

## Program Evaluation of Total Force Fitness Programs in the Military

Joan A. Walter, PA\*; Ian Coulter, PhD†; Lara Hilton, MPH‡; Amy B. Adler, PhD‡;  
LTC Paul D. Bliese, MSC USA§; Maj Rena A. Nicholas, USAF BSC||

---

**ABSTRACT** The Chairman of the Joint Chiefs of Staff has charged the military's medical and research communities to join with the line community in rapidly and efficiently building a framework to promote and measure total force fitness (TFF). The need to identify the elements of promising TFF programs, and to evaluate whether and how they work, has taken on a new urgency, as we witness the heavy toll of combat exposure on our military members and families. We propose a rigorous methodology for conducting program evaluation, including the study of structure, process, and outcomes. The proposed model combines both quantitative and qualitative methods, to assess the effectiveness and replicability of holistic, multidimensional programs as they are implemented. We describe an additional model for the assessment of efficacy through group randomization trials. The result is a comprehensive model that can be used to assess and compare TFF programs.

---

### INTRODUCTION

The concept of total force fitness (TFF), when applied to the military, is defined as follows: A state in which the individual, family, and organization can sustain optimal well-being and performance under all conditions; well-being encompasses physical, mental, social, and spiritual well-being and not merely the absence of disease or infirmity. Total force fitness promotes resilience: the resources to withstand recover and/or grow in the face of stressors and changing demands.

At a time when our all-volunteer force is under extreme stress due to multiple and frequent deployments in two theaters, with short dwell time and heavy enemy engagement, maintaining the health of our fighting force is of critical importance. At the same time, Department of Defense health-care costs are projected to rise at twice the rate of overall U.S. healthcare spending between 2001 and 2011.<sup>1</sup> Still, perhaps the most compelling reason to commit to total force fitness for the military is the moral imperative to safeguard the health of individuals who have volunteered to be put in harm's way in service to their country.

Many warriors enter the military in top physical and mental form; others reflect lifestyle behaviors that include poor sleep hygiene, unhealthy diet, and inattention to spiritual and social support. The military has the unique opportunity to establish standards for millions of Americans and to achieve them through the implementation of total force fitness programs.

If the goal of total force fitness is "universal" health strength and vigor, a single program may be inadequate to meet all service members' and units' fitness needs. It is the nature of human capital that "one size fits all" is not a viable approach. Unlike tanks, aircraft carriers, and mechanized weaponry, human beings are not built to military specifications, and getting them to achieve a standard level of fitness and health is a daunting challenge. Therefore, total force fitness endeavors must also include a method for evaluating and comparing any programs, including the ability to identify which elements work best for which subpopulations and in which environments. Laying out the fundamental contributors to total force fitness, along with standardized metrics to assess progress toward the goal of total force fitness, is insufficient if it does not include the tools for evaluating and comparing the effectiveness of programs according to these dimensions. Only then can policy be set and resources be allocated in the most efficient manner.

In developing this evaluation framework, it is important to consider both the need for rigorous methods and for relevance of results. Rigor is established through the use of valid research methods and design that are accepted by the scientific community. Relevance addresses the more subjective elements of the particular values and needs that the research is intended to meet. Thus, a virtual "evidence house" is necessary in which multiple methods and research designs are included to provide a complete view of the value of the studied program. Unlike the classical "evidence hierarchy," the evidence house recognizes that various research methodologies are necessary and no less valuable when they are properly designed to address matters of relevance, feasibility, and related elements of the "does it work?" question. The relative timing of effectiveness and efficacy research can be determined on a case-by-case basis; where the 80% solution exists, we posit that the mixed methods, continuous process improvement program evaluation might well be the most efficient way to drive toward the 100% goal.<sup>2-4</sup>

---

\*Samueli Institute, 1737 King Street, Suite 600, Alexandria, VA 22314.

†RAND Corporation, 1776 Main St, Santa Monica, CA 90407.

