% scheduled), but wide unexplained scheduling variability was observed over 9 months (some weeks 0% scheduled, some weeks 100%), and the gap between scheduling and completing 2 visits showed the 40% goal was never achieved (new mean = 14%). In SMART terms, PSD proposes that if the goal was never Achievable, because it was not Realistic in that Timeframe with the available resources, then it may do more harm than good by 1) generating further instability of the EBP-related system behaviors (see Aim 3 Background), and 2) undermining psychological safety, or willingness to learn and try out new solutions,^{44,56,65} exacerbating staff burnout and risk for turnover (see Aim 2).^{55,68,118}

Preliminary R21 Data in Support of Aim 1 – PSD Effectiveness

Our statistical process control analyses indicate our two R21 pilot clinics each demonstrated a three standard-deviation increase above their pre-intervention EBP reach using PSD (*Figure 1*; $\alpha < .003$). In *Figure 1*, purple = lower control limit; red = clinic 12-month pre-intervention EBP reach; green = upper control limit. Control limits are three standard deviations above and below the pre-intervention mean. These clinics have maintained this improvement for 12 months and 8 months respectively, whereas the other seven VA clinics from the same regional health care system (HCS) did not improve EBP reach over this period. Moreover, we observed no secular trend toward improved EBP reach in VA national AF measures over this period. These R21 quasi-experimental findings support the proposed R01 effectiveness test: a cluster randomized trial (CRT) to evaluate for PSD superiority over standard AF.

Figure 1. R21 Effectiveness – Statistical Process Control of EBP Reach Improvement (+3 SD)



Summary (Aim 1). AF and PSD research show data review is insufficient for effective change.^{43,125} AF effectiveness is diminished³² due to choosing ineffective strategies that leave system causes unaddressed.

Partnering to Define EBP Reach using Systems Thinking – PSD Theory of Change (TOC) – Aim 2

Why else doesn't AF work better? PSD explains that AF learning is also attenuated by delays between making changes and observing their real-world effects.⁴⁷ A recent scoping review of ‘Healthcare Learning Organizations’ found the majority principles outlined by System Dynamics (the field coined ‘Learning Organization’),^{121,123} including the central idea, in which systems thinking coheres mental models, shared vision, and team learning.¹²⁶ But, narrowed down from 263 keyword- identified articles, *only 2 published studies measured systems thinking*, the cornerstone of PSD research on change.

The PSD TOC defines learning as a system feedback process in which mental models are formed from feedback in the real or virtual world, which shape the rules used for decision-making, which then shapes the real world. In other words, 'Seeing is believing and believing is seeing,'¹³⁹ especially as we act to change the real world. This learning process is more effective with PSD modeling, called 'double loop learning.'^{124,127,128} From the first writings of the discipline,¹²⁹ PSD practices are guided by decision science.^{46,130}

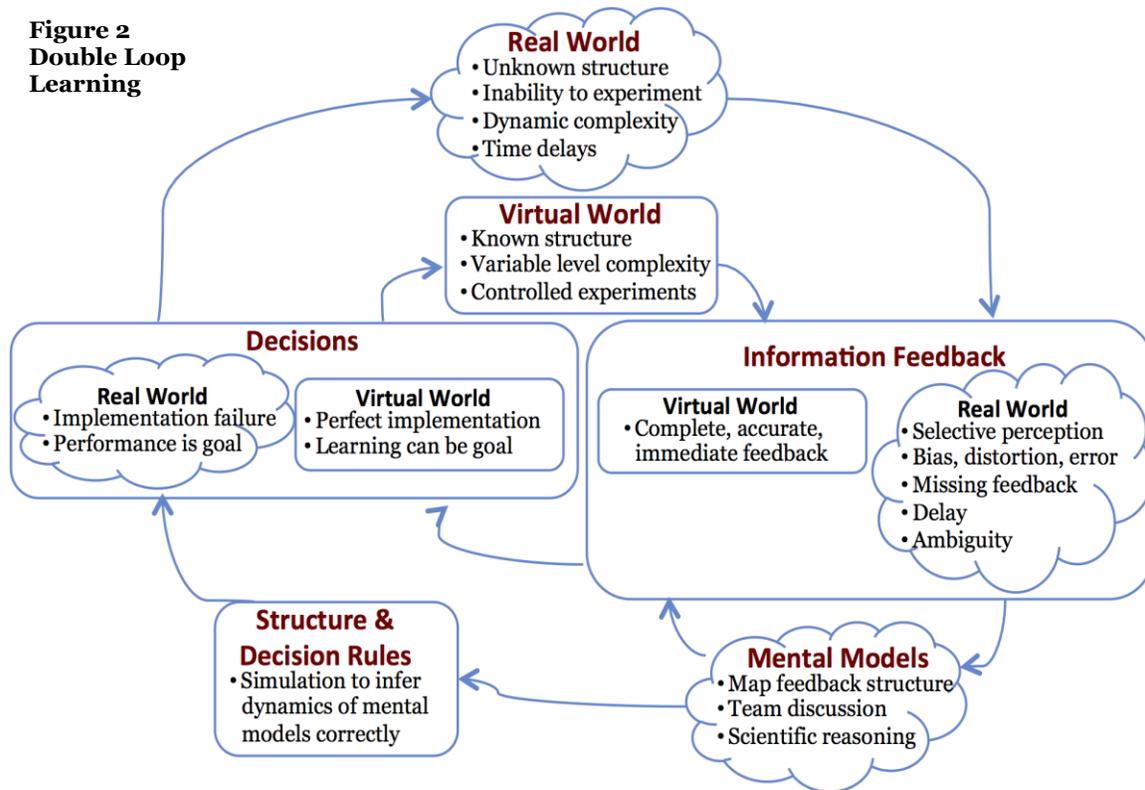
PSD research identifies several cognitive biases and limitations that lead to poor decisions when facing complexity: use of heuristic mental models that seek minimally satisfying solutions rather than optimal solutions (bounded rationality),¹³⁰ rules of thumb that wrongly attribute the state of a system solely to inflows rather than outflows (correlational heuristic),¹²⁴ and inability to solve accumulation or delay problems (stock-flow failure).^{113,124,131} A PSD insight is that in a complex system, cause and effect may not be closely related in time and space,¹¹² rendering learning from AF alone unlikely.

Systems Thinking as a Mediator of Change. Figure 2 depicts the difficulty of learning from the complex, real world, as compared to the virtual world of modeling. PSD recommends simulation to improve mental models with systems thinking. Without PSD, defective causal decision rules impact explicit, effortful implementation planning (system 2 cognition) and implicit, automatic day-to-day decisions about EBP coordination and continuity (system 1 cognitions).¹³²⁻¹³⁴ An effectiveness review of 107 PSD projects, identified more efficient improvements (33%), increased consensus (49%) and commitment to change (33%), including systems change guided and evaluated by modeling (42%).^{43,125} A recent mediation study showed *increases in systems thinking due to PSD*, that led to increased psychological safety, which increased information sharing, explaining performance improvements.⁴⁴

'Deep change in mental models, or *double loop learning*, arises when evidence not only alters our decisions within the context of existing frames, but also feeds back to alter our mental models. As our mental models change, we change the structure of our systems, creating different decision rules and new strategies. The same information, interpreted by a different model, now yields a different decision. Systems thinking is an iterative learning process in which we replace a reductionist, narrow, short-run, static view of the world with a holistic, broad, long-term, dynamic view, reinventing our policies and institutions accordingly.'

-Sterman, 2006 p. 509

Figure 2
Double Loop Learning



Preliminary PSD Mixed-Methods Facilitation Data in Support of Aim 2 – Systems Thinking Mediator
Systems Thinking during PSD model development. We used PSD best practices, including nominal group technique to generate 12 categories of 131 'limited EBP reach' related issues.^{148,149} Over a period of six months of modeling we used several rounds of 'dot voting' to converge on PSD priorities: 'issues' were

translated into ‘variables’ and ‘decisions’ in models.^{136,150} Our purpose was to build models that explain ‘limited EBP reach’ using providers’ new systems thinking skills. Our prompt question was: ‘What simulation will help teams within the VA learn to manage the tradeoffs in how to provide evidence-based care to Veterans?’ Four Focus areas, ‘Care Coordination,’ ‘Management Concerns,’ ‘Provider Capacities and Constraints,’ and ‘Provider Quality of Work Life’ were narrowed to 9 specific priorities (Table 2).

Summary (Aim 2). *Systems thinking* is the ability to recognize, understand, and synthesize interactions and interdependencies, including how actions and components can reinforce or counteract each other.¹³⁵ Simulation learning upgrades mental models guiding daily system 1 (fast) and strategic system 2 (slow) EBP decisions in multidisciplinary teams, overcoming AF limits for learning how to improve local EBP reach^{58,130}

Table 2. Staff Priority Experiments

- Patients' needs/preferences
- Reduce extra stops for Veterans
- Initiating a specific treatment
- Allocations of time (not enough time)
- Actual time (what we really do)
- Misunderstanding provider functions
- Morale & burnout
- Staff turnover

Partnering to Define EBP Reach as a System Behavior – PSD Mechanisms of Change (MOC) – Aim 3

The ability to infer general principles from observations (generality) is foundational to scientific development. Use of PSD in Mayo Clinic identified substantial savings in the treatment of renal disease by recognizing that oscillating hemoglobin measures were caused by a mismatch between the measurement and its use in guiding clinician medication decisions.^{151,152} Although hired to develop an AF system, use of PSD with a multidisciplinary team recognized the problem was across hematology and nephrology. Based on the measures available, clinicians were administering a second dose before the first took effect, causing clinically acute adverse effects. *Identification of this underlying biophysical dynamic was generalizable and led to Mayo Clinic-wide implementation of PSD to guide individualized dosing,* which brought population-level hemoglobin within range, increased well-being of patients, and reduced costs.^{151,152}

Aim 3 will test whether R21 models have potential to replicate this type of advance in addiction and mental health. We hypothesize the structural-behavioral validity of our R21 models will generalize across AF/PSD arms.

