

Evidence-Based Policy Making: Balancing Rigor With Real-World Health Care for Veterans and Military Personnel

Amy Kilbourne, David Atkins

The US health care system is undergoing unprecedented policy transformations that will impact veterans and returning service members. Rigorous program evaluations will be crucial as policy makers make tough decisions regarding the value of new policies. Also, more partnerships between scientists and policy makers will be required to implement these designs in light of trade-offs between scientific rigor and political realities.

Less than \$1 out of every \$1,000 that the government spends on health care this year will go toward evaluating whether the other \$999-plus actually works [1].

Military service has a long tradition in North Carolina. Historically, the state sent more soldiers to the Civil War than any other Confederate state, and now, with nearly 800,000 veterans, North Carolina ranks 8th among the states with the largest veteran populations [2]. Health care for veterans has been frequently in the public eye, with concerns about the recent problems at the US Department of Veterans Affairs (VA), combined with recognition of the special needs of recent veterans, including post-traumatic stress disorder, other mental health issues, traumatic brain injury, and the challenges of chronic pain and opiate addiction [3]. In North Carolina, about 30% of adult men are military service veterans, many of whom have substantial mental or physical health problems [4].

Health care for veterans is also now at the intersection of several important policy changes. These include changes under the Patient Protection and Affordable Care Act of 2010 (ACA), which includes Medicaid expansion in some states, as well as recent changes to VA care including the Veterans Access, Choice, and Accountability Act (Choice Act) of 2014 [5]. The latter allows veterans waiting for longer than 30 days or living further than 40 miles from a VA facility to use non-VA providers, and it makes VA a payer as well as a provider of health care for veterans. Together, Medicaid and VA represent the largest providers of health care services in the United States for low-income individuals and/or those with chronic mental or physical health conditions. Many veterans will seek VA care, but some will also be eligible for Medicaid under the ACA.

Policy changes are often controversial, both because they usually involve important trade-offs (eg, access versus cost, choice versus efficiency) and because new policies usually reflect the specific philosophy of those who write them. Another reason policy changes are often so heatedly debated is that, unfortunately, there is rarely strong evidence in support of the changes being proposed. In clinical medicine and among organizations that promote best practices, such as VA's Evidence-based Synthesis Program, careful randomized trials are the standard for demonstrating benefits. Similarly, regulatory bodies such as the US Food and Drug Administration determine when sufficient proof of safety and effectiveness is available to market a drug or device. In contrast, policies are often promoted based on minimal evidence, and the evidence that is available is often weak.

Evidence-Based Policy

To this end, there has been an increased call to implement programs or policies that have a proven, evidence-based track record [6, 7]. Thus, there has been a greater emphasis on the use of stronger evaluation designs—particularly randomized program evaluations, or at least ones with parallel comparison groups—in order to determine whether new policies or programs are effective before they are fully implemented. This approach to data-driven decision making has been referred to as evidence-based policy [1]; in essence, this is program implementation informed by rigorously established objective evidence, and it is an extension of the concept of evidence-based medicine. Recent bills introduced in Congress [8, 9], including the Regulatory Accountability Act and the Sound Science Act, apply concepts of evidence-based policy as outlined by the US Office of Management and Budget and other agencies.

Evidence-based policy studies, especially those that employ randomized designs, can inform state and federal

Electronically published November 13, 2015.

Address correspondence to Dr. Amy Kilbourne, Health Services Research and Development, US Department of Veterans Affairs, 810 N Vermont Ave, Washington, DC 20420 (Amy.Kilbourne@va.gov).

N C Med J. 2015;76(5):339-342. ©2015 by the North Carolina Institute of Medicine and The Duke Endowment. All rights reserved.
0029-2559/2015/76517

policy makers regarding where to invest resources, thus preventing wasted effort and expense on ineffective programs or policies, and ultimately producing a greater return on investment. As with evidence-based medicine, strong evidence is derived from sound study designs, notably randomized controlled trials [10, 11].

From a program or policy perspective, a key advantage of randomized evaluations is that they offer patients equitable access to new programs or policies, via a lottery or similar system, when there is insufficient funding for total implementation of a new program upfront [12]. Randomization also provides the best opportunity to determine the effect of a new program or policy compared to a control group, since it reduces the potential influence of variations across sites, patients, and providers. When planned in advance, randomization does not add significant cost to the study, especially if program or policy leaders can control allocation of new initiatives to regions, sites, or individuals [11]. There is also growing criticism that decisions regarding new federal funding rely on observational or quasi-experimental designs that do not account for historical trends or the effects of individual or site self-selection (eg, those most enthusiastic about the program or policy are more likely to use it first) [13]. Still, only 18% of studies of health care programs used randomized designs in the period 2009–2013, compared to 36% of education studies and 49% of international development studies [14].

