








© Health Research and Educational Trust  
DOI: 10.1111/1475-6773.12986  
POLICY-MANAGERIAL IMPACT ARTICLE

# Overcoming Challenges to Evidence-Based Policy Development in a Large, Integrated Delivery System

*Austin B. Frakt* , *Julia C. Prentice* , *Steven D. Pizer* ,  
*A. Rani Elwy* , *Melissa M. Garrido* , *Amy M. Kilbourne* ,  
and *David Atkins* 

---

**Objective.** To describe a new Veterans Health Administration (VHA) program to foster the learning health system paradigm by rigorously evaluating health care initiatives and to report key lessons learned in designing those evaluations.

**Principal Findings.** The VHA's Quality Enhancement Research Initiative and its Health Services Research and Development Service are cooperating on several large, randomized program evaluations aimed at improving the care veterans receive and the efficiency with which it is delivered. The evaluations we describe involve collaborative design, outcomes assessment, and implementation science through partnerships between VHA operations and researchers. We review key factors to assess before committing to an evaluation. In addition to traditional design issues (such as ensuring adequate power and availability of data), these include others that are easily overlooked: the stability of intervention financing, means of controlling and commitment to adhering to randomized roll-out, degree of buy-in from key implementation staff, and feasibility of managing multiple veto points for interventions that span several programs, among others.

**Conclusions.** Successful program implementation and rigorous evaluation require resources, specialized expertise, and careful planning. If the learning health system model is to be sustained, organizations will need dedicated programs to prioritize resources and continuously adapt evaluation designs.

**Key Words.** Evaluation design and research, randomized program evaluation, health care organizations and systems, VA health care system

---

The United States spends over \$3 trillion on health care, but less than 0.1 percent of that total on evaluating health care programs and policies (Bridgeland and Orszag 2013). A smaller fraction is devoted to evaluations that include randomization (Finkelstein and Taubman 2015a). Because they confer strong

control for observable and unobservable characteristics of subjects, randomized controlled trials (RCTs) are widely regarded as the evidentiary gold standard. They are also relatively easy to understand compared to rigorous observational approaches. These features of RCTs strengthen credibility and facilitate dissemination of results, accounting for their influence in medical science (Bothwell et al. 2016). Their relatively limited use for health systems interventions is a missed opportunity.

However, there is a burgeoning movement to promote evidence-based policy and randomized evaluations, including within the Veterans Health Administration (VHA) (Atkins, Kilbourne, and Shulkin 2017; Kilbourne et al. 2017). In 2012, the National Academy of Medicine called for enhanced digital infrastructure to facilitate research and evidence-based practice (National Academy of Medicine 2012). Several bills have been introduced in Congress (U.S. House of Representatives 2013, 2017) that would have required agencies to base decisions on evidence. The Evidence-Based Policy Commission Act of 2016 was a bipartisan effort to increase data available to support evidence-based policy (U.S. House of Representatives 2016).

Researchers nonetheless face challenges in initiating program evaluations with randomized designs. The requirements of RCTs—in particular, that intervention is delayed or withheld for some units—are not always consistent with health system managers' goals of rapidly deploying an initiative; a number of pragmatic questions arise in trial design, and obstacles often emerge during implementation. Design questions include those pertaining to trial size and duration, the unit of randomization (e.g., individual vs. site), inclusion or exclusion of participants, whether consent or institutional review board oversight is required, and timing of randomization (e.g., classic, two-arm RCT vs. stepped wedge) (Newhouse and Normand 2017). Moreover, comprehensive randomized program evaluations (RPEs) involve a broad set of stakeholders and researchers with diverse skills and differing priorities, often requiring delicate coordination throughout program roll-out and evaluation.

---

Address correspondence to Austin B. Frakt, Ph.D., PEPRc, VA Boston Healthcare System, 150 S. Huntington Ave. (Mailstop 152H), Boston, MA 02130; e-mail: Austin.Frakt@va.gov. Julia C. Prentice, M.S.P.H., Ph.D., is with CAPER, VA Boston Healthcare System, Boston, MA. Steven D. Pizer, Ph.D., is with the Department of Health Law, Policy & Management, Boston University School of Public Health, Boston, MA. A. Rani Elwy, Ph.D., is with CHOIR, VA Boston Healthcare System, Boston, MA. Melissa M. Garrido, Ph.D., is with the James J Peters (Bronx) VA Medical Center, Bronx, NY. Amy M. Kilbourne, M.P.H., Ph.D., is with QUERI, Veterans Health Administration, Ann Arbor, MI. David Atkins, M.D., M.P.H., is with HSR&D, Veterans Health Administration, Washington, DC.

We have encountered and overcome several of these obstacles in our efforts to launch four large RPEs that focus on VHA national priorities: (1) home- and community-based care for veterans at risk of nursing home placement, (2) opioid prescription risk mitigation, (3) suicide risk assessment and prevention, and (4) a telehealth tool for improving access to dermatology services. As the largest U.S. integrated delivery system, serving a population of 8.9 million enrolled veterans per year, the VHA offers unique opportunities to employ rigorous evaluation designs for new policies. This paper describes lessons learned in the design phase of these randomized program evaluations within the VHA. Because program roll-out has not been completed, our assessment does not allow for examination of problems encountered during design implementation.