‡US Army Medical Research Unit-Europe, CMR 442, APO AE 09042.

§Walter Reed Army Institute of Research, 503 Robert Grant Ave., Silver Spring, MD 20910.

||Lackland AFB, 2200 Bergquist Dr., Ste 1, Lackland AFB, TX 78236.

The views expressed in this article are those of the authors and do not necessarily represent the official policy or position of the U.S. Army Medical Command, U.S. Air Force, or the Department of Defense.

While the individual and group metrics defined elsewhere in this journal will allow us to determine outcomes using a rigorously derived toolkit, they do not address evaluation of the structure and process of the programs and therefore consider neither the relevance to particular populations nor the effectiveness of the programs as implemented. The quantitative metrics do not speak to the issue of replicability of the programs nor do they foster continuous quality improvement that comes from a feedback process about the program's operation in real time. Last but not least, the success of any program is highly dependent on contextual factors, those that arise from the unique features of the setting and the implementation. The latter is a significant issue in the military, which permits considerable leeway at both the service level and installation level for implementing programs. What is required therefore is a systematized process, including the collection of both quantitative and qualitative data as programs are developed and implemented, to determine what works to approximate the goals, to understand how and why certain elements contribute to success, and to identify those that act as barriers. Such a standardized process for rigorous data collection, with a robust dataset and the ability to measure how changes in programs lead to changes in achieving the goal, is termed program evaluation.

Another advantage of the proposed program evaluation methodology is that it is sufficiently flexible to allow multiple programs with perceived value to be implemented while being rigorously studied and compared using standardized metrics, and to cycle important information on lessons learned back into the program, thus promoting performance improvement and timely progress toward TFF. That approach accommodates the urgent need for TFF programs now, while ensuring that each iteration yields valuable information on how to improve the program for the next cycle. The mixed methodology we propose is essential to gauge key factors, such as scalability, facilitators and obstacles to success, and midcourse changes that often occur in real-life, real-time implementation; moreover, the mandate to produce total force fitness requires both the assessment of each individual's fitness, using the metrics outlined in the accompanying articles in this journal and the study of group dynamics and interactions that work to either enhance or degrade the effectiveness of a program. By combining the collection of both quantitative and qualitative data in a program evaluation model with group randomized trials to determine efficacy, the necessary information to achieve total force fitness can be obtained.

### **RELEVANCE TO THE MILITARY**

Leadership at all levels of the military recognizes the challenge of total force fitness, to identify its key elements, to describe how they interact, and to select validated and relevant outcome measures to assess achievement in each domain and in totality. Similarly, each of the armed services has performance challenges and goals that require total force fitness of its members. For example, the sustained, asymmetrical warfare against an elusive insurgency, which characterizes Operations

Iraqi Freedom and Enduring Freedom has changed much of the way the Air Force experiences warfare and widened the pool of airmen who experience direct combat conditions. Today, nearly every Air Force career may be exposed to direct enemy engagement. Furthermore, the reliance on the all-volunteer military has required Air Force members to fill roles traditionally filled by members of sister services. Consequently, airmen are increasingly asked to push the boundaries of their traditional roles and training needs—conditions that create risk for physical and mental health problems.<sup>5,6</sup>

Like the other services, the Air Force is interested in strengthening service members' resilience by ensuring that they have skills to promote physical health and restedness, healthy thinking and stress management, leadership and group cohesion, and organizational commitment. These factors have been identified since World War II to be helpful for enabling troops to continue to perform under adverse conditions.<sup>7</sup> Efforts are ongoing to develop credible programs that can increase knowledge of these skills; increase confidence and motivation to use them; and results in improved resilience and decreased adverse effects, such as burnout, injuries, mental health distress, and casualties. While there is evidence that psychoeducational briefings can have a positive impact,<sup>8,9</sup> they may not be sufficient to produce desired outcomes<sup>10</sup> and there is growing support for more comprehensive skills-based training programs. This emphasis on program evaluation is especially important to ensure that the time and energy spent on resiliency-enhancing programs are truly effective rather than just efficient.