Program Evaluation: Strong Science Versus Political Realities

Increasing the number of rigorous program evaluations takes concerted effort by leaders, policy makers, and scientific investigators who must plan the study design in advance as well as communicate and garner input from the parties involved (eg, consumers and providers) about the benefits of randomization (eg, ensuring equity). Having investigators involved upfront in study design decisions is a crucial step in order for designs to be used effectively.

Different design options for rigorous program or policy evaluation are presented in Table 1. One recent example of a statewide, randomized program evaluation was the Oregon Health Insurance Experiment [12, 14, 15], in which state officials chose to expand Medicaid coverage in Oregon using a lottery system, given insufficient resources to enroll all eligible individuals right away. State officials were able to involve researchers in the evaluation early enough to implement longitudinal assessments, which allowed them to compare key health and economic indicators for individuals randomized to expanded Medicaid versus those who did not receive insurance.

From the perspective of researchers, key challenges included the need to react to swift time lines created by state officials rolling out Medicaid expansion, to secure funding and institutional review board approvals in time, and to accommodate further expansion during the experiment

when the lottery was reopened to those in the control group [12]. After overcoming these challenges, investigators found key differences among those receiving Medicaid [14, 15] that challenged many assumptions in regards to the benefits of insurance expansion, while also showing unexpected improvements in economic indicators. State officials were credited for their willingness to participate in this randomization, which occurred because of their prior experience with strong partnerships between researchers and operational/policy leadership [12].

In many situations, there will not be the political will or feasibility to randomize new policies or programs. For new programs at a facility or site level, for example, policy makers may want to offer the intervention to everyone because of pressure to quickly respond to a public health or services need. However, rolling out the intervention to all sites simultaneously may not be feasible, especially if there are complex logistics involved in its start-up and scale-up. In this case, rollout could be staggered, which would provide an opportunity to collect data on both an intervention group and a control group. This type of approach, called a stepped wedge design, compares random and sequential crossover of groups from control to intervention until all groups are exposed.

A key example of a stepped wedge design used to evaluate a state-level program was the Depression Improvement Across Minnesota, Offering a New Direction (DIAMOND) initiative [16]. Under DIAMOND, primary care providers who provided care management for depression were offered a reimbursement from state insurance companies in the form of a bundled payment. The payment covered costs of care managers as well as supervision time from mental health specialists. State and insurance officials agreed to implement the reimbursement in staggered sequences, which allowed greater feasibility in training and orienting sites to the new reimbursement policy. The use of repeated outcome measures over time also enabled researchers to control for historical trends as well as to provide for a comparison group. Still, randomized trials may be preferable to stepped wedge designs for policy or program interventions because the actual time to conduct stepped wedge designs ends up being greater than that needed for a traditional randomized program trial, due to the number of randomization points. In addition, stepped wedge designs may require more outcomes data collection points, which adds to respondent burden and introduces the potential for contamination across intervention arms [17].

In most situations, however, randomization or stepped wedge approaches may not be an option, since the policy or program has already been implemented, or the desired outcome might be too rare to warrant randomization (eg, cause-specific mortality). Instead, an observational study design that allows for a more generalizable population-based study is often the best option [18]. In these situations, well designed observational studies are desirable, especially

those that apply selection effect adjusters such as propensity scoring or instrumental variables [19]. At the same time, observational designs may still have lingering selection effects when comparing exposure to different programs or policies, and analyses may not be able to control for these effects. In the Oregon Health Insurance Experiment, when data were reanalyzed using observational methods, investigators found results that were opposite to the original findings reported from the randomized design [14].

Evaluations After the Fact: Value-Added

Alternatively, when the policy or program has already been implemented, the question for policy makers may focus on how to improve its implementation and sustainability. In these situations, randomized trials of specific strategies [20] to improve the uptake of programs or policies could be used to inform improvements [21]. A key example is the Community Partners in Care study in greater Los Angeles, which randomized sites to either an enhanced implementation strategy involving community engagement or to standard dissemination of a depression care management program [22]. This choice in study design was based on the desire to provide all community-based sites with the opportunity to implement depression care management, which was already shown to be evidence-based; the policy question was what it would take to improve its implementation and ultimate sustainability. Another option would be to

employ adaptive designs or sequential multiple assignment randomized trial (SMART) designs [23], in which a program or policy is augmented in direct response to its limited adoption among specific sites. Adaptive designs are more efficient to implement because they focus on sites in need of intervention rather than ones already implementing the policy or program.