In common with Community Based System Dynamics¹³⁶ we locate PSD within the continuum of participatory research.^{48,137} PSD is a partnership approach that equitably involves all stakeholders’ expertise, in all aspects of the research development process, using shared decision-making activities that are designed to produce system change.^{137–140} We committed to equitable PSD resource development, valuing local staff knowledge in PSD models and activities. Unlike most AF systems, we co-created our new shared PSD assets with frontline staff.⁴⁸ **PSD activities elicited stakeholders’ mental models about how they think EBP implementation works, with those interconnections made explicit in PSD models.**

A system is set of elements interconnected in such a way that they produce their own internal dynamics.¹²² The dynamics of a system problem cause its behavior. Many IS frameworks, and stakeholders, view a system as an external setting or organizational context (exogenous).^{98,141–145} PSD does not. EBP reach behavior emerges from internal causes.^{122,124} Using AF, these system dynamics are hidden ‘black boxes.’ With simulation, PSD makes causal dynamics transparent in real-time. The PSD endogenous theory of ‘limited EBP reach’ is qualitatively refined with stakeholders and rigorously assessed for structural-behavioral validity.^{49,51,146,147} The PSD endogenous view is empowering, proposing that local teams engage in mutual learning to co-create solutions that change the dynamics of EBP reach.

PSD Model Structural-Behavioral Validity is present when the model represents its purpose defining ‘limited EBP reach,’ and the accuracy of its formulation is rigorously confirmed using reliable data.^{49,51,124,147,153} All of these conditions were present in our R21 model validation using VA data.^{15,48,50} Tests of structural-behavioral validity show how PSD reconciles the equifinal/multifinal IS paradox. The equifinal and multifinal columns in Table 3 display explanatory causal operators (structure of the equation), accounting for specific numerical values to derive the result.

Table 3. Causality

Equifinal	Multifinal
1 + 1 = 2	2 x 3 = 6
1 - 1 = 0	3 x 4 = 12
1 x 1 = 1	5 x 1 = 5
1 / 1 = 1	0 x 6 = 0

In the equifinal column of Table 3, mathematical operators demonstrate *different* causal relationships sometimes achieve the *same* result (2 out of 4 times), and sometimes don’t, even with the *same* numerical values. The multifinal column demonstrates the *same* cause achieving *different* results as a function of *different* numerical values. These mathematical facts demonstrate the key import of understanding causes.

PSD formulates generalizable feedback and stocks-and-flow equations that produce EBP reach according to system theory, when parameterized with local data to guide change. The fundamental

insight of PSD, ‘Dynamics (equations) before details (specific parameter values),’ is drawn from the fundamental theorem of calculus.

Preliminary PSD Structural-Behavioral Validity Data in Support of Aim 3.

Causal Stock and Flow Dynamics. We found that although stakeholders may disagree about details, rarely did they disagree about dynamics. For example, several insights are visual and clear in Figure 3, which shows how changing the ‘Medication Management (MM) Appointment Frequency’ and ‘MM Return Visit Interval’ are related, and influence the interdependent ‘New and Existing Patient Services,’ and how the units of appointments (‘MM Appts scheduled’), ‘MM New Patient’ start rate and ‘MM Patients waiting for 1st visit’ are all related. Before calibration, these interdependencies were refined qualitatively to ‘saturation.’ After several iterations with multiple teams and stakeholders no new key dynamics were identified. These issues are generic across a variety of clinical teams.

Improving team understanding of these dynamics was deemed critical by staff, particularly with regard to OUD EBPharm, which requires a very strict return, visit interval for an adequate dose among existing patients. This tradeoff with ‘Decision to start Patient in MM’ is typically managed with mental heuristics, which staff agreed would be better optimized with PSD modeling. NOTE: Figure 3 is static, but PSD simulations are temporally dynamic and show the impact of decisions in 1) real-time, on 2) EBP reach over a variety of future time horizons. We noted intent to submit a video for the R01 to show how these causal dynamics are displayed in the ‘Modeling to Learn’ Team Training (described below).

Calibration and Validation of PSD Models. PSD models are calibrated against existing EBP reach time-series data for each decision-making team, using Kalman Filtering and Monte Carlo methods for optimization.⁴⁹ Sensitivity analyses test the structural-behavioral validity of the model, which also was formulated to be consistent with seminal service system PSD models.¹⁵⁴

Figure 3 and Table 4 and Table 5 (on the next page) show PSD model dynamics and the associated parameter inputs are expected to be generic across a variety of services, patient populations and EBPs. But, parameter values are team-specific to guide local decision. Like the Mayo Clinic example, this enables participatory ‘modeling to learn’ activities at scale. Table 5 displays example model inputs, feedback dynamics represented as provider decisions, and model equations. In addition, based on feasibility and acceptability of PSD in the R21, NCPTSD funded the development of ‘Modeling to Learn.’ This workshop and user-interface enables this specification and sophistication under the hood, and real-time causal learning simulations to frontline EBP decision-makers. NOTE: Demonstration of real-time dynamic interface used in team training is available in the 2-minute ‘Modeling to Learn’ simulation video.

Summary (Aim 3). Generalizability PSD effectiveness of improving reach in a variety of settings is enhanced by the use of standard data definitions for model inputs and the generality of using mathematical principles to identify the common dynamics of EBP services systems. Based on the multiple iterations of

Table 4. VA Corporate Data Warehouse (CDW) Source for Model Inputs

Patient Cohorts	ICD diagnostic information from encounters with providers
Clinic Hours	Clinic scheduling hours & availability from VA scheduling system
Provider Specialties	Provider disciplines with user input to map to encounter types
Clinic Appointments	Completed CPT encounters. Appointment data = ‘no shows’
EBP reach	EBPsy Template usage or VA SAIL EBPharm data

ICD = International Classification of Disorders. CPT = common procedural terminology. EBPsy = Evidence-based psychotherapy. EBPharm = Evidence-based pharmacotherapy. SAIL = Strategic Analytics for Improvement Learning (quality measures, see also Table 1).

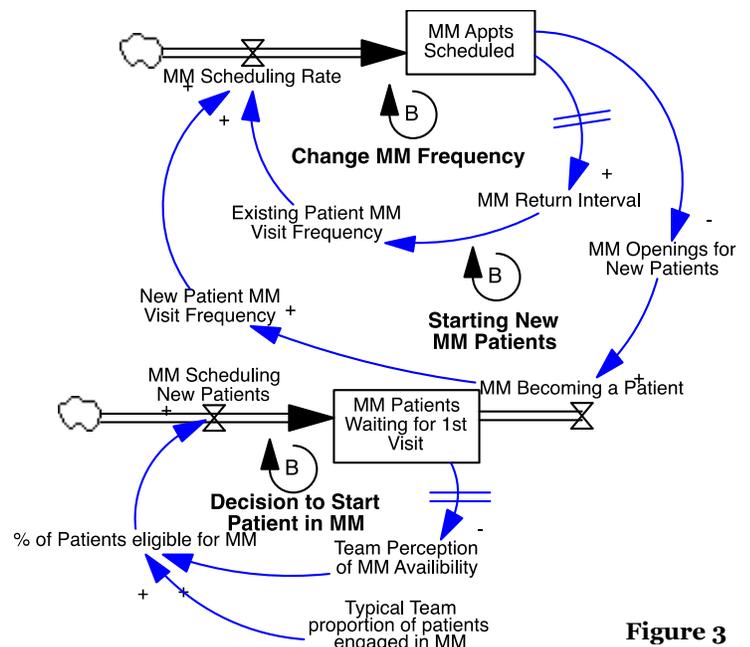


Figure 3

stakeholder engagement and structural-behavioral validity testing, we expect that our four PSD models will explain reach in any R01 study clinic.

Summary of Preliminary Data in Support of R01 Specific Aims

We propose a Phase III Clinical Superiority design with falsifiable hypothesis tests (see Table 6). We propose the two-arm CRT to test theoretical, confirmatory effectiveness (Aim 1), causality (Aim 2) and generality/generalizability (Aim 3) hypotheses. The proposed R01 is responsive to calls to develop AF strategies that are theoretically rigorous due to ineffectiveness of AF to date. We test AF (described below), against very rigorous theory-based approach (PSD) with a 60-year track record of effectiveness for improving organizations. Building from the PSD research tradition, proposal to conduct robust tests of internal validity and external validity is relatively rare in the IS field, which is often exploratory, especially regarding causality. We focus on EBP reach, due to solid EBPsy/EBPharm Effectiveness, and hefty VA Adoption, Implementation and Maintenance efforts (Re-AIM).⁹³ We locate R21 insights in implementation science (IS), outlining our rationale for expected PSD superiority (Aim 1), based on PSD theory of change (Aim 2), and PSD mechanisms of change (Aim 3).

Table 6. Summary of Proposed R01 AF/PSD CRT Design, Hypotheses, Tests & Measures for Specific Aims

Aim	Purpose	Hypothesis	Test	Measure
1	Effectiveness	PSD will be superior to AF for improving EBP reach	Cluster Randomization	Existing Patient Data; 6 & 12 month EBP initiation and course
2	Causality	Effect of PSD/AF on EBP Reach explained by Systems Thinking	Multilevel Mediation	6 mo. Team STS Mean (Mediator); Existing Patient Data (Outcome)
3	Generality	SD models will explain EBP reach across clinics in PSD AF arms	Structural-Behavioral	Match of CC, MM, Psy, Agg Model Structures to Observed EBP reach

RESEARCH DESIGN AND METHODS

Overview. Five standards guided study design and will be followed for reporting (Table 7).