The VHA is an emerging leader in advancing evidence-based policy within a learning health care system, in which research helps drive practice and policy change (National Academy of Medicine [formerly Institute of Medicine] 2012; Atkins, Kilbourne, and Shulkin 2017; Etheredge 2007). The Quality Enhancement Research Initiative (QUERI) has priority goals of rapidly disseminating best practices, increasing the impact of VHA research through rigorous evaluation design—with a preference for randomization—and promoting innovative implementation science (Kilbourne et al. 2017). To help plan, prioritize, and coordinate randomized program evaluations, in 2015, QUERI established the Partnered Evidence-based Policy Resource Center (PEPREc), which has also received support from the VHA's Health Services Research and Development (HSR&D) Service. In 2017, QUERI also established the Center for Access Policy, Evaluation and Research (CAPER). This article draws on our experience with RPEs facilitated by these two centers.

The VHA is just one of several government agencies and private organizations embracing learning health care system activities, marrying data, research, operations, and the interests of other stakeholders to facilitate a cycle of improvement and innovation. For example, a consortium of health systems, including Kaiser Permanente, with clinical data on about 15 million patients, created the HMO Research Network (now known as the Health Care Systems Research Network) to collaborate on comparative effectiveness research (Klein and Hostetter 2013), with funding through the NIH Collaboratory for pragmatic health systems trials (Weinfurt et al. 2017). UnitedHealth Group's Optum Labs has engaged health systems (including the Mayo Clinic), academic medical centers, and research organizations to facilitate the rapid translation of evidence from electronic health data on 150 million people into clinical practice (Wallace et al. 2014). The FDA's Sentinel Program works with

partnered health systems to conduct post-market drug safety surveillance (Findlay 2015). Efforts like these have gained the support of policy makers, with grant funding specifically earmarked for learning health system initiatives from the Patient-Centered Outcomes Research Institute, the National Institutes of Health, and the Agency for Healthcare Research and Quality (Wallace et al. 2014). To the extent participants in these efforts wish to implement randomized program evaluations, our study provides guidance about how researchers can navigate relationships with operations partners and balance scientific rigor with practical necessities.

## METHODS

As part of its Learning Health Care System Initiative, in 2015, VHA QUERI and HSR&D leadership solicited input from operational leaders for high-priority programs amenable to randomized implementation and evaluation. After receiving 36 responses, six finalists were sent to PEPRc for review and prioritization (Atkins, Kilbourne, and Shulkin 2017). These included proposals from the National Center for Health Promotion and Disease Prevention; Geriatrics and Extended Care (two proposals); Office of Mental Health and Suicide Prevention (two proposals); and Office of Connected Care.

PEPRc applied a pre-specified rubric in prioritizing proposed programs (see Appendix SA2). One criterion was that the supporting program office has adequate resources for program implementation because Learning Health Care System funds could be applied only to program evaluation, not to the intervention itself. This requirement also increased the chances of the sustainability of program intervention but did not guarantee it. Other criteria included the addressed problem's significance (e.g., number of veterans affected), the intervention's likely efficacy and impact, whether the program office and other stakeholders would support randomized roll-out, and the feasibility of matching the proposed intervention with researchers who had the subject matter expertise to evaluate it. PEPRc staff conducted interviews with program offices to receive input across these domains, and then independently scored each proposed program in each domain. Scores were tallied and averaged to guide prioritization.

After prioritization, QUERI and HSR&D leadership selected four programs. PEPRc worked with the corresponding program offices and teams of VHA researchers to finalize randomized intervention and evaluation designs. Below, we describe the four selected programs and lessons learned in

developing their randomized roll-out. To inform the lessons, we conducted interviews with selected collaborators associated with the four ongoing programs and evaluations. Interview questions were prepared in advance and varied by program to address issues specific to each. As this work was to inform internal VHA programs and policies, VA leadership considered it to be non-research (Department of Veterans Affairs 2011).

## RESULTS

### *Selected RPEs*

The four programs selected for randomized roll-out and evaluation are indicated in Table 1, along with design details and methods, and described more fully in the subsequent subsections. All programs have completed the design phase, but none have finished data collection or entered analysis.

*Veteran-Directed Home and Community Based Services.* In response to the long-term care rebalancing movement (Kaye 2012), in 2009 the VHA began offering participant-directed services to veterans—Veteran-Directed Home and Community Based Services (VD-HCBS)—at several medical centers. The program targets veterans with functional and cognitive limitations who are at risk of nursing home placement. Participants receive a monthly allotment to pay for personal care workers they select, medical equipment or supplies, or home modifications. The services are coordinated by Aging and Disability Network Agencies (ADNAs), which are overseen by the US Department of Health & Human Services' Administration for Community Living. The VHA's early experience with VD-HCBS suggests that it may increase the number of days a veteran remains safely at home while reducing health care costs (Mahoney and Kayala 2012). However, evaluation of this program has been limited by the lack of a control group. There was insufficient preliminary data to firmly establish an impact of VD-HCBS on nursing home placement, so there is equipoise on that question. Since insufficient funding existed to roll it out at all VHA medical centers simultaneously, a randomized approach was viewed by operations as equitable and ethical.