Even without the ability to randomize new programs or policies, evaluations can provide important information and support towards understanding a program or policy's long-term implementation and value. Notably, deep-dive systematic or formative evaluation approaches that assess whether an initiative delivered what was intended are important because they help inform sustainability. It not only asks, "Does the program work?" but also asks, "What makes it work?"

For example, in response to the Choice Act, the Office of Management and Budget asked VA to support rigorous internal evaluations of its implementation. While the requirements of the Choice Act made it impossible to randomize who would get access to non-VA care, VA's Quality Enhancement Research Initiative funded 7 projects that are taking a deep-dive observational look at the impact of the Choice Act on several factors: patient outcomes among veterans with mental health or chronic conditions, access to women's health care, access to pain management services, and wait times [24]. These projects are informing future

TABLE 1.
Examples of Evaluation Study Designs for New Programs or Policies

Design options	Randomized-effectiveness	Cluster randomized-effectiveness	Cluster randomized-implementation	Rollout: stepped wedge	Adaptive or SMART design	Observational
Typical unit of recruitment, intervention, and analysis	Individual	Provider or site	Provider or site	Site	Individual, provider, or site	Individual, provider, or site
Description	Test program or policy on effectiveness compared to control/usual care/alternative intervention	Compare program or policy (versus no program/policy) randomized among groups of individuals (eg, by providers, sites)	Compare implementation or rollout strategy of program or policy (program/policy without strategy) randomized among groups of individuals (eg, by providers, sites)	Compare random and sequential crossover of groups from control to intervention until all groups are exposed	Those not responding to initial program/policy randomized to receive augmented or different version	Compare program or policy across non-randomized groups
Advantages	Equity in chance of receiving new program; balancing of individual characteristics	Ideal for site-based interventions; reduced potential contamination across sites	Ideal when program/policy must roll out to all subjects; evaluates sustainability strategies	Staggered program or policy rollout precludes one-time randomization	Efficient distribution of new program or policy to non-responders	Allows larger study needed for rare outcomes
Example	Oregon Health Insurance Experiment: http://www.nber.org/oregon/	Head Start Impact Study: http://www.acf.hhs.gov/sites/default/files/opre/executive_summary_final.pdf	Community Partners in Care: http://www.communitypartnersincare.org/	DIAMOND: https://www.icsi.org/health_initiatives/mental_health/diamond_for_depression/	VA SMI Re-Engage study: http://www.science.com/content/8/1/136	Value in Health Study: http://www.valueinhealthjournal.com/article/S1098-3015(14)004638-5/abstract

Note. SMART, Sequential Multiple Assessment Randomized Trial.

evaluations by improving the collection of crucial data on use of non-VA care as well as helping with the development and validation of metrics to measure veterans' care experiences.

Conclusion

Major transformations in US health care policy—the greatest in over a generation—will markedly impact the care delivered to many individuals, especially veterans and returning military service members. Rigorous program evaluations will be crucial as state and federal policy makers make tough decisions regarding the value of new programs and policies. The growing interest in randomized designs, as well as proposed legislation that would mandate that laws be based on evidence-based policy, will call for a greater investment in rigorous evaluations. These will require more partnerships between scientists and policy makers to implement these designs in light of trade-offs between scientific rigor and political realities. *NCMJ*

Amy Kilbourne, PhD, MPH director, VA Quality Enhancement Research Initiative, Health Services Research and Development Service, US Department of Veterans Affairs, Washington, District of Columbia.

David Atkins, MD, MPH director, VA Health Services Research and Development Service, US Department of Veterans Affairs, Washington, District of Columbia.

Acknowledgments

Financial support. A.K. and D.A. are paid employees of the US Department of Veterans Affairs.

Disclosure. The views expressed in this article are those of the authors and do not necessarily represent the views of the VA.

Potential conflicts of interest. A.K. and D.A. have no relevant conflicts of interest.