Protocols for AF and PSD Implementation:

Protocols were developed to be implemented and evaluated for fidelity (see Table 9) using the Guidelines for Reporting Evidence-based Practice Educational interventions and Teaching (GREET). Led by the R21/R01 team (Table 8), National Center for PTSD (NCPTSD), Office of Mental Health and Suicide Prevention (OMHSP), Office of Strategic Integration (OSI) and Veterans Engineering Resource Center (VERC) developed a website for reviewing team data (AF/PSD), and a second website for team simulation (PSD). AF *‘Team Feedback’* clinics will be trained to use the data tools in 2 team huddles (month 1), with weekly emailed data reports linking to the data (months 2-6). PSD *‘Modeling to Learn’* clinics will use the data website, simulation website, and sessions with Care Coordination (CC), Medication Management (MM), Psychotherapy (Psy), and Aggregation (Agg) models, ending with facilitated wrap-up reflection (months 2-6).

AF Dynamic Data Tools. The AF condition is comprised of tools for reviewing data before synthesis in PSD simulation models is graphically displayed as retrospective trends overtime for the team. Trends graphs

Table 5. Model Inputs, Feedback/Decisions & Equations (example) Inputs (care coordination model)

New Patients Per Week	Mean Usage (appointment/week)
One and Done Patients (%)	Median Usage (appointment/week)
Missed Opportunities Rate (no show)	Seventy Five Percent Usage (appt/wk)
Mean Return Visit Interval (weeks)	Mean CC Engagement (duration)
Median Return Visit Interval (weeks)	Median CC Engagement (duration)
Mean Wait to First Visit (weeks)	
Median Wait to First Visit (weeks)	

Feedback & Decisions (medication management model)

Focus on Actively Engaged MM Patients, Work more MM hours, Change MM Frequency, Starting New MM Patients, Decision to Start Patient in MM

Model Equations (aggregate model)

Ideal Patient Start Rate for Available Hours[service]=(Ideal Completion Rate for Available Hours[service]/Demand per Patient per Week[service])/Actual Time in Service[service]; **Units** = patient/Week

Maximum Start Rate[service]=Patients Waiting to Start Service[service]/Minimum Time to Schedule[service]; **Units** = patient/Week

Ideal Completion Rate for Available Hours[service]=Baseline Service Capacity[service]/Time per Appointment[service]; **Units** = appt/Week

Table 7. Design and Reporting Standards

Cluster Randomized Trial Design	CONSORT*
PSD Simulation Model Documentation	SDM-DOC
PSD Simulation Model Output	SIMULATE
PSD/AF EBP Education Intervention	GREET
MH/Addiction Common Data Elements	PHENX

* 2010 update for 'parallel group randomized' trials. MH = Mental Health.

also produce searchable reports. The team can review their typical screening practices to note where gaps in quality exist. They can also review the underlying report to follow-up with specific patients who require care coordination and follow-up. Teams value graphical review of EBP template data.

PSD Simulation Tools. Data review improves provider confidence in the data that feeds the simulation models. Augmenting AF data review, PSD model simulations are designed to more precisely guide decisions through exploration of the expected causes of current EBP reach and simulations of the impacts of changes in future-oriented projections or scenarios.

Team Training Sessions. Virtually facilitated AF/PSD workshop sessions (see *Table 9*) are designed for teams of providers (typically 5-10/team) from each clinic, with no cap on total providers (typically ~20-40 staff). Frontline leadership and one ‘champion’ from each service delivery team will receive additional PSD resources to operate as an internal facilitator for their team. Mental health staff will be eligible to receive continuing education credits for AF (2 credit hours) and PSD (12 credit hours) workshops from VA Employee Education Service (EES) for six primary frontline disciplines: psychiatry, psychology, social work, counseling, nursing and certified peer-support.

Table 9. ‘Team Feedback’ (AF) and ‘Modeling to Learn’ (PSD) Protocol Implementation and Fidelity

Arm	Resources	Session	Use	Objective
AF & PSD	*EES Virtual Workshop	Intro (1a)	# Learners	Increase knowledge of current EBP reach
	Team Data Sharepoint	Data (1b)	# Unique Visits	
AF	Weekly Email w/Data Report	Self-directed (2-6)	# Clicks Min. on Site	Identify EBP reach gaps for improvement
PSD	Team Data Sharepoint	CC (2)	# Clicks	Apply systems thinking & simulation
	Simulation Interface	MM (3)	Min. on Site(s)	to see causes of EBP reach & run tests to find improvements
	Weekly Email Data & Sim. Report	Psy (4)	# Simulations	
		Agg (5)	# Sims Shared	
		Wrap (6)	Sim. Inputs & Outputs	

*EES – Employee Education System accreditation for physicians, psychologists, social workers, nurses, counselors and certified peer specialists (AF = 2 credit hours; PSD = 12 credit hours;). Min. = Minutes. Sim = Simulation. Models: CC = Care Coordination. MM = Medication Management. Psy = Psychotherapy. Agg = Aggregate.

Project Management Plan

Study Team and Stakeholders.

See *Table 8* in **Section 3.5 – Structure of the Study Team – Human Subjects and Clinical Trials.**

Study Timeline and Feasibility

See *Table 10* in **Section 2.7 – Structure of the Study Team – Human Subjects and Clinical Trials.**

We propose 60 total months of study activities and 30 months of active PSD facilitation or post-training technical assistance (phases 3 and 4). This leaves 24 months of flexibility for delays across pre (phases 1 and 2) and post (phases 5 thru 8) activities (*Table 10*).

Outcomes and Analysis Plan – Definitions and VA Corporate Data Warehouse (CDW) Data Use

Evidence-based Practices (EBPs). We selected EBPs for highly prevalent AUD, OUD, PTSD and depression⁴ based on demonstrated clinical efficacy and effectiveness^{76–82,87,88} and limited EBP reach (*Table 1*).

EBP Reach definition. We propose to use PSD to improve reach of 7 EBPs in the outpatient system. We define reach as the proportion of patients diagnosed with OUD, AUD, PTSD, or depression (ICD-10 codes) who meet EBPsy and EBPharm *1a) initiation* and *1b) course* measures (numerator) divided by the total number of patients with these diagnoses (denominator) at that clinic. Initiation of an EBP is indicated by EBPsy template or EBPharm prescription after intake. Adequate course is based on receiving an adequate number of EBPsy sessions to be a “completer” (typically 8 sessions) or enough refills for a guideline-recommended adequate trial of each medication (varies by medication) (see CDW/SAIL detail in *Table 1*).

EBPsy Definition. We will study five **EBPsy**. Three for depression (Cognitive Behavior Therapy [CBT-D], Acceptance and Commitment Therapy [ACT], Interpersonal Psychotherapy [IPT]), and two for PTSD (Prolonged Exposure [PE], Cognitive Processing Therapy [CPT]).

EBPsy Measure. Our EBPsy measure is completion of EBP templates during sessions with a relevant CPT code.¹⁵⁶ EBP templates may not reflect overall EBP delivery without templates. However, the templates confer several significant advantages, including 1) national standardization, 2) maximizing use of existing CDW data for AF, 3) integration with SAIL quality measures (PTSD 56) and EBP dissemination programs.

EBPharm Definition. Our EBPharm measures follows VA SAIL definitions (SUD 16, MDD43h, MDD47h): a combination of prescriptions placed with the VA pharmacy and sessions with a relevant CPT code.

EBPharm Measure. We will also improve reach of eight EBPharm measures - two for depression (84 and 180 days therapeutic continuity at new antidepressant start), two for OUD (methadone and buprenorphine)

and four for AUD (Acamprosate, Disulfiram, Naltrexone, and Topiramate).

Team Survey Measures

Participatory Measure Selection and Evaluation Process. AF and PSD theories of change and prior empirical research led us to target measures of: 1) systems thinking, 2) learning, 3) team decision-making, 4) psychological safety, 5) burnout/morale, and 6) pragmatic measures of PSD/AF implementation. We reviewed Society for Implementation Research Collaboration and Agency for Healthcare Research and Quality Measure Repositories, measures in *Implementation Science* and *System Dynamics Review*, and key IS review manuscripts,¹⁵⁷⁻¹⁵⁹ noting measures validated with the VA staff. After identifying the most valid and reliable measures (see *Table 11*), our team of patients, providers, manager and program leads further reached consensus about the measures based on face and content validity, or factor structure. Higher weight was given to brief measures and measures preferred by providers, reaching a final survey items to be completed by staff pre/post AF/PSD.

Measurement Schedule. *Table 11* shows the plan to collect 3 measures at baseline (BL) and 6 months (52 items), our mediator (STS) and measures to describe team (PACT) and clinic (LOS-27) contexts. Two measures of AF/PSD practicality (AIM, IAM, FIM), and AF/PSD effect on team decision-making (TDMQ) are added at 6 months (83 items).

Systems Thinking Scale (STS Mediator).

We selected the STS as mediator measure due to demonstration of criterion-validity. In addition to the measure details reported in the STS manual, the PI spoke to the authors to confirm appropriateness for use testing PSD theory of change in this R01 (Aim 2). The

STS was developed as part of the Advance the Science of Quality Improvement Research and Evaluation (ASQUIRE) initiative and validated among health care staff for purposes of quality improvement, including evidence the STS was manipulated by systems thinking education (sensitivity to change)^{160,161} and STS scores discriminated 3-groups based on their level of systems thinking education (Cohen's effect size $d = .78$).¹³⁵ The authors used the Quality Improvement Knowledge Application Tool (QIKAT),¹⁶² as a measure of convergent validity. But, the relatively low STS/QIKAT correlation (.28-.46) provides discriminant validity for our purposes. Based on this finding, the STS should distinguish general quality improvement knowledge from AF) from systems thinking (from PSD). After dimension reduction, the 20-item single-factor STS has strong internal consistency (.89) and test-retest reliability (.74).