A new, randomized program evaluation will develop evidence of changes in health care use, days at home, and health care costs associated with VD-HCBS availability (see Garrido et al. 2017 and trial registries [isrctn.com/ISRCTN12228144](http://isrctn.com/ISRCTN12228144) and [clinicaltrials.gov/ct2/show/NCT03145818](http://clinicaltrials.gov/ct2/show/NCT03145818)). In 2016, PEPReC worked with the VHA's Office of Geriatrics and Extended Care, the

Table 1: Selected Programs for Randomized Evaluation

<i>Program</i>	<i>Program Office</i>	<i>Brief Description</i>	<i>Study Design</i>	<i>Type of Randomization</i>	<i>Expected Number of VA Sites (Medical Centers) and Patients</i>	<i>Study Duration</i>	<i>Primary Outcomes</i>
Veteran-Directed Home and Community Based Services	Geriatrics and Extended Care	Veterans at risk for nursing home placement receive resources and support to receive care at home	Stepped wedge cluster-randomized trial	Restricted*	77 sites, ~2,300 patients	3 years	Any hospitalization Any emergency department visit Any residential or post-acute nursing home admission Total health care costs per veteran per month alive Death or other serious adverse event
Risk Mitigation for Patients Receiving VA Opioid Prescriptions Targeting Care for Patients at High Risk for Suicide	Office of Mental Health and Suicide Prevention	VHA facilities review the cases of patients prescribed opioids and at high risk of adverse events Additional outreach and support provided to patients at high risk of suicide	Stepped wedge cluster-randomized trial Non-randomized stepped wedge	Simple None <sup>†</sup>	140 sites, ~100,000 patients 144 sites (28 sites at 7 Veterans Integrated Service Networks will receive virtual external facilitation)	1.5 years 2 years	Reach, adoption, and implementation fidelity of REACH-VET <sup>§</sup>
Mobile Tele dermatology	Office of Connected Care	Use of mobile apps to increase access to dermatology services	Stepped wedge cluster-randomized trial	Restricted <sup>‡</sup>	60 sites, ~16,000 patients for VA Telederm and ~70,000 for My Telederm	2 years	Consult completion and follow-up time, no show rates, and travel distance for dermatology services

\*Covariate constrained per Ivers et al. (2012).

<sup>†</sup>The randomized design was discontinued in response to changing priorities of VA leadership. Please see text for details.

<sup>‡</sup>Optimization-randomization per Bertsimas, Johnson, and Kallus (2015).

<sup>§</sup>See Landes (2016).

Administration for Community Living and an interdisciplinary team of researchers from the Center of Innovation in Long Term Services and Supports at the VA Providence Medical Center and the Center for Health Services Research in Primary Care at the Durham VA Medical Center to design a cluster-randomized, stepped wedge roll-out of VD-HCBS at 77 VHA medical centers. Over a 3-year period, program referral start dates will be staggered every 3–5 months. This roll-out began in March 2017.

*Risk Mitigation for Patients Receiving VA Opioid Prescriptions.* The VHA Office of Mental Health and Suicide Prevention (OMHSP) developed the Stratification Tool for Opioid Risk Mitigation (STORM) to prioritize review of patients receiving opioids based on their risk for overdose-, accident-, or suicide-related events and to inform providers of the risk factors and risk mitigation strategies. The STORM dashboard is a clinical decision support tool that uses the VHA electronic medical record to (1) estimate patient risk for these serious adverse events (SAEs) and (2) provide and track actionable information for risk-stratified intervention (such as reduced opioid dosage or naloxone kit provision) (Oliva et al. 2017).

The effect of STORM on SAE risk mitigation will be evaluated with a cluster-randomized, stepped wedge trial (see the trial registry at [isrctn.com/ISRCTN16012111](https://www.isrctn.com/ISRCTN16012111)). The VHA will release a policy memo that requires medical centers to review the cases of patients at high risk of opioid prescription-related SAEs, responding to the demands of the Comprehensive Addiction and Recovery Act of 2016. All medical centers will be required to review patients in the top 1 percent of predicted risk. Over 15 months, medical centers will be randomly staggered to expand their focus to the top 5 percent of predicted risk. This design permits the evaluation of the impact on patient outcomes when medical centers are required to review patients who fall under an expanded risk threshold.

PEPReC is collaborating with investigators with the Center for Health Equity Research and Promotion at the VA Pittsburgh and Philadelphia health care systems to evaluate STORM and the related policy roll-out. Outcomes include serious opioid-related adverse events, use of risk mitigation strategies (e.g., naloxone kits), and an assessment of facility and patient characteristics' impact on effectiveness.

*Targeting Care for Patients at High Risk for Suicide.* Suicide prevention is one of the VA's top priorities (Shulkin 2017). In an effort to identify veterans at increased risk of suicide, the VHA has instituted the Recovery Engagement

and Coordination for Health Veterans Enhanced Treatment (REACH VET) program. REACH VET includes a dashboard that ranks the risk of suicide for each VHA patient. Risk rankings are predicted by an algorithm that accounts for patients' health status, contextual factors (e.g., service-connected disability, homelessness), and socio-demographic factors (McCarthy et al. 2015). Beginning in February 2017, the use of REACH VET to identify veterans in the top 0.1 percent of risk at each VHA medical center is mandated. On-site coordinators assist providers with dashboard use, patient identification, and re-evaluation of treatment plans.