The recent TFF workshop acknowledged that TFF is multifactorial, encompassing individual, relational, group, and environmental factors. Developing a construct of total force fitness, in which the interdimensional effects are predictable and measurable, is a complex endeavor; ultimately the need is line driven. Key terms and definitions from the literature must be translated into line and operationally relevant language and the outcomes must be measured using military performance metrics. Program evaluations attach to pragmatic implementations and provide relevant, timely information that is usable by the line.

To generate a model capable of producing meaningful, actionable results, an optimal evaluation framework must provide an understanding of logistics of implementation, environmental conditions, subtle group dynamics, and the other salient features of any program. The only way to do this is to apply mixed methods, including both qualitative and quantitative research.

Herein, we make the case for a standardized program evaluation methodology as the best framework for assessing the real-life effectiveness of programs instituted to promote total force fitness. The power of program evaluations derives from a number of features:

- Its incorporation of both qualitative and quantitative research methods creates a rich tapestry of data.

- Because it is conducted in real-life settings, as opposed to controlled research environments, it can identify and accommodate variations from the intended program and address feasibility factors.
- It can easily be performed repeatedly as program improvements are implemented and can provide a real-time feedback loop, facilitating continuous performance improvement in design and delivery.
- It provides meaningful information in the event that the program intervention proves ineffective in meeting the performance goals.
- Its focus on effectiveness in real-world settings as opposed to efficacy in controlled environments is consonant with the chairman's call for immediate solutions.

We posit that program evaluation should be a required element of all programs being implemented under the rubric of TFF. There are several compelling reasons to make standardized evaluation a requisite. Doing so would ensure that every program yields valuable data that inform future effort. Further, such evaluations produce a catalog of data that enables comparison of programs, populations, and environments.

### PROGRAM EVALUATION

Too often there is a disjuncture between the goals of research and the operational goals of the military. The gold standard of research is the randomized controlled trial (RCT), which focuses on efficacy in controlled environments. However, such trials are not always feasible or pragmatic; they may require large numbers of soldiers, and strict adherence to intervention guidelines that may be easily disrupted by unanticipated military operational events. Beyond determination of what works under ideal trial conditions is the need to understand why it works, whether it is practical or affordable in the real world of the line and how best to implement it. In that regard, the standard RCT may be too narrow a design, and the results may yield limited conclusions that only tentatively support a "go/no-go" decision. Properly constructed mixed-methods research can provide more actionable information, both on what not to do again and what to do more of in the future. Depending on the nature of the program and the outcomes of interest, program evaluations may be completed in as short a period as 2–4 months or may extend over a period of years.

Health, education, and training programs have as their ultimate goal preventing or treating disease, changing behavior, attitudes, beliefs, and increasing knowledge and skills. More often than not, trainings target significant and difficult changes in attitudes and risk/protective behavior such as those related to stress and resiliency. Because the stakes are so high, there is a pressing need for accountability to stakeholders, to include assessing effectiveness, documenting implementation challenges and solutions, developing and applying standard metrics, and comparing outcomes across programs, to inform resource allocation.

Toward this end, the Toolkit for the Evaluation of Resilience training programs has been developed by the Samuelli Institute in conjunction with a consulting team from RAND Corporation. The toolkit incorporates the theoretical elements outlined below and adds a preliminary set of validated metrics to assess outcomes of interest to assess resilience. It is currently in use in an evaluation being conducted by the Samuelli Institute of a skills-based resilience training program in a brigade combat team that has deployed to Afghanistan. As designed, this particular evaluation will be completed over a 2-year period; assessments of structure and process of a discrete training intervention require only a 4- to 6-month period, but quantitative data on resilience and related health outcomes is being collected over a period that begins predeployment and extends to the Post-Deployment Health Reassessment (PDHRA) at 3–6 months postreturn. (As noted below, this longer period of follow-up is desirable to assess whether the training's effects on health are sustained.)

