

Enhanced Stress Resilience Training for Faculty Physicians (ESRT-Faculty)

IRB STUDY # 18-24596 (approved 4/03/2018)

Project Summary:

Mounting evidence shows that burnout, a critical metric for dissatisfaction and distress, is a growing problem within medicine. Burnout is a syndrome associated with worse physician performance, patient outcomes, and hospital economics. Furthermore, investigators are coming to understand that burnout, diminished performance and the development of mental and physical illness are related. It has been proposed that chronic and overwhelming stress, in the absence of adequate coping skills, promotes performance deficits from surgical errors to poor professionalism due to the effects of stress on cognition.

Notably, in small studies of physicians and other high-stress/high-performance groups mindfulness-based interventions have shown exceptional promise in improving burnout and distress symptoms, protecting cognition, and enhancing meaningfulness and satisfaction in work. Nevertheless, in spite of promising results in various populations the translation of mindfulness-based interventions to real-world settings has been slow. There is a paucity of quality research examining individually-based interventions, formal mindfulness training in physicians, or either of these things in the high stakes world of surgeons and anesthesiologists. To address these gaps, investigators have developed Enhanced Stress-Resilience Training (ESRT) based on MBSR, but streamlined and tailored for surgeons and anesthesiologists. Investigators have administered a needs assessment and found almost 100 faculty interested in such training, and concluded beta-testing of the curriculum which allowed for further refinement of logistics, dose and delivery. Finally, to establish a baseline in both departments, investigators administered a comprehensive well-being survey which captures multiple facets of psychosocial well-being and will be used along with executive function testing and measures of performance to evaluate the efficacy of our training in a stepped-wedge trial of faculty surgeons and anesthesiologists at multiple UCSF hospital sites.

Design:

ESR Training for Faculty is a stepped-wedge trial in Surgery and Anesthesia faculty. It will evaluate the feasibility of formal mindfulness-based training in these groups and how this impacts mental health, stress and performance at baseline, post-intervention, 6-month and 18-month follow-up.

The primary hypothesis is that formal mindfulness-based training is a feasible and objectively effective method for enhancing personal and professional well-being in physicians. The secondary hypothesis is that the benefits of such training will extend to improvements in physician cognition and performance.

Aim 1: To show that formal mindfulness training is feasible for surgery and anesthesia faculty. Investigators will recruit and retain faculty and provide ESRT using two highly experienced instructors and a

codified curriculum. Using mixed methods, investigators will evaluate comprehensive feasibility (comprised of demand, implementation, acceptability, practicality, adaptation and integration) on individual, departmental and institutional levels, as recommended by the NIH.

Aim 2: To show that mindfulness training will enhance well-being and performance in surgeons and anesthesiologists. Using a stepped-wedge trial design and objective measures, investigators will evaluate well-being (burnout, stress, mental health, work balance and work strain) and performance (executive function, professionalism and team work) at baseline, post-training and at 6 and 18-months follow-up.

Background and Significance:

The significance of this work lies in evaluating a process-centered skill believed to promote resilience, defined as the ability to thrive under adversity and predicated on the perception of stressors as challenges rather than overwhelming threats. The transformation of how humans experience stress is a learned skill that can be applied across career stages, practice trajectories and life. The potential to protect individuals, while investigators work for mandatory institutional and systemic change, is powerful. Moreover, the resultant tendency for self-awareness and equipoise has been contagious in other settings, providing fuel for the greater culture change that is inarguably necessary and holds great promise for us and our patients.

The innovation of this work is in bringing a mind-body intervention to bear not only on well-being but also on the fundamental cognitive processes believed to sub-serve performance, such as the impact of attention and working memory capacity on medical decision-making, and of emotional regulation and self-awareness on professionalism and team work. The potential to improve both the operative environment and surgical or medical errors is unprecedented. Moreover, the use of a manualized curriculum specifically crafted for physicians could pave the way for translation to other specialties and institutions.