One way to address suicide risk is via a series of eight "caring letters" sent over 1 year. Outside of the VHA, similar brief contact interventions have reduced rates of suicide attempts (Milner et al. 2015). However, less is known about the effect of brief contact interventions among veterans who are engaged in mental health care and among those who are identified as being at elevated risk for an initial or repeat suicide attempt. Additionally, clinicians within the VHA are developing a more intensive way to cover these topics—a telephone coaching intervention that includes three to four sessions over a 6-week period. In 2016, PEPRc worked with the VHA's OMHSP, VHA clinicians with expertise in treating patients at risk of suicide, and researchers from the Center for Mental Healthcare and Outcomes Research (CeMHOR) at the VA Little Rock Medical Center to design an evaluation of the effectiveness of caring letters among patients in the top 5 percent of risk and of telecoaching among patients in the top 0.1 percent of risk. Outcomes of interest include suicide attempts, hospitalization, outpatient mental health care use, mortality (suicide and all cause), and costs. The original evaluation design called for a cluster-randomized stepped wedge trial of caring letters in 84 medical centers over 3 years. Within participating sites, a random sample of patients in the top 0.1 percent of risk would receive telecoaching.

Due to increased focus on suicide prevention and to budgetary concerns, VHA leadership imposed a change in the evaluation design (Office of the Inspector General, Department of Veterans Affairs 2017). All sites are required to perform reviews of all patients in the top 0.1 percent of predicted suicide risk. Then, if warranted, providers contact patients with efforts to mitigate that risk. This change also reflected a concern about having sufficient capacity to do something for all patients identified as high risk. Operations partners implementing the initiative believed it was essential that high acute risk would be contacted by suicide prevention coordinators (standard practice), but that there was equipoise regarding the best way to manage that risk. Using an adaptive design, sites not meeting targets for outreach to high-risk



patients have been assigned to staggered times at which they will receive additional facilitated support to enhance program uptake by addressing provider and site-level barriers.

*Mobile Tele dermatology.* The VHA Office of Connected Care oversees a store-and-forward tele dermatology program intended to improve access to dermatological services by forwarding images from outlying clinics for review by a dermatologist at a VHA medical center. However, it is perceived as burdensome and inefficient. Consequently, the VHA Office of Connected Care has developed two mobile tele dermatology applications (apps)—one for providers, another for patients—with the aim of improving VHA patients’ access to dermatological care, particularly for veterans in rural areas.

One app, *VA Telederm*, is designed to simplify and expedite the process of primary care providers’ requests for tele dermatology consultation. The app will facilitate the image acquisition process and will be integrated with VHA’s clinical workflow, allowing the images to be uploaded into the patients’ electronic health record for review by the dermatology consultant. A second app, *My Telederm*, offers eligible patients an alternative to in-person follow-up care. Selected dermatology patients can use *My Telederm* to follow-up remotely, using their own mobile device to submit interval history and skin images.

Both apps will be made available across a subset of VHA medical centers over a 2-year period in a randomized, stepped wedge design. In all, 36 medical centers were randomized to receive the *VA Telederm* app and 24 medical centers were randomized to receive the *My Telederm* app. Medical centers were selected for randomization to only receive one app or the other but not both (see Done et al. 2018 and the trial registry at [clinicaltrials.gov/ct2/show/NCT03241589?term=NCT03241589&rank=1](https://clinicaltrials.gov/ct2/show/NCT03241589?term=NCT03241589&rank=1)). Outcomes of interest include time to completed consultations and number of in-person dermatology visits.

### *Lessons Learned*

In this section, we discuss challenges we encountered in designing and initiating the randomized evaluations described above and how we addressed them. To preserve anonymity, in many cases, we have concealed the specific project(s) to which each lesson pertains.

*Complex Research Design Challenges.* In our work, two challenges with stepped wedge designs arose. First, in the VD-HCBS evaluation, we could not establish the roll-out schedule for all sites up front. Randomization could only occur for sites with buy-in from local leadership, that had available funding to hire a local coordinator to oversee the program, and that had gone through a VHA-ACL interagency preparatory process that lasted several months. The initiation and completion of this process were outside our control. For these reasons, randomization was to occur every 6–10 months and limited to sites that were ready for the intervention. With a small number of sites available for randomization in each wave, we were concerned that simple randomization (assigning treatment/control by coin flip) might lead to imbalance across study arms. To address this, we used covariate constrained randomization (Ivers et al. 2012), in which we ranked every possible allocation of sites within each wave to start times according to their ability to balance observed site-level confounders across earlier and later start times. We randomly selected one allocation from the top quartile of potential combinations.

In addition, wave-specific randomization may give rise to heterogeneity across sites in different waves. Sites that are ready to roll-out the program in year 1 may differ in meaningful ways from sites that are not ready to roll-out the program until year 3, diminishing the benefits of randomization. If we find this to be the case, we will need to treat our study as if it had an observational design.

The Mobile Tele dermatology trial also implemented a constrained randomization procedure to avoid imbalance across the stepped wedge trial steps. By adapting an optimization-randomization approach proposed by Bertsimas, Johnson, and Kallus (2015), we first allocated study units (facilities) to groups such that the difference between the groups was minimized. Random assignment of the order in which the groups will receive the intervention was then followed.