### PROGRAM EVALUATION THEORY

Standard program evaluation involves three elements: (i) structural evaluation, (ii) process evaluation, and (iii) outcome evaluation (Fig. 1). This information helps to determine the value of a program and assists those who may wish to expand, change, or replicate it in other environments. Such an approach is based on the discovery of unique factors within a program (leadership, relationships, culture, structure, rewards) that bring about the outcomes.

Process evaluation assesses the extent to which the intervention components are implemented as planned. Summative evaluation measures the extent to which program goals and objectives were achieved and the intermediate and longer-term impact of the program.

### DATA COLLECTION TECHNIQUES

Data are collected using multiple methods,<sup>11</sup> a strategy that yields higher quality results<sup>12</sup> and that allows for triangulation<sup>12–14</sup> to increase the reliability and validity of the findings.<sup>14</sup> It will combine both qualitative and quantitative techniques, a combination that has long been recognized as a powerful research strategy, especially for exploratory studies of new programs, such as TFF programs.<sup>11,12,14</sup>

### THE MODEL

We reviewed several classic models of program evaluation. These included the following: Donabedian,<sup>15</sup> Andersen,<sup>16</sup>

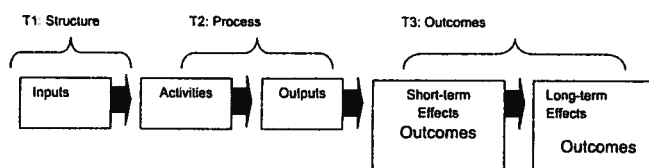


FIGURE 1. Logic model with evaluation components.

Chinman,<sup>17</sup> Green and Kreuter's Precede/Proceed Model,<sup>18</sup> RE/AIM,<sup>19</sup> CDC Self-Guided Protocol for Program Evaluation,<sup>20</sup> and Scrimshaw's rapid ethnographic assessment.<sup>21,22</sup>

We have drawn on this wide body of work to construct our program evaluation toolkit. From Donabedian's model we incorporated the foundations of evaluation, adding Andersen's approach for rigorous measurement and the social/community interaction. From Chinmann's "Getting To Outcomes" model, we took the practical application of evaluation so that practitioners can plan, implement, and evaluate their own programs. From Green's "Precede/Proceed" model we took the logic model and feedback loop, which addresses midcourse corrections better than a linear logic model (Fig. 2). The RE/AIM model added the maintenance and reinforcement components, which are essential in training/education programs. The conceptual and functional frameworks were adapted from the CDC evaluation toolkit because they provide a simple snapshot of evaluation. The RAP model added the qualitative data collection component. Logic models are used extensively throughout; all the models we reviewed contributed essential components to our evaluation model. Finally, Ryan's model added the systems perspective.

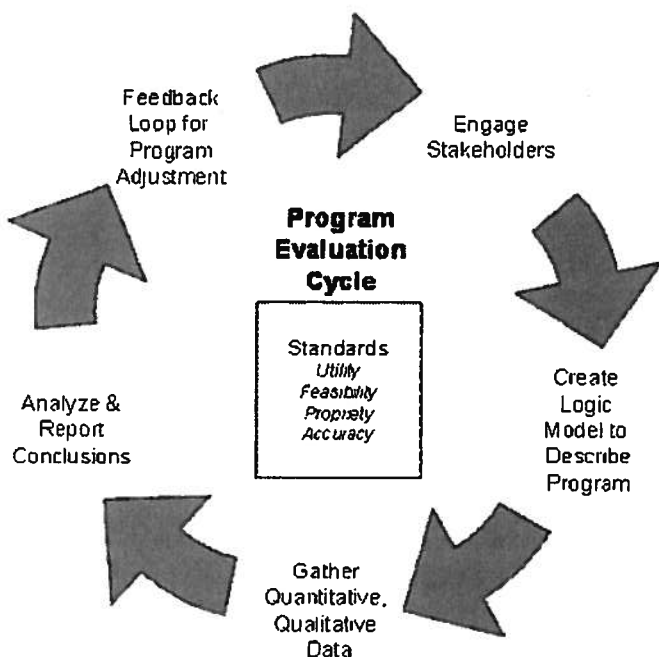


FIGURE 2. The program evaluation cycle.