Team Decision Making Questionnaire. The TDMQ asks, 'To what extent does AF/PSD help you to...' This four factor scale was validated to assess the impact of a team intervention on team decision-making, support, learning and development of quality services.⁶⁶

Exploratory Aim. Context Measures for Description of Study Clinics and Posthoc Analyses

Patient Aligned Care Team - (PACT). From VA team-based primary care, this 4-item descriptive measure tracks 1) years of experience with the team, 2) working on more than one team, 3) turnover/change in team staff, 4) team overwork, and the single-item 5) self-reported burnout (sensitivity 83.2 % and specificity 87.4 %)^{67-70,118}

Learning Organization Survey (LOS-27). Developed out of the learning organization tradition,⁶⁵ the 27-item Learning Organization Survey demonstrated good psychometric properties during VA validation,⁶⁵ and we include it to assess 7 clinic context factors: a) supportive learning environment (including psychological safety), b) leadership that reinforces learning, c) experimentation, d) training, e) knowledge acquisition, f) time for reflection, and g) performance monitoring. Response scales range from 0 (never) to 4 (always) for subscales a, and c-g. Subscale b ranges from 0 (highly accurate) to 7 (highly inaccurate). Items will be reverse scored as necessary and summed for analyses (in *Appendix*).

Measures of Feasibility (FIM), Appropriateness (IAM) and Acceptability (AIM) will assess for differences in team perceptions of PSD and AF on these three factors (12-items). Despite similarities, AF and PSD may be perceived differently in pragmatic terms. These scales have strong psychometric properties (see

Table 11. Team Measures

Scale	Aim	Time	Items	Valid	Reliable	Factors
1. Systems Thinking Scale (STS)	Aim 2	BL & 6 mo.	20	0.89	0.74	1
2. Patient Aligned Care Team (PACT) Survey-Burnout	Desc.	BL & 6 mo.	5	0.89	n/a	n/a
3. Learning Organization Survey (LOS-27)	Desc.	BL & 6 mo.	27	0.75 to 0.93	0.67 to 0.92	7
4. Team Decision Making Questionnaire (TDMQ)	Desc.	6 mo.	19	0.96	0.52 to 0.94	4
5. Pragmatic Measures (AIM/IAM/FIM)	Desc.	6 mo.	12	0.85 to 0.91	0.73 to 0.88	3

Table 11) for implementation research. Use in the R01 contributes to replication of reliability and dimensionality, and public release of data from the proposed sample of 720 providers adds to future measure norming.

Systems Thinking Mediator: Multilevel Psychometric Analytic Plan

Baseline Measures and Descriptive Analyses. STS means, standard deviations, and correlations will be calculated for individuals and teams. Intraclass correlation coefficients (ICCs) will be calculated. ICC values represent team 'traits,' estimated as the proportion of total measure variance attributed to variance between teams. An ICC near zero indicates highly variable STS within teams. We will do the same calculation for clinics within regional VA health systems.

Multilevel Confirmatory Factor Analysis (MCFA). Factor analyses will be confirmatory. Psychometrics and factor structure for the STS was determined in prior research (Table 11). Our unit of interest is teams. Teams are the unit for EBP implementation and PSD/AF training. As a result, we will use the STS team-mean. MCFA will be used due to nested measurement (teams in clinics), and potential for differential measure performance across arms in Aim 2 analyses due to PSD/AF assignment. MCFA with R package 'lavaan,'¹⁶³ will determine whether the same covariance (factor) structure holds for clinics in each arm (PSD/AF) via closeness of fit (CFI), root mean squared error (RMSEA) and information criteria (BIC/AIC). MCFA reduces risk for inflated error variance associated with non-independent measures (teams in clinics), which is known to inflate type I errors in association tests (mediation). MCFA ensures valid reliability and validity estimates of theorized constructs, and ensures valid tests of Aim 2 mediation hypotheses. The single-factor structure will be specified consistent with prior STS psychometric validation and results reported.

Criterion/Predictive Validity. We expect systems thinking to be sensitive to change and discriminate between our PSD/AF arms at 6 and 12 months. We will conduct a cross-arm manipulation check of PSD/AF assignment on systems thinking. Measures will be standardized and baseline assignment will be assessed for an increase in systems thinking among PSD teams. We expect an effect size consistent with STS validation research showing a difference between groups based on receipt of systems thinking training ($d = .78$).¹³⁵

Reliability. Pre/Post Measurement Invariance will be examined with generalizability coefficients (GC)¹⁶⁴ in R package 'lavaan,'¹⁶³ which extend tests of internal consistency to designs with multiple sources of error. GC values measure the true variability of STS team-means in each arm as a proportion of total variance to determine the reliability of the scales across arms over time (Factor loadings and 95% confidence intervals). A standard set of increasingly constrained SEM Models will assess factor loadings, intercepts and residual variance for equivalence¹⁶⁵ and will detect the magnitude of measure non-invariance to obtain correct mediation inferences. Variance-covariance matrices of within- and between-arm latent factors will be plotted.

Qualitative and Observational Measures

Systems Thinking Codebook and Session Observations. Workshop Coding. We selected the STS self-report measure due to validation among health care staff receiving quality improvement training.¹³⁵ However, the primary systems thinking measures in *System Dynamics Review* are observation of systems thinking in language/explanations, and performance demonstrating system thinking skills (competence). Potential for coding bias precluded use of coding measures in mediation analyses, compared to the effect size estimate available for the STS. However, observational coding of online AF and PSD sessions, confers strengths for describing use or non-use of systems thinking to specific EBP-related decisions across AF/PSD arms.

AF/STS Fidelity and Use. Workshop Coding. We will review fidelity with qualitative checks against AF/PSD facilitator scripts for session learning objectives, 'key idea' and 'definitions,' including tracking the proportion of AF/PSD session activities (in minutes) on these components (see Table 9). Web-based Observation. Use of online AF and PSD resources will be tracked for criteria listed in Table 9 across the active AF/PSD phases, and for ongoing AF/PSD sustainment (see Timeline Table 10).

Design for Tests of Specific Aims

We propose a randomized, two-arm, parallel group CRT to test theoretical, confirmatory effectiveness (Aim 1), causality (Aim 2) and generality (Aim 3) hypotheses. To avoid inflating type I error for Aims 1 and 2 we will conduct two omnibus tests for improved EBP reach (initiate and course). Due to monthly fluctuations in EBP reach, we will calculate 12-month period pre/post EBP reach averages, which removes clustering of EBP reach observations *within* clinics over time in tests of specific aims.

See **Table 10** in **Section 4.2 – Statistical Design and Power – Human Subjects and Clinical Trials** for detailed information on our calculation of Intracluster Correlation Coefficient (ICC), Average Clinic Cluster Size, Power Analyses and plans for Attrition.

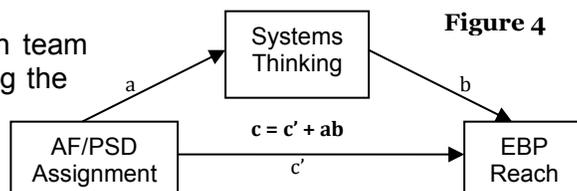
Analyses for Tests of Specific Aims

Aim 1. Test for superiority of PSD over AF for increasing EBP initiation and course. We will use R for tests of Aims 1a and 1b to establish PSD superiority. First, we will assess ICCs for within and between clinics. We will test for a significant difference between PSD and AF arms in increasing EBP reach using two (initiation and course) generalized estimating equation analyses for differences in proportions (reach).⁵⁷ We will estimate two (initiation and course) generalized linear models that account for clustering. We will assess distribution of EBP reach to identify the appropriate link function for robust standard errors, and test for a significant difference in EBP reach between the PSD and AF arms. We will report the effect size and 95% confidence intervals for the difference between arms at $\alpha = .05$.⁵⁷

Aim 2. Test the PSD theory of change that increased EBP reach is via systems thinking.^{47,58,59}

We hypothesize that the effect of PSD/AF

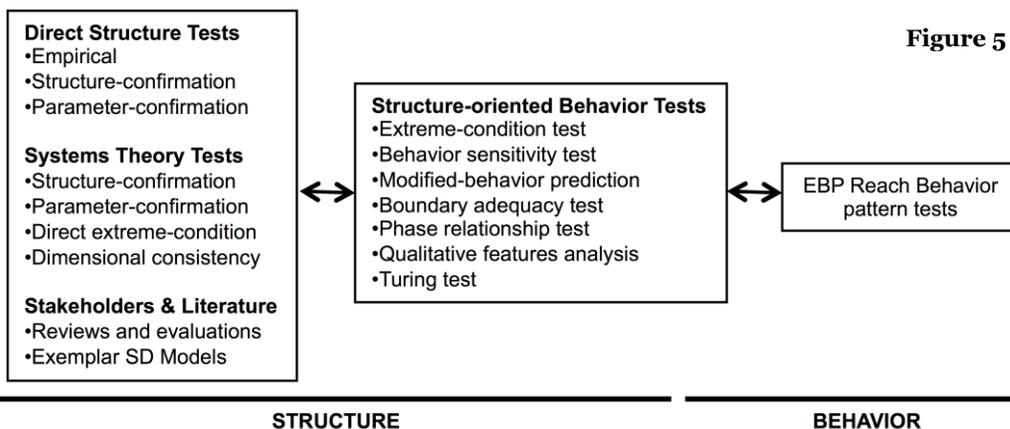
on 12 month period EBP reach will be explained by 6 month team systems thinking (STS), adjusting for baseline covariates using the ratio of mediator probability weight (RMPW).^{60–62} Our CRT design uses AF/PSD assignment (independent variable) to experimentally manipulate STS (mediator) on clinic-level EBP reach (outcome) in each arm. But, STS is not randomized.



RMPW uses sensitivity tests to address potential bias due to an interaction between the intervention and the mediator, or due to the operation of systems thinking through unhypothesized mechanisms (confounders) using multilevel mediation analyses in R package 'MultisiteMediation'⁶² a confirmatory hypothesis test using multilevel mediation with partial variance in 6 month STS and 12 month period EBP reach across arms. *Figure 4* displays the mediation model from strategies (baseline), to mediators (Baseline to 6 month STS change), to outcomes (12 months), where **a** = the standardized beta coefficient of PSD/AF assignment on systems thinking, **b** = the effect of systems thinking on EBP reach, **c** = the total effect of PSD/AF assignment on EBP reach, **c'** = the direct effect of PSD/AF on EBP reach, and **ab** = the indirect effect of PSD/AF on EBP reach through STS (hypothesized mediation). We will use a bootstrapping, asymmetric confidence interval approach to balance power and type I error.