A second issue that can arise with stepped wedge is keeping on schedule. Each phase of the roll-out is another opportunity for delay. A consequence is that steps may be of unequal length. If earlier steps are delayed, but the overall trial still needs to occur within a fixed time period, as is the case with VD-HCBS and STORM, later steps will become compressed. Compression of later steps may make it more difficult to detect the treatment effect of interest; there may not be sufficient time for sites to fully implement a program (VD-HCBS) or respond to a policy intervention (STORM) before measurement needs to occur. Delays in the roll-out can also cause delays in planned implementation evaluations. In the Mobile Tele dermatology trial, delays in app

development were handled by preserving the length of each step planned initially (to allow the sites enough time for implementation) but increasing the number of sites randomized to each step.

*Fragile Financing.* A necessary condition for a randomized program evaluation is sufficient funding for both the program and effort to evaluate it. Although HSR&D supported the cost of designing and planning the randomized roll-outs, implementing programs over multiple years can add costs and uncertainty. In a large organization with shifting priorities, sustained operational funding for new initiatives is not always guaranteed.

One solution is to inquire early about funding commitments for operations and research before deeply engaging in design work. Of course, even when there is strong early interest in providing it, that commitment can weaken if other priorities become salient. Another approach is to develop contingency plans—smaller or shorter studies—when funding is not adequate for the optimal design, but this will sacrifice statistical power.

Conversely, the design can sometimes influence the scope of the planned program in cases where the budget and scale of the program are not yet fixed. We encountered this chicken-and-egg problem on one of our projects for which a programmatic budget request was driven by the evaluation design. A more ambitious design (e.g., with higher recruitment numbers) would require more resources and a greater budget request. However, exactly what request would be approved was uncertain in advance. Therefore, so was the design.

There are limits to what one may be able to do with smaller and cheaper evaluations. “Planning a randomized program evaluation requires a level of flexibility that can be difficult to achieve while maintaining scientific rigor,” one of our collaborators said. “When funding is decreased and the number of sites is impacted, there may be less ability to randomize.”

Ultimately, research planners should expect some randomized initiatives to fail due to shifting priorities that threaten budgetary support.

*Premature or Overly Expansive Roll-out.* Intervention roll-out can get ahead of the randomized design, threatening the study. In a stepped wedge design, for example, some sites may want to initiate an intervention before their assigned start date, and it can be difficult to hold them back. This is akin to cross-over in a standard RCT and can weaken power. Another variant is when national (or cross-site) leadership favors a broader and more rapid implementation. If the intervention is offered to all sites and

patients at once as a matter of national policy, that can completely disrupt randomized evaluation.

Communication is the key to head off problems like this, as one of our research partners told us. If staff are sure that the intervention will be beneficial, “then we would not want to stop [a full, immediate] roll-out,” he added. “But if there is some uncertainty, then an RCT is a rigorous way to figure out what works and what doesn’t.” Though accommodating an RCT takes time, it can ultimately help defend a program that works or support withdrawing resources from one that does not. Communicating that should include advocacy up the chain in slowing roll-out to accommodate rigorous evaluation.

*Insufficient Buy-in at All Levels.* Few programs are implemented solely by organizational leaders. Therefore, even with strong support from the top, a program can fail to be implemented as designed without buy-in and cooperation from subordinates. Unfortunately, it is not easy to tell in advance when boots-on-the-ground staff are not invested in a randomized design.

However, even if commitment cannot be assessed before implementation, it may be revealed early enough in the process to allow for corrections. Regular communication with key implementation staff and leadership, with requests for feedback on preliminary steps, can help identify problems. Pay close attention to what is not going according to plan and where in the chain things are going wrong, and try to identify who the problematic actors are. Including them on calls together with leadership can help deliver a unified message of the importance of the design.

“Many programmatic decisions are made at the local level—not the national level,” said one of our research collaborators. “As a result, building a network of engaged local leaders is a critical step to ensuring the development and initiation of the program and the measurement of the implementation as well.”

*Discomfort with Randomizing.* A good design can be threatened if staff or leadership is not comfortable with randomization. One source of discomfort is that not all units will get the intervention, or at least not right away. However, when there is uncertainty about the effectiveness of an intervention, “randomization is often viewed as a fair way of allotting the initial spots,” one of our collaborators said.

Stepped wedge designs can help because they plan for full, if staggered, roll-out. However, they require orchestration of a gradual roll-out—which

necessitates engagement with program implementers over a long period of time. Gradual roll-out also opens up the possibility that some organizations hire staff too early, who are then potentially idle while waiting to enter the treatment group. This can arise if funding and hiring cannot be timed to coincide with the design. This tension between when an organization is ready for implementation and when it is scheduled to receive it in a randomized roll-out can also raise discomfort with the design.

“Implementation [in the VHA] has traditionally been all at once in response to a directive. Facilities are not accustomed to being told to wait before implementing,” said one of our research collaborators. “There is a lead time for local programs to build a case for a program at a medical center. Then, the local program wants to act once approval is gained.”

The upshot is that designs that require significant local investments in staff or systems may be harder to randomize. A potential solution, where possible, is to seek centralized staffing. Another is to focus on interventions in which randomization can be controlled centrally. In one of our projects, resources to implement the required technology are delivered only to sites randomized to treatment.