## STEPS IN EVALUATION PRACTICE<sup>20</sup>

### Structural Evaluation

The determination of the structure of the program and its sites requires interviews with key personnel and a review of official descriptions of the program and the environment. It also involves the logic of the program, the goals/milestones, and values.<sup>23</sup> The aim is to get as complete a description of the

program as possible from those who are implementing it. For evaluation of TFF programs we suggest the following components be included:

### Organization

- Were the implementation sites optimal? Were the facilities adequate to support the training?
- How was the program organized? Was the per-person time allocation sufficient to convey the concepts and/or skills? Was the schedule feasible and appropriate to the military duty environment?
- Did this training program integrate well with existing programs? Was its design consistent with other military programs? Did it support/enhance institutional principles and culture?
- Was the implementation site properly selected to target the intended population and outcomes stated?
- If there was a needs assessment, did the program's design meet the identified needs?

### The Program

- Who did the training? How many trainers were there and were they sufficiently knowledgeable about military life to be credible teachers?
- What was the content of the training program? Did it contain didactic and experiential elements? What skills were trained? Did the principles support the skills, and vice versa?
- Were the training goals clearly stated?

### Staff/Participants

- What were their qualifications?
- Were other faculty/staff involved in the program? Was the staff augmented by military?
- What are the affiliations of the staff (e.g., institution, department, community), and how were they selected?
- What are their backgrounds, credentials/certifications?
- Was the staff adequate for the sessions?

### Process Evaluation

The primary roles of the process evaluation are to identify deviations from the intended program, barriers that need to be overcome if the program is to work effectively and to assess the impact of midprogram modifications. The secondary goal is to provide a clear description of how the program worked so that successful programs can be replicated, and less successful programs can be improved.

We suggest using a two-stage evaluation, beginning with an initial assessment to identify how the program really operates (the lived program as opposed to the program on paper), followed by a more expansive evaluation to identify factors that may explain the program's success or the actual or potential barriers that interfere with its success.

Process evaluations should also be designed to record important implementation decisions and document change

throughout the history of the program; moreover, consideration should be given to providing timely identification of barriers to the decision makers in a timely fashion, to allow for midcourse corrections if needed. In summary, the process evaluation is conducted to characterize and understand:

- The extent to which the program is being implemented according to plan, including assessment and documentation of the degree of integrity, fidelity, variability in the program implementation, expected or unexpected, planned or unplanned.
- What components of the intervention appear to be responsible for outcomes.
- The relationship between the program context (i.e., setting characteristics) and the program processes (i.e., implementation).
- The quality of staff implementing program and the impact of that quality on the successful delivery.
- Frequency of program sessions, intensity, and duration.
- Size of group receiving program.
- The correlation of the program's design, goals, and outcomes.

### **Outcome Evaluation**

Outcome evaluation involves comparing what the original program objectives were with what has been achieved before and after the program, using both quantitative and qualitative measures. For TFF programs, the metrics presented in the accompanying manuscripts would form the standardized dataset for functional and health outcomes, allowing for comparisons of various programs, based on their relative success in promoting the desired outcomes. Wherever possible this should be done with objective measures such as medical records or survey data on the following types of quantitative outcomes, measured pre- and postprogram:

- Changes in risk or protective behaviors (i.e., healthy sleep patterns, nutrition, stress reduction, use of alcohol).
- Changes in trends in morbidity and mortality (i.e., anxiety, depression, anger management, blood pressure, other health indices and biomarkers).
- Changes in the environment, including policies, formal and informal enforcement of regulations, and influence on social norms and other societal forces (i.e., allocation of down time to practice self-regulation).
- Additional objective metrics relevant to the goals/objectives of the particular program being evaluated, such as those outlined in this journal.