Aim 3. Test the generality of mechanisms of change in EBPsy/EBPharm SD models. We will use structural-behavioral validation tests^{51,63,64} to evaluate whether causal mechanisms formulated in our Psy/EBPsy and MM/EBPharm models generalize to explain EBP reach as a function of local data across PSD and AF clinics. *Figure 5* displays the analytic plan we will use to establish the Structural-Behavioral validity of the two models against observed 12 month EBP Reach in both arms (AF and PSD), using the Theil Statistics module in Vensim DSS^{49,170}

PSD Models. We will document our models using SDM-DOC and report findings using SIMULATE.^{49,50}



Limitations and Alternatives

Limitations. Random assignment to usual AF or PSD is the best comparator for guiding decisions about scaling/sustaining PSD infrastructure. However, clinics in our 'usual AF' comparator will likely engage in highly variable AF activities. We take steps to address this with CDW-based stratification.

Alternatives. We will complement our primary tests of Aim 1a and 1b using autoregressive integrated

moving average (ARIMA) to mitigate potential limitations.^{171,172} This helps to guard against threats from secular trends (overall patient demand or VA EBP adoption may be increasing), and regression to the mean (given these are lower performing clinics). We will track national trends in EBP reach and estimate the percentage of regression towards the mean.^{173,174} We will also graphically display 12 month period averages of EBP reach related to Aim 1a and 1b at the clinic level, using statistical process control (SPC). SPC p-charts will display the pre/post intervention EBP reach proportions for all PSD and AF clinics. SPC is a standard, healthcare quality tool robust for non-normal data and unbalanced samples.¹⁷⁵ The p-chart centerline corresponds to the mean proportion of patients who meet EBP criteria, controlling for the number of patients in each observation.^{176,177} We will use SQUIRE standards for reporting SPC.^{177,178}

INNOVATION

Generalizable theoretical and empirical mechanisms to empower everyday EBP decision-makers.

This proposal tests the effectiveness, causes and generality of PSD against a highly prevalent comparator in the largest integrated health network in the U.S. The PSD approach empowers frontline teams to directly test their own local causal attributions for EBP reach, using theory-guided, mathematically specified simulations and existing health system data from their own team.^{38,85} Due to NCPTSD development of 'Modeling to Learn,' PSD learning can be scaled with online resources and virtual facilitation.^{39,44}

PSD is innovative meeting several needs for advancing the science of EBP implementation (Table 12).

Health care data systems are increasingly standardized (e.g., ICD and CPT codes).¹⁷⁹ Lessons learned in VA outpatient addiction and mental health can be translated to other VA health services,¹⁸⁰ and health care systems.¹⁸¹ The public health care and education missions of NCPTSD enable dissemination of free, fully transparent, open source PSD facilitation scripts, PSD models, and SQL code examples for standard EHR data extraction. The proposed CRT focuses on the significant problem of improving EBP reach, and advances IS with a paradigm that has potential for a significant pay-off. Due to the use of high-quality tests, *null findings for Aims 1-3* will also inform the IS discipline increasingly turning to operations and systems science.^{103,182-188} Our ability to formulate robust hypothesis tests is based on the translation of analogous fields to IS, rigorous qualitative and quantitative R21 preliminary research, and our ability to capitalize on the VA AF infrastructure, enabling a powerful comparison of PSD against one of the most commonly used strategies in the world.

Table 12. PSD Insights for Stakeholders & Implementation Science

1. Range of system behaviors explained by causal dynamics
2. Generalizable variable relationships across settings; Standard data inputs
3. Primary role of dynamics in producing EBP behavior
4. Secondary role of specific variable values in producing EBP behavior
5. Qualitative co-development and mutual learning with stakeholders
6. Quantitative testing of causal assumptions (feedback dynamics)
7. Causal mechanisms matched to local capacities/constraints (local data)
8. Intuitive visuals & 'system stories' for frontline teams (EBP implementers)
9. Without PSD underlying causal EBP dynamics is not obvious (black box)
10. Simulation an efficient, highly-effective way to select implementation plans
11. Experiential learning improves day-to-day EBP-related decision making

DESIGN

We propose a parallel two-arm, 24-site (12 clinics/arm) cluster randomized trial (CRT) to establish superiority of PSD over usual AF for increasing EBP reach. CRTs are best for complex interventions like PSD, with many interacting components, and in which the unit of intervention and observation is the clinic.⁵⁷ Number of clinics and number of patients per clinic define total CRT size. Our primary aim is to increase the proportion of the patient-population within PSD clinics that receive EBPs (reach: initiate/course). Cluster size is defined by the eligible depression, PTSD, OUD patient cohorts. Clinic proportion (reach), is the sum of patient-level EBP reach in usual care: patients receive (1) or do not receive (0) an EBP (binary).

Inclusion/Exclusion Criteria for Study Clinics. We will randomize and complete the PSD ‘*Modeling to Learn*’ virtual workshop with VA divisions and community-based outpatient clinics (CBOCs) or ‘clinics.’ Eligible clinics will be from regional VA health care systems (HCS) below the overall SAIL quality median, and 3 of 8 SAIL measures associated with 4 EBPsy and 3 EBPharm for depression, PTSD, and OUD in *Table 1*. Inclusion criteria balance sensitivity and specificity in identifying clinics from lower performing HCS.

Once identified via SAIL/CDW, recruitment will occur via OMHSP, NCPTSD, EBP Coordinator, and VISN Mental Health Lead networks (see Section 2.5 - Recruitment and Retention Plan). Clinics must have HCS director assent to randomization. Each clinic will be from a separate health care system (1 clinic/health care system). Analyses will control for clinics nested within VA.⁵⁷ Given known PSD interest, and R01 partner support, recruiting 12 clinics from these networks should be highly feasible, enhanced by 18 months total pre/post flexibility (see *Table 10*).

Stratification. Using baseline CDW data, we will conduct computer-assisted stratified block randomization with the R package ‘blockrand’¹⁵⁵ to balance arms for region, and 5 factors expected to influence clinic-level EBP reach: baseline EBP reach, clinic size (defined by two measures, both patients and providers), urban vs. rural location³ and medical division vs. community-based outpatient clinic (CBOC). We stagger start dates every 6 months (3 PSD clinics/3 AF clinics per wave), to mitigate cohort effects, making management of multi-site relationships feasible.

EBP Pre/Post Operationalization. For each wave, we use 12-month period average of EBP reach before AF/PSD start (pre-measure) and 12-month period average of EBP reach after AF/PSD end (post-measure).

ANALYSES FOR TESTS OF SPECIFIC AIMS

We propose a randomized, two-arm, parallel group CRT to test theoretical, *confirmatory* effectiveness (Aim 1), causality (Aim 2) and generality (Aim 3) hypotheses. To avoid inflating type I error for Aims 1 and 2 we will conduct two omnibus tests for improved EBP reach (initiate and course). Due to monthly fluctuations in EBP reach, we will calculate 12-month period pre/post EBP reach averages, which removes clustering of EBP reach observations *within* clinics over time in tests of specific aims.

Intraclass Correlation Coefficient (ICC). The ICC is the degree of similarity among EBP reach observations over time and within the same national VA health care system. We used the ICC to estimate the variance inflation factor/design effect for clustered data, a parameter for determining the number of clinics needed for tests of specific aims. Using R and the ‘ICC’ package,^{166,167} we analyzed a 2-year database extracted from the CDW to estimate the ICC of EBP reach for clinics nested within the VA health care system, and repeated observations of clinics over time. Across EBP initiation and course, values were low,^{168,169} ranging from low ICCs (< 0.0001) between clinics, to higher ICCs (~ 0.006) for repeated within-clinic observations. These ICC values indicate near independence,^{168,169} and are consistent with pilot work identifying significant differences between clinics, and high within-clinic variability in EBP reach over time.⁴⁸

Mediation analyses. Tests of Aim 2 are assessed on same between-clinic level: Clinics are assigned randomly to AF or PSD. We will use the clinic observation of team-means across arms as our mediator, and the clinic EBP reach as our outcome. Our mediator is adequately powered ($d = .78$).¹³⁵

Average Clinic Cluster Size. ICC and average cluster size are key to determining CRT power, accounting for variance inflation due to clustering. Our pre/post EBP reach measure is the 12-month period average of 12 monthly EBP reach observations. In our 2-year CDW database, median clinic size was ~800 unique patients/month summed across diagnostic cohorts (depression + PTSD + OUD + AUD diagnoses).

Power analyses. We used 'CRTSize'^{168,169} in R to calculate the number of clinics necessary to balance type 1 error (alpha), type 2 error (beta), and power, for a CRT with a binary outcome (difference in proportions). Omnibus EBP_{psy}/EBP_{pharm} *initiation* (Aim 1a) is the limiting power analysis.^{168,169} Effect size: We expect a 5% *initiation* increase in PSD clinics (from ~5 to 10%), which would meaningfully double omnibus EBP reach to exceed the national median. We expect little/no change in AF clinics (~ 5 to 5%). Thus, we expect PSD intervention will to lead to a 5% difference in EBP reach between PSD and AF arms. With 11 clinics/arm we have power = 0.80 to detect a 5% difference in EBP initiation (two-tailed test, alpha = 0.05, average clinic/cluster size of 800, and ICC = 0.02). However, we will use an even 12 clinics/arm, better for stratified block randomization.¹⁵⁵ Power analyses used ICC = .02, because our STS team-mean measures should become more similar after PSD/AF, leading to higher ICC values than in our CDW clinic population database.⁵⁷

Attrition. Attrition of providers or patients will not impact analyses for specific aims. Provider PSD participation will be tracked as a PSD fidelity check, but patient attrition is included by definition in EBP reach measures, and use of provider attrition in team-average (mean) survey measures means the only loss of data would be due to loss of an entire clinic team. Our R21 pilot testing indicates it is unlikely that care teams will attrit. Should clinics attrit after randomization, they will be included in intent-to-treat analyses using CDW data.