*Multiple Veto Points.* In a large organization like the VHA, system change is implemented with new policy or technology. These often require concurrence across multiple leaders and offices, each of whom can take significant time for deliberations and also has a veto. The initiation of two of our randomized programs was delayed for several months precisely for this reason.

A collaborator on one effort told us, “A policy intervention requires the development of the policy itself. When that pertains to a large organization, it can require a lengthy, deliberative process of obtaining approval from multiple organizational leaders.”

It is important to note that, though it can make planning difficult and threaten the evaluation, delay is not all bad. The concurrence process can help achieve buy-in from stakeholders, which can mitigate potential future threats to the program. Still, intervention sponsors need a strategy for obtaining all required concurrences. This will often take support from top leadership, coupled with a working coalition of program offices to push policy or technology through the process.

## DISCUSSION AND CONCLUSION

By fostering multidisciplinary operations and research partnerships in large, randomized programs, the VHA has emerged as a leader, along with several other organizations (Greene, Reid, and Larson 2012), in implementing learning health care system concepts. We are directly involved in four such efforts in high-priority areas—home- and community-based care, opioid prescribing risk mitigation, suicide prevention, and telehealth. In developing these randomized program evaluations, we encountered and attempted to overcome a host of issues likely common to similar efforts. We found that successful development of these collaborations required more than attention to traditional design issues, such as ensuring adequate power and availability of data. It also required the stability of intervention financing, means of controlling and commitment to adhering to randomized roll-out, degree of buy-in from key implementation staff, and feasibility of managing multiple veto points for interventions that span several programs.

Our study has several limitations. Chief among them is that we could only assess (and attempt to overcome) challenges that arose during the design phase of the randomized programs discussed. It is possible that other challenges will arise in subsequent phases of implementation and evaluation that we cannot anticipate at this stage. A second limitation is that we only considered the randomized program evaluations with which we are intimately familiar—the several that are being conducted by PEPReC and CAPER in the VHA. Other evaluations in other settings may encounter different sets of issues.

Indeed, others have successfully melded randomization with health system operational requirements and serve as models for future evaluations. Classic examples of evaluations facilitated by person-level randomization include the RAND Health Insurance Experiment (Manning et al. 1987) and the Oregon Health Insurance Experiment (Baicker et al. 2013). Within the VHA, the Serious Mental Illness (SMI) Re-Engage study combined site-level randomization with an adaptive design, in which the implementation intervention provided to sites was adjusted in response to site outcomes (Kilbourne et al. 2013). Other interventions across a diverse range of subject areas have been evaluated with cluster-randomized, stepped wedge designs (Medge et al. 2011). Finally, the National Institutes of Health (NIH) Health Care Systems Collaboratory has helped integrate rigorous evaluation with real-world clinical settings by fostering pragmatic trials (Johnson et al. 2016; Simon et al. 2016; Mor et al. 2017).

Other approaches help facilitate randomization. Where possible and permitted by institutional review boards, site randomization, waiver of informed consent, an intent-to-treat design, and use of routinely collected clinical or administrative data can obviate some of the costly elements of RCTs. Administrative data offer a variety of other advantages: they are less likely to suffer non-response bias (though can suffer ascertainment and selection biases) and are often available over long time spans (Finkelstein and Taubman 2015b). As more health information becomes electronically available, opportunities for more rapid, low-cost, randomized trials of system interventions will expand (Saleem et al. 2016; Choudhry 2017).

Evidence-based policy is a worthy goal, but, as we have learned, it takes more than rhetoric. Sustainability of the learning health system model requires institutional commitment and funding for implementation, as well as dissemination of the lessons learned from each effort. Recognizing the complex challenges involved in developing evidence-based policy, the VHA has institutionalized its commitment to a learning health care system in the form of the Partnered Evidence-based Policy Resource Center. This is where lessons learned can be maintained, along with more obvious needs like data access and evaluation design expertise.

## ACKNOWLEDGMENTS

*Joint Acknowledgment/Disclosure Statement:* We thank Joseph Doyle, Walid Gellad, Sara Landes, Nicolae Done, and James Rudolph for contributions to early drafts. This work has been supported by HSR&D grant SDR 16-196 and CDA 11-201/CDP 12-255 and QUERI grants PEC 16-001 and PEC 15-167. The contents do not represent the views of the US Department of Veterans Affairs or the United States Government and were deemed to be the product of non-research operations activities under Veterans Health Administration Handbook 1058.05.

*Disclosure:* None.

*Disclaimer:* None.

## REFERENCES

- Atkins, D., A. M. Kilbourne, and D. Shulkin. 2017. "Moving from Discovery to System-Wide Change: The Role of Research in a Learning Health Care System:

- Experience from Three Decades of Health Systems Research in the Veterans Health Administration." *Annual Review of Public Health* 20 (38): 467–87.
- Baicker, K., S. L. Taubman, H. L. Allen, M. Bernstein, J. H. Gruber, J. P. Newhouse, E. C. Schneider, B. J. Wright, A. M. Zaslavsky, and A. N. Finkelstein. 2013. "The Oregon Experiment—Effects of Medicaid on Clinical Outcomes." *New England Journal of Medicine* 368 (18): 1713–22.
- Bertsimas, D., M. Johnson, and N. Kallus. 2015. "The Power of Optimization over Randomization in Designing Experiments Involving Small Samples." *Operations Research* 63 (4): 868–76.
- Bothwell, L. E., J. A. Greene, S. H. Podolsky, and D. S. Jones. 2016. "Assessing the Gold Standard—Lessons from the History of RCTs." *New England Journal of Medicine* 374 (22): 2175–81.
- Bridgeland, J., and P. Orszag. "Can Government Play Moneyball?" *Atlantic*. July/August 2013 [accessed on September 20, 2017]. Available at <http://www.theatlantic.com/magazine/archive/2013/07/can-government-play-moneyball/309389/>
- Choudhry, N. K. 2017. "Randomized, Controlled Trials in Health Insurance Systems." *New England Journal of Medicine* 377: 957–64.
- Department of Veterans Affairs. 2011. "VHA Operations Activities That May Constitute Research." VHA Handbook 1058.05. October 28.
- Done, N., D. H. Oh, M. Weinstock, J. D. Whited, G. L. Jackson, H. A. King, S. B. Peracca, A. R. Elwy, and J. C. Prentice. 2018. "The VA Telederm Study: Protocol for a Stepped-Wedge Cluster Randomized Trial to Compare Access to Care for a Mobile App Versus a Workstation-Based Store-and-Forward Teledermatology Process." *British Medical Journal Open*, under review.
- Etheredge, L. M. 2007. "A Rapid-Learning Health System." *Health Affairs* 26 (2): w107–18.
- Findlay, S. 2015. "The FDA's Sentinel Initiative." *Health Affairs Health Policy Brief* [accessed on December 7, 2017]. Available at <https://www.healthaffairs.org/action/showDoPubSecure?doi=10.1377%2Fhpb20150604.936915&format=full>
- Finkelstein, A., and S. Taubman. 2015a. *Using Randomized Evaluations to Improve the Efficiency of U.S. Healthcare Delivery*. Cambridge, MA: J-PAL North America.
- . 2015b. "Randomize Evaluations to Improve Health Care Delivery." *Science* 347 (6223): 720–2.
- Garrido, M. M., R. M. Allman, S. D. Pizer, J. L. Rudolph, K. S. Thomas, N. R. Sperber, C. H. van Houtven, and A. B. Frakt. 2017. "Innovation in a Learning Health Care System: Veteran-Directed Home- and Community-Based Services." *Journal of the American Geriatrics Society* 65 (11): 2446–51.
- Greene, S. M., R. J. Reid, and E. B. Larson. 2012. "Implementing the Learning Health System: From Concept to Action." *Annals of Internal Medicine* 157 (3): 207–10.
- Ivers, N. M., I. J. Halperin, J. Barnsley, J. M. Grimshaw, B. R. Shah, K. Tu, R. Upshur, and M. Zwarenstein. 2012. "Allocation Techniques for Balance at Baseline in Cluster Randomized Trials: A Methodological Review." *Trials* 13: 120.
- Johnson, K. E., G. Neta, L. M. Dember, G. D. Coronado, J. Suls, D. A. Chambers, S. Rundell, D. H. Smith, B. Liu, S. Taplin, and C. M. Stoney. 2016. "Use of



- PRECIS Ratings in the National Institutes of Health (NIH) Health Care Systems Research Collaboratory.” *Trials* 17 (1): 32.
- Kaye, H. S. 2012. “Gradual Rebalancing of Medicaid Long-Term Services and Supports Saves Money and Serves More People, Statistical Model Shows.” *Health Affairs* 31 (6): 1195–203.
- Kilbourne, A. M., K. M. Abraham, D. E. Goodrich, N. W. Bowersox, D. Almirall, Z. Lai, and K. M. Nord. 2013. “Cluster Randomized Adaptive Implementation Trial Comparing a Standard Versus Enhanced Implementation Intervention to Improve Uptake of an Effective Re-Engagement Program for Patients with Serious Mental Illness.” *Implementation Science* 8: 136.
- Kilbourne, A. M., A. R. Elwy, A. E. Sales, and D. Atkins. 2017. “Accelerating Research Impact in a Learning Health Care System: VA’s Quality Enhancement Research Initiative in the Choice Act Era.” *Medical Care* 55(7 Suppl. 1): S4–12.
- Klein, S., and M. Hostetter. 2013. “In Focus: Learning Health Care Systems.” The Commonwealth Fund [accessed on December 7, 2017]. Available at <http://www.commonwealthfund.org/publications/newsletters/quality-matters/2013/august-september/in-focus-learning-health-care-systems>
- Landes, S. J. 2016. “Risk Stratified Enhancements to Clinical Care: Targeting Care for Patients Identified Through Predictive Modeling as Being at High Risk for Suicide, with the Office of Mental Health Operations.” VA HSR&D Study SDR 16-195 [accessed on January 30, 2018]. Available at [https://www.hsrdr.research.va.gov/research/abstracts.cfm?Project\\_ID=2141704560](https://www.hsrdr.research.va.gov/research/abstracts.cfm?Project_ID=2141704560)
- Mahoney, E., and D. Kayala. 2012. “Veteran-Directed Home and Community Based Services: A Program Evaluation [accessed on September 20, 2017]. Available at <http://www.appliedselfdirection.com/file/446/download?token=IMY2DJyu>
- Manning, W. G., J. P. Newhouse, N. Duan, E. B. Keeler, A. Leibowitz, and M. S. Marquis. 1987. “Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment.” *American Economic Review* 77 (3): 251–77.
- McCarthy, J. F., R. M. Bossarte, I. R. Katz, C. Thompson, J. Kemp, C. M. Hanneman, C. Nielson, and M. Schoenbaum. 2015. “Predictive Modeling and Concentration of the Risk of Suicide: Implications for Preventive Interventions in the U.S. Department of Veterans Affairs.” *American Journal of Public Health* 105: 1935–42.
- Medge, N. D., M. S. Man, C. A. Taylor, and D. J. Torgerson. 2011. “Systematic Review of Stepped Wedge Cluster Randomized Trials Shows That Design Is Particularly Used to Evaluate Interventions during Routine Implementation.” *Journal of Clinical Epidemiology* 64 (9): 936–48.
- Milner, A. J., G. Carter, J. Pirkis, J. Robinson, and M. J. Spittal. 2015. “Letters, Green Cards, Telephone Calls and Postcards: Systematic and Meta-Analytic Review of Brief Contact Interventions for Reducing Self-Harm, Suicide Attempts and Suicide.” *British Journal of Psychiatry* 206 (3): 184–90.
- Mor, V., A. E. Volandes, R. Gutman, C. Gatsonis, and S. L. Mitchell. 2017. “PRagmatic Trial Of Video Education in Nursing Homes: The Design and Rationale for a Pragmatic Cluster Randomized Trial in the Nursing Home Setting.” *Clinical Trials* 14 (2): 140–51.