Other important outcomes that are frequently included are changes in beliefs/attitudes/knowledge, which are subjective measures, as opposed to changes in behavior or biomarkers. The purpose of the additional qualitative assessments is to elucidate issues that cannot be answered by the quantitative analyses and to explore additional areas that are difficult to address through quantitative data, including the following,

obtained through interviews with participants, staff, and administrators:

- Participant acceptability and receptivity (satisfaction).
- Participant descriptions of the program, including things they liked or disliked, improvements they suggest.
- Perceived side effects or undesired effects.
- Perceived unmet needs.
- Economic outcomes (cost versus benefits).
- Change in beliefs, attitudes, knowledge about a particular topic.
- Anticipated change in behaviors, expected degree of use and applications.
- Participants' comparisons of the program to others they have learned or experienced.

Once the program evaluation has been completed, the data analyzed and interpreted, the final report should address the overall utility of the program for its intended users and outcomes; the feasibility of the program as designed and implemented (accounting for such elements as acceptance, total resource costs, successful completion, etc.); recommendations for future programs, addressing the weaknesses and strengths of the program, and lessons learned about reparative measures; and a thoughtful, scientific analysis of how the program compared overall to others with similar goals. Both the operational line community and the research community can extract important, pragmatic information from this report.

## **THE EVIDENCE HOUSE BEYOND PROGRAM EVALUATION**

### **Measuring Efficacy**

Program evaluation provides some measure of program coherence and effectiveness; however, as noted above, ideally research should also assess program efficacy. One way this has been done in the Army is through randomized trials in which individuals were randomly assigned to a program of intervention or some other comparison condition (either an active comparison intervention or no intervention at all). Adequate sample sizes and randomization are designed to ensure that differences between the intervention and comparison conditions can be reliably attributed to the intervention. In the military context, fitness-type interventions can be designed to leverage unit-level strengths such as cohesion and leadership. Because programs are typically implemented in existing units, it is often preferable to randomize by intact unit (i.e., conduct a group randomized trial or GRT). The Army's GRTs generally entail randomizing 30 or more groups and conducting analyses that account for pre-existing group-level differences.<sup>8</sup>

The main goal of a GRT and other randomized trials is to establish with some certainty that positive outcomes are caused by the targeted intervention. Research designs are strongest when they measure and evaluate outcomes (such as mental health status) months after the intervention because such designs test for evidence of long-lasting effects. Designs

that assess outcomes immediately after the intervention are weaker because they fail to provide evidence of lasting impact. The strongest GRTs also test an intervention against an active control. That is, the targeted intervention is contrasted to an alternative intervention that is face valid to those undergoing the intervention. Finally, within the context of the military, universally applied programs are more realistic in many cases than targeted interventions that focus only on high-risk individuals.

Importantly, design elements that strengthen the scientific credibility and/or applicability of the research also reduce the effect sizes (e.g., the measurable impact of the intervention). This concept can be understood in the context of the earlier discussion on the need for both rigor and relevance in the evidence house. For example, it may be more difficult to detect changes months after the program implementation; therefore, this design often leads to small effect sizes. Similarly, contrasting an intervention to an active control is often associated with smaller effect sizes than contrasting an intervention to a passive control condition.<sup>24</sup> Finally, universally applied programs generally have lower effect sizes than targeted programs.<sup>24</sup>

For these and other reasons, authors such as Bliese et al.<sup>25</sup> have argued that in conducting research on efficacy, focusing on statistical significance is more important than focusing on effect sizes in large-scale universally applied resilience programs. Fundamentally what is important to demonstrate efficacy is that resilience-based programs have a statistically significant probability of improving mental health outcomes and that such programs do not demonstrate harm.