STRUCTURE OF THE STUDY TEAM

Serving more than 8.9 million Veterans each year, VA is the largest health care system in the U.S., providing care at 1,053 outpatient clinics. Due to our participatory system dynamics (PSD) philosophy of science, interdisciplinary PSD application, and the scale of VA, a large team with a range of investigator, methodologist, and advisory expertise, ensures R01 success.

PI and Co-Investigators NCPTSD principal investigator and co-investigators (*Table 8*) have collectively facilitated many, multi-site implementation studies. R01 feasibility is further enhanced by our established partnerships. Eleven of seventeen partners collaborated effectively as R21 co-investigators or NCPTSD-funded PSD development partners.⁴⁸ Three new partners were added to our R01 team due to study relevant expertise with implementation science in addiction and mental health (McGovern), cluster randomized implementation studies (Wiltsey Stirman) and scaling system dynamics trainings (Snyder).

Advisory and Safety Monitoring Board In addition to our VAPOR partners, four study advisors lead national VA improvement initiatives: VA system organization (Rust), SAIL audit and feedback quality measures in mental health (Trafton), national evidence-based psychotherapy coordinator program (Collie) and the national evidence-based pharmacotherapy program (Wiechers). Rust and Trafton (R21 co-investigators) worked with us to develop *Modeling to Learn* at scale in VA. The study relevant expertise and contributions of each R01 study team member or partner is listed below in *Table 8*.

During the R21, we collaborated effectively across our nationally distributed team through regular meetings on the Lucid Meetings platform, project management via Basecamp 2.0, an open data science workflow on GitHub, and an integrated set of servers behind the VA firewall. We also benefit from co-location of NCPTSD, Office of Mental Health and Suicide Prevention (OMHSP), Veterans Advisory Partnership for Operations and Research (VAPOR), and Stanford partners at VAPAHCS (see included letters of support). NCPTSD, Georgia Health Policy Center (GHPC) and VAPOR staff will co-facilitate '*Modeling to Learn*' workshops. NCPTSD staff will provide post-workshop technical assistance.

Table 8. R01 Study Team & PSD Partners

Name	Role	Organization	Responsibilities during the R01 Trial	Specific Contributions
Zimmerman*+	PI	NCPTSD	Implementation, PSD, Mediation, Qualitative	Aims 1, 2 & 3
Lounsbury*	co-I	Einstein	Participatory System Dynamics (PSD)	Aims 1, 2 & 3
McGovern+	co-I	Stanford	Addiction, Mental Health, Implementation	Aims 1 & 2
Rosen*+	co-I	NCPTSD	EBP Dissemination, VA Multi-site Studies	Aims 1 & 2
Kimerling*+	co-I	NCPTSD	Health Services Research, Staff Engagement	Aims 1 & 2
Wiltsey Stirman+	co-I	NCPTSD	Implementation Science/CRT	Aims 1 & 2
Lindley*+	co-I	VAPAHCS	Frontline Management	Field Expertise
Snyder	co-I	GHPC	PSD Facilitation Administration	PSD Trainings
Branscomb*	co-I	GHPC	PSD Facilitation/Facilitation Training	PSD Trainings
Hong*	key personnel	NCPTSD	CRT statistician, R and GitHub Expertise	Aims 1 & 2
Holbrook*	key personnel	VERC	Industrial Engineer/CDW SQL Programmer	CDW data/SQL code
Azevedo*+	key personnel	VAPAHCS	Qualitative Research and Coding	Aim 2 & Fidelity
Rust*	advisor	VERC	Systems Engineer/System Dynamicist	Aims 2 & 3
VAPOR*+	advisor	VAPAHCS	Peer Specialists: Patient Perspective	PSD Trainings
Trafton*+	advisor	OMHSP	SAIL Measures, VA Policy	AF/SAIL Code
Collie	advisor	OMHSP	VA VISN and EBP MH Leadership	AF/Clinic Engagement
Wiechers+	advisor	OMHSP	VA Psychotropic Drug Safety Initiative	AF/Clinic Engagement

Note: * Co-investigators on R21 and partners in NCPTSD development. + Co-located at VAPAHCS = VA Palo Alto Health Care System. NCPSTD = National Center for PTSD. Einstein = Albert Einstein College of Medicine. Stanford = Stanford University. OMHSP = Office of Mental Health and Suicide Prevention. VERC = Veterans Engineering Resource Center. GHPC = Georgia Health Policy Center. VAPOR = Veterans Advisory Partnership for Operations and Research. PSD = Participatory System Dynamics. AF = Audit and Feedback. CRT = Cluster Randomized Trial. CDW = Corporate Data Warehouse.

DATA SAFETY AND MONITORING PLAN

We propose an implementation trial focused on the comparative effectiveness of two health care quality improvement strategies (participatory system dynamics vs audit and feedback). We will compare these two strategies to test their relative effectiveness for increasing the proportion of the outpatient addiction and mental health patient population that receives the highest quality, evidence-based standard of care.

Our two-arm trial focuses on changing provider care decisions to expand evidence-based treatments in routine care. Therefore, although our proposed R01 meets criteria for a phase III clinical trial, we expect that it is low risk with regard to patients and providers. We expect it to be low risk because we will not interact with patients for the purposes of research during this trial. Our focus is on the reach of evidence-based practices as measured in the VA electronic health record system. These psychotherapies and pharmacotherapies are referred to as “evidence-based practices” based on multiple randomized controlled trials, meta-analyses, and pragmatic effectiveness research studies.

This is a large national implementation science trial, but we are not examining the effectiveness of evidence-based practices. We are focused on expanding their reach among patients. There is no prior data to suggest that audit and feedback or participatory system dynamics approaches to quality improvement have significant adverse effects for patients or providers. Given this, we will take the following steps to ensure adequate data safety and monitoring.

Data Safety and Monitoring Plan

All investigators and project staff will complete necessary coursework regarding protection of human subjects and will receive certification from the Collaborative IRB Training Initiative (CITI). All investigators and project staff will remain current on VA privacy and information security trainings. We will also submit all procedures and documentation/definitions for electronic health record data collected to the relevant IRB (Stanford University) and VA Offices (VHA National Data Systems, VA Informatics and Computing Infrastructure, VA Office of Research and Development) for review and oversight.

We will maintain ongoing communication with our data safety and monitoring board and will regularly review data management procedures to identify and address unintended problems, and address any unlikely, but possible, adverse events.

Data Safety and Monitoring Board

Commensurate with the low risk of the trial, and commensurate with the size and complexity of our trial, our advisory board of VA quality improvement leaders in VA will comprise our data safety and monitoring board. One member runs the VA national audit and feedback program for the VA Office of Mental Health and Suicide Prevention (OMHSP) as the Director of the Program Evaluation Resource Center. A second member runs the national VA OMHSP quality improvement program for evidence-based pharmacotherapy as director of the Psychotropic Drug Safety Initiative. A third member runs the national VA quality improvement program for evidence-based psychotherapy as an OMHSP Technical Assistance Specialist. A fourth member runs national VA systems engineering quality improvement programs using EHR data. Finally, the Veterans Advisory Partnership for Operations and Research (VAPOR) provides ongoing input in study plans and evaluation from the Veteran patient perspective. This advisory board, will help to provide oversight and monitoring of the trial through the independent activities of their programs, and through regular communication with the study team.

We will meet regularly with our advisory board and will perform data safety and monitoring activities every six months. Monitoring activities will include review of study data in light of overall VA quality improvement data with our complete multidisciplinary R01 study team. Our advisory board will provide assessments of trial progress based on the input of their independent teams of multi-disciplinary program evaluators, and will advise the R01 study team accordingly.

In addition to the advisory board, the R01 study team includes clinical trial experts, VA health services research experts, and clinicians who are experts in the evidence-based psychotherapies and pharmacotherapies that this trial seeks to make accessible to more VA patients.

RECRUITMENT AND RETENTION PLAN

Selection of Study Clinics – Inclusion/Exclusion Criteria

We will randomize and complete the participatory system dynamics ‘*Modeling to Learn*’ (PSD) or audit and feedback ‘*Team Feedback*’ (AF) virtual workshop with VA divisions and community-based outpatient clinics (CBOCs) or ‘clinics.’ Eligible clinics will be from regional VA health systems below the overall VA quality median (as assessed by the Strategic Analytics for Improvement and Learning or SAIL). This includes 3 of 8 SAIL measures associated with four evidence-based psychotherapies and three evidence-based pharmacotherapies for depression, PTSD, and opioid use disorder in (see *Table 1* in *Strategy*).

Inclusion criteria balance sensitivity and specificity in identifying clinics from lower performing health care systems. We focus on the lower half of our national distribution in EBP reach, which is the relevant population for inference from the R01 study sample, and for interpreting R01 results within the literature.

Recruitment of Study Clinics

Once identified via SAIL, recruitment will occur via the networks and training programs of the VA Central Office quality programs run by the study team. These offices and programs include the Office of Mental Health and Suicide Prevention (OMHSP), the National Center for PTSD (NCPTSD), Dissemination and Training Division, the NCPTSD National Mentorship Program, the NCPTSD Practice-based Implementation Network, the OMHSP EBP Coordinator program, the OMHSP Technical Assistance Specialists, OMHSP Program Evaluation Resource Center, OMHSP Psychotropic Drug Safety Initiative, VA Academic Detailing Program, the Behavioral Health Integrated Plan (BHIP) Coaches Program, and VISN Mental Health Lead networks (Veterans Integrated Service Network [VISN] - multi-state regional health care system networks).

Clinics must have regional health care system director assent to randomization. Analyses will control for clinics nested within VA. Given known interest in participatory system dynamics among health care systems due to the activities of our quality improvement programs, recruiting from these networks should be highly feasible. Feasibility is further enhanced by 18 months total pre/post flexibility (see Section 2.7 of Human Subjects and Clinical Trials Information - *Timeline, Table 10*).