References

1. Bridgeland J, Orszag P. Can government play moneyball? The Atlantic. July/August 2013. <http://www.theatlantic.com/magazine/archive/2013/07/can-government-play-moneyball/309389/>. Accessed August 17, 2015.
2. National Center for Veterans Analysis and Statistics. US Department of Veterans Affairs website. http://www.va.gov/vetdata/veteran_population.asp. Accessed August 17, 2015.
3. Kessler RC, Heeringa SG, Stein MB, et al; Army STARRS Collaborators. Thirty-day prevalence of DSM-IV mental disorders among nondeployed soldiers in the US Army: results from the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *JAMA Psychiatry*. 2014;71(5):504-513.
4. Gizlice Z. Health conditions and behaviors among North Carolina and United States military veterans compared to non-veterans. State Center for Health Statistics website. <http://www.schs.state.nc.us/schs/pdf/schs-133.pdf>. Accessed August 17, 2015.
5. Veterans Choice Program. US Department of Veterans Affairs website. <http://www.va.gov/opa/choiceact/>. Accessed August 17, 2015.
6. Institute of Medicine. Best Care at Lower Cost: The Path to Continuously Learning Health Care in America. Washington, DC: National Academies Press; 2012.
7. Executive Office of the President, Office of Management and Budget. Next steps in the evidence and innovation agenda. Memorandum M-13-17. White House website. <https://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-17.pdf>. Accessed August 17, 2015.
8. US House of Representatives, 113th Congress. H.R. 1287 - Sound Science Act of 2013. Congress.gov website. <https://www.congress.gov/bills/113th-congress/house-bill/1287>. Accessed August 17, 2015.
9. US House of Representatives, 114th Congress. H.R. 185 - Regulatory Accountability Act of 2015. Congress.gov website. <https://www.congress.gov/bills/114th-congress/house-bill/185/text>. Accessed August 17, 2015.
10. Owens DK, Lohr KN, Atkins D, et al. Grading the strength of a body of evidence when comparing medical interventions. In: *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*. Rockville, MD: Agency for Healthcare Research and Quality; 2009.
11. Finkelstein A, Taubman S. Health care policy. Randomize evaluations to improve health care delivery. *Science*. 2015;347(6223):720-722.
12. Allen H, Baicker K, Taubman S, Wright B, Finkelstein A. The Oregon Health Insurance Experiment: when limited policy resources provide research opportunities. *J Health Polit Policy Law*. 2013;38(6):1183-1192.
13. Kolata G. Method of study is criticized in group's health policy tests. *New York Times*. February 2, 2014. http://www.nytimes.com/2014/02/03/health/effort-to-test-health-policies-is-criticized-for-study-tactics.html?_r=2. Accessed August 17, 2015.
14. Taubman SL, Allen HL, Wright BJ, Baicker K, Finkelstein AN. Medicaid increases emergency-department use: evidence from Oregon's Health Insurance Experiment. *Science*. 2014;343(6168):263-268.
15. Baicker K, Taubman SL, Allen HL, et al. The Oregon experiment—effects of Medicaid on clinical outcomes. *N Engl J Med*. 2013;368(18):1713-1722.
16. Crain AL, Solberg LI, Unützer J, et al. Designing and implementing research on a statewide quality improvement initiative: the DIAMOND study and initiative. *Med Care*. 2013;51(9):e58-66.
17. Kotz D, Spigt M, Arts IC, Crutzen R, Viechtbauer W. The stepped wedge design does not inherently have more power than a cluster randomized controlled trial. *J Clin Epidemiol*. 2013;66(9):1059-1060.
18. Frakt AB. An observational study goes where randomized clinical trials have not. *JAMA*. 2015;313(11):1091-1092.
19. Prentice JC, Conlin PR, Gellad WF, Edelman D, Lee TA, Pizer SD. Capitalizing on prescribing pattern variation to compare medications for type 2 diabetes. *Value Health*. 2014;17(8):854-862.
20. Powell BJ, Waltz TJ, Chinman MJ, et al. A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. *Implement Sci*. 2015;10:21.
21. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. *Med Care*. 2012;50(3):217-226.
22. Chung B, Jones L, Dixon EL, Miranda J, Wells K; Community Partners in Care Steering Council. Using a community partnered participatory research approach to implement a randomized controlled trial: planning community partners in care. *J Health Care Poor Underserved*. 2010;21(3):780-795.
23. Kilbourne AM, Almirall D, Eisenberg D, et al. Protocol: Adaptive Implementation of Effective Programs Trial (ADEPT): cluster randomized SMART trial comparing a standard versus enhanced implementation strategy to improve outcomes of a mood disorders program. *Implement Sci*. 2014;9:132.
24. QUERI - Quality Enhancement Research Initiative. US Department of Veterans Affairs website. <http://www.queri.research.va.gov/choiceact.cfm>. Accessed August 17, 2015.