At a theoretical level, it is critical to understand that even seemingly small effect sizes can have meaningful effects when (a) the process is cumulative and (b) when large populations are involved. In terms of a cumulative effect, for instance, we would reject a treatment for the common cold that had a low effect size for preventing the next single exposure to cold virus; however, we would accept a treatment with a low effect size (on a single occasion) if it would lower the risk for the next 10 years of exposure. Intervention programs such as those that would comprise a total force fitness program are often designed to be cumulative—teaching skills that generalize across time and situations—and these effects may be missed in traditional randomized trials that have few follow-up measurement occasions. In terms of populations, the fact that resilience programs are being designed to impact hundreds of thousands of individuals means that even small effect sizes will result in positive outcomes for hundreds if not thousands of service members. Thus, the value of even small effect sizes is compensated for by the magnitude of the institutional benefit.

Despite the fact that GRTs are resource intensive, it is possible to assess the efficacy of an intervention using these techniques.<sup>8</sup> Such studies demonstrate not only the feasibility of conducting a GRT but also the possibility of demonstrating that interventions designed to leverage the military social context can result in positive effects.

## CONCLUSIONS

In tandem with the development and adoption of a toolkit containing the elements and metrics for the military-wide commitment to total force fitness, a rigorous method of evaluation will provide not only a framework to show that total force fitness programs are successful, but also the basis for continuous process improvement. It is vital that such methods be harnessed to deliver TFF programs now when the need is urgent, and that they can yield timely results that are relevant to the line, producing actionable evidence of effectiveness and efficacy and the evaluation of the facilitators and barriers to successful implementation, expansion, and replication. The proposed mixed-methods program evaluation will enable leaders to understand why and how a program was or was not successful to tailor the next effort thoughtfully and on the basis of scientific information. Conducting structural, process, and outcomes evaluation will provide the results required to justify spending precious resources on TFF programs, guide the allocation of resources to those with evidence-based success, and will ultimately yield a catalog of TFF programs suitable for implementation at various locations among troops with a variety of needs. For the assessment of efficacy, group randomization trial design, a comprehensive, scientifically rigorous and operationally meaningful variation on program evaluation has been described. Because universally applied programs may demonstrate small treatment effect sizes, the information obtained through program evaluation, including such things as feasibility, replicability, subjective value ratings, cost analysis, and cultural/environmental fit may be the determining factors in deciding which program(s) to implement. Importantly, the pragmatic methodology, in which programs are studied as they are actually implemented in real-life, brings the research function into closer alignment with the needs of the line and ensures that research lessons can be cycled into practice more efficiently. At a time when stress-related illness and dysfunction are exacting a heavy toll on our military communities, the need for identifying holistic programs that work both in principle and in fact is vital.

## ACKNOWLEDGMENTS

The Total Force Fitness conference on which this article was based was supported by award number MDA 905-03-C-0003 (Uniformed Services University of the Health Sciences). The writing of this article was supported by award number W81XWH-08-1-0615-P00001 (United States Army Medical Research Acquisition Activity).

## REFERENCES

1. Zoroya G: Military's health care costs booming. *USA Today*, April 22, 2010.
2. Jonas W, Lewith G: Toward standards of evidence for CAM research and practice. In: *Clinical Research in Complementary Therapies*, Ed 2. Edited by Lewith G, Jonas W, Walach H. Littleton, NH, Churchill Livingstone, 2010 (in press).
3. Jonas W: The evidence house: how to build an inclusive base for complementary medicine. *West J Med* 2001; 175(2): 79-80.