Engagement and Retention of Teams in Study Clinics

AF participants will engage during two regular team meetings in 1 month (2 hours), and receive weekly emails for 6 months; PSD participants will engage during two regular team meetings over 6 months (12 hours), with weekly emails for 6 months (*Strategy AF/PSD Workshop Training and AF Dynamic Data Tools*). AF/PSD will occur during normal meetings, substituting only the activities used to improve quality objectives.

Continuing Education Licensure Credit. All frontline addiction and mental health disciplines will have the opportunity to participate for licensure credit provided by VA Employee Education Services. This includes psychiatrists, psychologists, social workers, nurses, counselors and certified peer support specialists. Providers randomized to the audit and feedback arm will have the opportunity to receive two hours of licensure credit for two facilitated team meetings. Providers randomized to participatory system dynamics will have the opportunity to receive twelve hours of licensure credit for twelve facilitated team meetings over six months.

Attrition. Attrition of providers or patients will not impact analyses for specific aims. Provider PSD participation will be tracked as a PSD fidelity check, but patient attrition is included by definition in EBP reach measures, and use of provider attrition in team-average (mean) survey measures means the only loss of data would be due to loss of an entire clinic team. Our R21 pilot testing indicates it is unlikely that care teams will attrit. Should clinics attrit after randomization, they will be included in intent-to-treat analyses using CDW data.

INCLUSION OF WOMEN, MINORITIES, AND CHILDREN

Inclusion of Women and Minorities:

The proposed research will be open to adult men and women of all ethnic and racial backgrounds. Women and members of minority groups from *two populations* will be included in the proposed R01 study: the VA patient population served in outpatient mental health and addiction services (existing VA health system data) and the VA outpatient mental health and addiction provider population (survey data; see second page). Our trial plans are inclusive for both patients and providers.

The R01 patient and provider samples will be representative of VA patient and provider populations. The VA patient population skews male. We address sex/gender, race, and ethnicity in our proposed trial design with inclusive eligibility criteria. All VA addiction and mental health patients with a primary diagnosis of depression, PTSD, alcohol use disorder and opioid use disorder are eligible. All VA addiction and mental health providers mapped to a care team in a study clinic are eligible. We will randomize provider participants to each arm. Our evaluation of the trial outcome measure is drawn from existing health care records (see *Patients* below).

Prior studies neither support nor negate the potential for significant differences in participatory system dynamics or audit and feedback effectiveness for improving the reach of evidence-based practices among sex/gender, racial or ethnic subgroups. Our plans for valid analysis include reporting results by subgroup, and exploring for significant differences in our outcome of EBP reach (patients), and our mediator, systems thinking (providers). The targeted/planned distribution of subjects by sex/gender, racial, and ethnic groups for each proposed sample is provided in two Targeted/Planned Enrollment Tables below.

Patients. By VA policy, all VA patients should have equal access to EBPs. Existing health system data (e.g., means/median number of patients receiving specific services, means/median of scheduled clinic appointments) from a 2-year cohort of patient services data will be available for audit and feedback to frontline staff in order to achieve higher quality care (arm 1), or available for review by frontline staff and synthesized in participatory system dynamics models for simulation testing (arm 2). Analyses to test specific aims include a 24-month observation of health services delivery (12 months pre-/12 months post-) across 24 clinics.

Enrollment - Unique Patients in Multisite Mental Health Cohort					
	Not Hispanic or Latino		Hispanic or Latino		Total
Racial Categories	Female	Male	Female	Male	
American Indian/Alaska Native	314	3,157	50	504	4,025
Asian	64	649	10	104	827
Native Hawaiian or Other Pacific Islander	35	351	6	56	448
Black or African American	291	2,929	46	468	3,734
White	1,039	10,455	166	1,670	13,330
More than One Race					
Total	1,743	17,541	278	2,802	22,364

Providers. Teams (typically 4-6 staff) of frontline multidisciplinary mental health and addiction staff from 24 participating clinics will be randomized to six months of team audit and feedback or six months of team participatory system dynamics training, each designed for improvement in delivery of high-quality, evidence-based care. Pre-post provider survey measures will be collected at baseline and at the end of the improvement strategy (six months). All addiction and mental health providers mapped to a care team in our study clinics will have the opportunity to participate.

Provider Demographics					
	Not Hispanic or Latino		Hispanic or Latino		Total
Racial Categories	Female	Male	Female	Male	
American Indian/Alaska Native	8	5	1	1	15
Asian	32	20	3	2	57
Native Hawaiian or Other Pacific Islander	8	5	1	1	15
Black or African American	78	52	9	6	145
White	245	163	27	18	453
More than One Race	19	13	2	1	35
Total	390	258	43	29	720

Inclusion of Children:

All participants will be 18 years of age or older. There will be no children involved. This is a study of adult outpatient mental health and addiction services in the VA. The purpose of the study is to understand how to expand the reach of evidence-based psychotherapies and evidence-based pharmacotherapies determined to be effective for adult diagnosis with depression, PTSD, alcohol or opioid use disorder. The study will take place in the national VA health care system, which serves adult male and female patients.

This trial does not enroll patients and will not interact with patients for the purposes of research. Rather the evidence-based practice outcome measure will be observed in existing VA electronic health record systems.

The comparators in this two-arm cluster randomized trial are participatory system dynamics and audit and feedback. These two interventions are each designed to improve evidence-based practice implementation in the VA health care system by intervening with the professional (adult) frontline addiction staff of the VA.

Therefore, due to the trial target to expand the reach of adult treatments (i.e., evidence-based practices), and due to the focus on intervention with professional health care staff, participation of children in this trial is not scientifically or ethically justified.

STUDY TIMELINE

Timeline and Feasibility.

We propose 60 total months of study activities and 30 months of active PSD facilitation or post-training technical assistance (phases 3 and 4). This leaves 24 months of flexibility for delays across pre (phases 1 and 2) and post (phases 5 thru 8) activities (*Table 10*).

Table 10. Timeline of Proposed Study Activities

Year	1		2		3		4		5	
Months	6	12	18	24	30	36	42	48	54	60
Block 1										
Block 2										
Block 3										
Block 4										
CDW Data										
Surveys										
Qualitative										
Phase 1										
Phase 2										
Phase 3										
Phase 4										
Phase 5										
Phase 6										
Phase 7										
Phase 8										

- IRB, Recruit, Randomize 24 Clinics: 12 per Arm & 6 per wave
- Qualitative Codebook Developed/Adapted (Cbook)
- Pre 12-month EBP Reach; Team Data Websites & Models
- AF/PSD Trainings to Select Change Plans over 6 months
- AF/PSD Post-Training Technical Assistance
- Aim 1*** AF/PSD Post 12-month period EBP reach
- Aim 2*** Qualitative Coding Analyses
- Aim 2*** Survey Psychometric and Descriptive Analyses
- Explore* Online AF/PSD Sustained Use Monitoring
- Aim 2*** Systems Thinking Mediation Analyses
- Aim 3*** Model Generality Analyses for AF Clinics
- Dissemination of Findings, Make Study Data Accessible
- Disseminate PSD Workshop Scripts, Code & Models

PROTECTION OF HUMAN SUBJECTS

This Human Subjects Research involves an NIH-Defined Phase III Clinical Trial.

1. Risks to Human Subjects

a. Human Subjects Involvement, Characteristics, and Design

Human Subjects Involvement and Characteristics. Electronic data from the VA Corporate Data Warehouse (CDW) will be used to evaluate specific aims and comprise the *first of two* human subjects components to our study. Due to our proposal using existing data during this project, risks associated with this study to individual patients who use mental health services are minimal. There will be no interaction with current patients for the purposes of research. Patients will not be asked to sign a consent form. No new data will be collected beyond data generated during routine care. All individual patient data will stay on servers behind the VA firewall to prevent any potential risk for loss of confidentiality of protected health information. Data inputs in the models will be de-identified team aggregates and will not be individually identifiable.

Staff/Stakeholder Involvement. This project is a collaboration between the Principal Investigator and Co-Investigator team, and the leadership, front-line providers, and staff in the VA outpatient service system. Comprising the *second of two* human subjects components of our study, Mental Health staff will be engaged in the Audit and Feedback (AF) 'Team Feedback' and Participatory System Dynamics (PSD) 'Modeling to Learn' team trainings (i.e., they will participate in their actual team/workgroup) for no-cost, with continuing education credits provided by VA Employee Education Services toward licenses in psychiatry, psychology, social work, nursing, certified peer support specialty, and counseling. These educational trainings are commonly provided in VA and are necessary for maintenance of state licensure and VA hospital privileges. During workshop sessions, frontline teams of providers will partner to evaluate improvement scenarios via simulation and identify quality improvement changes in their clinics/teams.

Over the course of the proposed study, this research project expects to involve eight regional health systems randomized to either AF or PSD, in which the Medical Director has identified at least three clinics for participation in training (4 regional health systems/12 clinics per arm; 24 clinics total). We expect approximately 30 staff to participate in each outpatient clinic. Across 24 clinics that will include 720 staff. AF/PSD workshops with staff will be held during regularly scheduled staff meetings or team huddles. We anticipate that participation in PSD will include approximately two workshop hours per month for six months with optional self-directed learning. Participation in AF activities will include two hours in the first month, and less than one self-directed hour per month during months two through six. Participation will occur during work hours and require clinic manager workload credit approval to staff for participation. Staff will receive 12 continuing education credit hours for PSD and two continuing education credit hours for AF. Since randomization occurs at the VA regional health system level, the training opportunity will not vary among co-workers. Local staff will either all receive AF or all receive PSD. The leaders and staff of the outpatient service system helped to shape the goals of this study. Veteran staff with lived experience using the mental health and addiction service systems, who now work as VA patient navigators, will continue to participate and shape the development of the project through all phases, including their role as workshop co-facilitators. These certified peer specialists from the Veteran Advisory Partnership for Operations and Research complete CITI Training.

b. Sources of Materials

Sources of Materials – Patient Human Subjects. Administrative data and VA information systems will be used to evaluate and compare staffing allocations, patient referral flows, appointment timing and type, and pharmacy records. These electronic data will be drawn from the regional VA Corporate Data Warehouse (CDW).