- National Academy of Medicine (formerly Institute of Medicine). 2012. "Best Care at Lower Cost: The Path to Continuously Learning Health Care in America" [accessed on September 20, 2017]. Available at <https://www.nap.edu/catalog/13444/best-care-at-lower-cost-the-path-to-continuously-learning>
- Newhouse, J. P., and S. T. Normand. 2017. "Health Policy Trials." *New England Journal of Medicine* 376: 2160–7.
- Office of the Inspector General, Department of Veterans Affairs. 2017. "Overview of VA Suicide Prevention Efforts and Data Collection," Report No. 16-00349-369. Washington, DC [accessed on January 12, 2017]. Available at <https://www.va.gov/oig/pubs/VAOIG-16-00349-369.pdf>
- Oliva, E. M., T. Bowe, S. Tavakoli, S. Martins, E. T. Lewis, M. Paik, I. Wiechers, P. Henderson, M. Harvey, T. Avoundjian, A. Medhanie, and J. A. Trafton. 2017. "Development and Applications of the Veterans Health Administration's Stratification Tool for Opioid Risk Mitigation (STORM) to Improve Opioid Safety and Prevent Overdose and Suicide." *Psychological Services* 14 (1): 34–49.
- Saleem, J. J., L. G. Militello, A. L. Russ, and N. R. Wilck. 2016. "The Need for Better Integration Between Applied Research and Operations to Advance Health Information Technology." *Healthcare (Amsterdam)* 4 (2): 80–3.
- Shulkin, D. J. 2017. "Statement for Presentation Before the House Committee on Appropriations, Subcommittee on Military Construction, Veterans Affairs, and Related Agencies." May 3 [accessed on September 20, 2017]. Available at <http://docs.house.gov/meetings/AP/AP18/20170503/105899/HHRG-115-AP18-Wstate-ShulkinD-20170503.pdf>
- Simon, G. E., A. Beck, R. Rossom, J. Richards, B. Kirilin, D. King, L. Shulman, E. J. Ludman, R. Penfold, S. M. Shortreed, and U. Whiteside. 2016. "Population-Based Outreach Versus Care as Usual to Prevent Suicide Attempt: Study Protocol for a Randomized Controlled Trial." *Trials* 17 (1): 452.
- U.S. House of Representatives, 113th Congress. "H.R. 1287 – Sound Science Act of 2013" [accessed on August 17, 2015]. Available at <https://www.congress.gov/bill/113th-congress/house-bill/1287>
- U.S. House of Representatives, 114th Congress. "H.R. 1381 – Evidence-Based Policy-making Commission Act of 2016. Public Law No: 114-140. 2016" [accessed on December 7, 2016]. Available at <https://www.congress.gov/bill/114th-congress/house-bill/1381>
- U.S. House of Representatives, 115th Congress. "H.R. 5 – Regulatory Accountability Act of 2017" [accessed on May 30, 2017]. Available at <https://www.congress.gov/bill/115th-congress/house-bill/5>
- Wallace, P. J., N. D. Shah, T. Dennen, P. A. Bleicher, and W. H. Crown. 2014. "Optum Labs: Building a Novel Node in the Learning Health Care System." *Health Affairs* 33 (7): 1187–94.
- Weinfurt, K. P., A. F. Hernandez, G. D. Coronado, L. L. DeBar, L. M. Dember, B. B. Green, P. J. Heagerty, S. S. Huang, K. T. James, J. G. Jarvik, and E. B. Larson. 2017. "Pragmatic Clinical Trials Embedded in Healthcare Systems: Generalizable Lessons from the NIH Collaboratory." *BMC Medical Research Methodology* 17 (1): 144.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix SA1: Author Matrix.

Appendix SA2: Candidate Program Evaluation Rubric.