4. Jonas W: Building an evidence house: challenges and solutions to research in complementary and alternative medicine. *Forsch Komplementarmed Klass Naturheilkd* 2005; 12(3): 159–67.
  5. Gal R, Jones F: A psychological model of combat stress. In: *Textbook of Military Medicine, War Psychiatry*. Edited by Bellamy R, Jenkins D. Falls Church, VA, Office of the Surgeon General, U.S. Department of the Army, 1995.
  6. Wilcox V: Burnout in military personnel. In: *Textbook of Military Medicine, Military Psychiatry: Preparing in Peace for War*. Edited by Zajchuk R, Bellamy R, Jenkins D. Falls Church, VA, Office of the Surgeon General, U.S. Department of the Army, 1994.
  7. Manning F: Morale and cohesion in military psychiatry. In: *Textbook of Military Medicine, Military Psychiatry: Preparing in Peace for War*. Edited by Zajchuk R, Bellamy R, Jenkins D. Falls Church, VA, Office of the Surgeon General, U.S. Department of the Army, 1994.
  8. Adler A, Bliese P, McGurk D, Hoge C, Castro C: Battlemind debriefing and battlemind training as early interventions with soldiers returning from Iraq: randomization by platoon. *J Consult Clin Psychol* 2009; 77: 928–40.
  9. McKibben E, Britt T, Hoge C, Castro C: Receipt and rated adequacy of stress management training is related to PTSD and other outcomes among Operation Iraqi Freedom veterans. *Mil Psychol* 2009; 21(1): S68–81.
  10. Sharpley J, Fear N, Greenberg N, Jones M, Wessely S: Pre-deployment stress briefing: does it have an effect? *Occup Med (Chicago, IL)* 2008; 58(1): 30–4.
  11. Henry G, Melvin M: Beyond use: understanding evaluation's influence on attitudes and actions. *Am J Eval* 2003; 24(3): 293–314.
  12. Donaldson S, Scriven M: *Evaluating social programs and problems: visions for the new millennium*. Mahwah, NJ, Erlbaum, 2002.
  13. Tashakkori A, Teddlie C: *Mixed methodology: combining qualitative and quantitative approaches*. Thousand Oaks, CA, Sage Publications, 1998.
  14. Yin R, Bateman P, Moore G: *Case studies and organizational innovation: strengthening the connection*. Washington, DC, Cosmos, 1983.
  15. Donabedian A: Promoting quality through evaluating the process of patient care. *Med Care* 1968; 6(3): 181–202.
  16. Andersen R: Methods and problems in evaluation of primary health care: does it matter which way we go? Paper presented at Conference on Evaluation of Primary Health Care, Center for Primary Care Research, University of Uppsala, Uppsala, Sweden, October 28–29, 1991.
  17. Chinman M, Imm P, Wandersman A: *Getting to outcomes: promoting accountability through methods and tools for planning, implementation, and evaluation*. Santa Monica, CA, RAND Corporation, 2004.
  18. Green L, Kreuter M: *Health Program Planning: Precede/Proceed Model, Ed 4*. New York, McGraw Hill, 2005.
  19. Glasgow R, Vogt T, Boles S: Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *Am J Public Health* 1999; 89(9): 1322–1327.
  20. U.S. Department of Health and Human Services: *Introduction to program evaluation for public health programs: a self-study guide*. Atlanta, GA, Centers for Disease Control and Prevention, 2005.
  21. Beebe J: *Rapid assessment process: an introduction*. Walnut Creek, CA, AltaMira, 2001.
  22. Scrimshaw S, Hurtado E: *RAP: rapid assessment procedures, Vol 11*. Tokyo, United Nations University, 1987.
  23. Scriven M: *Evaluation Thesaurus*. Newbury Park, CA, Sage Publication, 1991.
  24. Brunwasser SM, Gillham JE, Kim ES: A meta-analytic review of the Penn Resiliency Program's effect on depressive symptoms. *J Consult Clin Psychol* 2009; 77: 1042–54.
  25. Bliese PD, Adler AB, Castro CA: Research-based preventive mental health in the military. In: *Deployment Psychology: Evidence-Based Strategies to Promote Mental Health in the Military*. Edited by Adler AB, Bliese PD, Castro CA. Washington, DC, American Psychological Association, (in press).
-