Sources of Materials – Staff/stakeholder Human Subjects. We will collect anonymized, qualitative and quantitative data from mental health staff. Routine data about team care patterns, and information about staff decisions regarding change will be collected throughout the project as part of PSD and AF

activities. Licensure accreditation bodies and VA employee education services each have educational course evaluation requirements that will not comprise research data, but will be required for staff to obtain licensure credits.

c. Potential Risks

The primary risk to human subjects/mental health patients is associated with potential breaches of confidentiality of patient health records. In addition, staff who participate may feel uncomfortable about the review of data and our focus on the performance of the mental health delivery system.

2. Adequacy of Protection Against Risks

a. Recruitment and Informed Consent

The research involves no more than minimal risk to the patient participants. Patient data will be extracted from VA administrative datasets. These data are collected during routine mental health care. We will seek a waiver of consent and a HIPAA waiver. We will not recruit, conduct informed consent, or interact with patients. Precautions will be taken to ensure that confidentiality is maintained. Individually identifiable patient data will not be synthesized in models. All files related to study data will be password protected and will only be accessible by those working directly on the study.

This research involves no more than minimal risk to the provider participants. AF is already standard practice in VA. PSD is an augmented version of AF for achieving greater improvements in quality. Eligible VA regional health systems will be below the overall VA national quality median. Regional VA health systems must have the Medical Director assent to randomization and have identified three clinics within their regional system that are willing to participate. All the clinics from the regional system, and the teams that comprise those clinics, will receive the same implementation strategy (Audit and Feedback or Participatory System Dynamics).

Recruitment will occur via the VA offices responsible for ensuring high-quality EBP delivery: Office of Mental Health and Suicide Prevention (OMHSP), National Center for PTSD (NCPTSD), Quality Enhancement Research Initiative (QUERI), Evidence-based Practice (EBP) Coordinators, and Regional Mental Health Leads. These national groups of directors, program leads and managers, each have listservs and regular meetings. We will use email and online/in-person presentations to introduce the AF/PSD team training opportunities, for the purposes of recruitment. The R01 study team includes co-investigators and advisors from OMHSP, NCPTSD, EBP Coordinators and Regional Mental Health Leads.

Providers will provide assent or implied consent to participate in the AF or PSD training workshop. Providers will be informed in online training materials that the team AF/PSD training opportunity is designed to improve the quality of addiction and mental health care with the opportunity to receive continuing education credits toward licensure. They will be informed the training is voluntary, and of the possible risks, such as discomfort talking about quality issues and team care coordination issues. Providers will also be informed of potential benefits, such as the ability to improve patient care and provider quality of work-life. Staff do not need to participate for continuing education credit to be involved in this team learning opportunity.

Staff online/written informed consent will be obtained for completion of a 52-item survey at baseline and an 83-item survey at six months for the purposes of research. Staff will be informed of the potential knowledge to be gained from the surveys. Staff will be informed of the opportunity to stop survey participation at any time without penalty. Any new findings developed during the course of the study, which may relate to staff willingness to continue to participate will be provided to staff. Providers will have the option not to participate in the survey measures or withdraw their participation at any time.

b. Protections Against Risk

Protection against Breaches of Confidentiality. Patient information in the VA administrative data systems will not be transported. Real-time aggregate data reports will be extracted from existing

quality assurance algorithms leaving the data on VA servers. Individually identifiable patient data will not be synthesized in models. All files related to study data will be password protected and will only be accessible by those working directly on the study.

Protection against Staff Discomfort. Staff, by agreeing to participate in the voluntary team training, will be giving assent and permission to review health system data to improve EBP implementation and overall mental health and addiction service quality. This will be stated clearly at the beginning of project activities and participants may withdraw from the training or decline to discuss without penalty; withdrawing from the team trainings will in no way impact access to consultation or data services developed or made available as part of this study. Based on past research collaboration with outpatient stakeholders, we expect that staff will appreciate the opportunity to share their experiences, challenges, and concerns. However, should they experience distress as a result of participating in this research partnership, we will refer them to a member of our VA Office of Research and Development, and will notify the IRB. Co-Investigators and advisors on this grant have extensive experience working with and addressing the problems of VA staff (Drs. Lindley, Rosen, Kimerling, Collie, and Trafton).

3. Potential Benefits of the Proposed Research to Human Subjects and Others

The purpose of this project is to test the use of PSD against usual AF. Based on large bodies of AF and PSD research, we expect PSD to enable frontline mental health staff to better identify improvements to mental health delivery that increase the proportion of the patient population who receive high-quality, evidence-based care. We will learn about causes of limited EBP reach, the explanation for findings regarding AF/PSD effectiveness, and we will learn how generalizable the PSD models are across a wide range of clinics/regional health systems. This project should improve Veterans' health and well-being by improving the quality of their care. It is possible that PSD/AF activities will also improve providers' quality of work-life on their teams as improvements are identified. The electronic data systems synthesized in system dynamics models will also enable us to prepare to expand this benefit more widely beyond VA through future research through online public dissemination of models, code and training resources. Finally, this project is designed to maximally support and increase the capacities of stakeholders in outpatient services delivering care via mental health, and therefore we aim to build the skills and knowledge of staff who are participating.

4. Importance of the Knowledge to be Gained

We're proposing a Phase III Clinical Superiority design with falsifiable hypothesis tests. We're testing the most commonly used, but often ineffective strategy in the world (AF), against a very rigorous theory-based approach (PSD) with a 60-year track record of effectiveness for improving organizations/business. Confirmatory effectiveness (Aim 1), causality (Aim 2) and generalizability (Aim 3) hypotheses are relatively rare in the field of implementation science. The ultimate anticipated benefit of this project lies in its potential to identify consistent ways to increase the reach of evidence-based practices (EBP) to patients in need of services. Toward this end, study aims are designed to inform replication of the use of PSD should R01 findings warrant further study. Our project activities are the first step toward creating a paradigm for ongoing quality improvements in health service delivery. Due to our use of VA nationwide data extraction approaches, this method can be scaled and applied to other VA health specialties beyond mental health. Use of standard data definitions for coding patient diagnoses or patient-provider encounters (visits) makes the study ready for replication in any U.S. healthcare system. This program of research could improve the quality of care in the future for large populations of patients and help the VA and other health systems to make better use of existing resources (i.e., staffing) to provide highly effective treatments.

IRB PLAN

Our local IRB is the Stanford University IRB and the Office of Research at VAPAHCs.

The study team at the National Center for PTSD will manage the R01 trial protocol from the campus of the VAPAHCs. All clinics enrolled in this R01 study will complete the same randomization protocol and participate in either audit and feedback (AF) or in participatory system dynamics (PSD). These interventions and all trial records and data will be managed centrally from one lead site at NCPTSD/VAPAHCs.

DISSEMINATION PLAN

Registration of Clinical Trial

We will register this R01 trial and follow clinical trials reporting standards and policies via the ClinicalTrials.Gov website. NCPTSD, VA Palo Alto Health Care System (VAPAHCS), and Stanford University, all have internal policies in place to ensure that clinical trials registration and results reporting are in compliance federal clinical trials policy requirements.

Data, Facilitation Guides and Code

A central component of each arm of this participatory research project is to increase data transparency and accessibility among local frontline providers. Should participatory system dynamics prove to be superior and/or effective for improving reach of evidence-based practices, then dissemination of system dynamics modeling scripts, code, and models will be made available for use in other implementation contexts, and for replication by other implementation researchers.

Data. Datasets meeting VA standards for disclosure to the public will be made available within 1 year of publication. A de-identified, anonymized dataset will be created and shared. Final data sets underlying all publications resulting from the proposed research will be shared publicly.

Prior to distribution, a privacy officer will certify that all datasets contain no PII/PHI. Final data sets will be maintained locally, until VA and NIH enterprise-level resources become available for long-term storage and access. The VA Office of Research Development (ORD) will provide guidance on request and distribution processes. Those requesting data will be asked to sign a Letter of Agreement regarding use.

Publicly Available Participatory System Modeling Resources. In addition to these datasets, model SQL code for retrieving data from generic health record systems will be made publicly available online. The system dynamics model files and group modeling scripts used in the “*Modeling to Learn*” workshop series will be posted online for transparent, public use.

Publications

Publications from this research will be made available to the public through the National Library of Medicine PubMed Central website within one year after the date of publication, in accordance with guidance provided by NIH and VA ORD.

RESOURCE SHARING PLAN

Data and Code

A central component of each arm of this participatory research project is to increase data transparency and accessibility among local frontline providers. Should participatory system dynamics prove to be superior and/or effective for improving reach of evidence-based practices, then dissemination of system dynamics modeling scripts, code, and models will be made available for use in other implementation contexts, and for replication by other implementation researchers.

Data. Datasets meeting VA standards for disclosure to the public will be made available within one year of publication. A de-identified, anonymized dataset will be created and shared. Final datasets underlying all publications resulting from the proposed research will be shared publicly.

Prior to distribution, a privacy officer will certify that all datasets contain no PII/PHI. Final datasets will be maintained locally, until VA and NIH enterprise-level resources become available for long-term storage and access. Guidance on request and distribution processes will be provided by the VA Office of Research Development (ORD). Those requesting data will be asked to sign a Letter of Agreement regarding use.

Publicly Available Participatory System Modeling Resources. In addition to these datasets, model SQL code for retrieving data from generic health record systems will be made publicly available online. The system dynamics model files and group modeling scripts used in the “Modeling to Learn” workshop series will be posted online for transparent, public use.

Publications from this research will be made available to the public through the National Library of Medicine PubMed Central website within one year after the date of publication, in accordance with guidance provided by NIH and VA ORD.