Procedures & Methods:

Intervention. The mindfulness-based intervention will be derived from Mindfulness-Based Stress Reduction (MBSR) but in the form of Enhanced Stress Resilience Training (ESRT). ESRT was developed with modifications to languaging, some activities and contextualization to tailor the training to physicians and in particular those involved in the operating room environment. The structure involves a 'learning component' consisting of 6 weekly 1.5-hour group classes and 2-4 hour retreat. Classes focus on developing mindfulness skills (i.e. sustained attention and open monitoring) in addition to training focused on stress and coping. 20 minutes per day of mindfulness exercises following guided meditation CDs or independent practice is assigned. A 2-4 hour

intensive retreat occurs at week 6. The central exercises of the training are the body scan, sitting meditation and limited yoga. The weekly teaching sessions will be offered at various days and times each week to try and accommodate changing and complicated schedules that will vary at each site. Participants can attend whichever session is most convenient. Daily practice will occur independently, unsupervised, with the duration reported weekly through text or email.

Evaluations and sample acquisition. Testing will occur over a week-long period on three different occasions: pre-intervention, post-intervention and at 3-months, 6-months and/or 18-months follow-up. Participants will receive a link to complete the psychological assessments (via Qualtrics) and will undergo neurocognitive testing consisting of the NIH EXAMINER (Executive Abilities: Measures and Instruments for Neurobehavioral Evaluation and Research) and/or CANTAB (the Cambridge Neuropsychological Test Automated Battery) on a tablet. Additionally, investigators will assess performance using Test of Performance Strategies-2 (TOPS-2) and Nasa Task Loader Index.

For the ESRT Study investigators focus on four general areas, well-being, emotional regulation, cognition and surgical performance, investigators will use tools validated in physicians where available. All measures described in this section are published.

Scales to assess well-being:

MBI-Human Services Survey for Medical Personnel (MBI) (West, Dyrbye, Sloan, Shanafelt, 2012). The 2 item MBI addresses 1. Emotional Exhaustion measures feelings of being emotionally overextended and exhausted by one's work and 2. Depersonalization measures an unfeeling and impersonal response toward patients. Each item is rated on a 7-point Likert scale ranging from never (0) to every day (6).

Perceived Stress Scale (PSS) (Cohen, Kamarck, Mermelstein, 1983). The 10 item PSS is the most widely used psychological instrument for measuring the perception of stress. It is a measure of the degree to which situations in one's life are appraised as stressful. Items were designed to assess how unpredictable, uncontrollable, and overloaded respondents find their lives to be. The scale also includes a number of direct queries about current levels of experienced stress. Moreover, the questions are of a general nature and hence are relatively free of content specific to any sub-population group. The questions in the PSS ask about feelings and thoughts during the last month. In each case, respondents are asked how often they felt a certain way. Each item is rated on a 5-point Likert scale ranging from never (0) to almost always (4). Items 4, 5, 7, and 8 are reverse scored.

Cognitive and Affective Mindfulness Scale- Revised (CAMS-R) (Feldman, Hayes, Kumar, et al, 2007). CAMS-R is a uni-dimensional, 10-item inventory that measures mindfulness during general daily occurrences on four components allegedly needed to reach a mindful state (i.e., attention, awareness, present-focus, and acceptance/nonjudgment). Each item is rated on a 4-point Likert scale ranging from rarely (1) to almost always (4). Question 5 is reversed scored.

The Alcohol Use Disorders Identification Test (AUDIT-C) (Bush, Kivlahan, McDonell, et al, 1998). The AUDIT-C is a 3-item alcohol screen that can help identify persons who are hazardous drinkers or have active alcohol use disorders (including alcohol abuse or dependence). Each item is rated on a 5-point Likert scale ranging from never (0) to 4 or more times a week (4).

Patient Health Questionnaire-9 (PHQ-9) (Kroenke, Spitzer, Williams, 2001). The PHQ-9 is 9 item a multipurpose instrument for screening, diagnosing, monitoring and measuring the severity of depression: The PHQ-9 incorporates DSM-IV depression diagnostic criteria with other leading major depressive symptoms into a brief self-report tool. The tool rates the frequency of the symptoms which factors into the scoring severity index. Each item is rated on a 4-point Likert scale ranging from Never (0) to 4 or more times a Nearly Every Day (3).

State-Trait Anxiety Inventory (STAI) (Spielberger, Gorsuch, Lushene, 1983). The STAI is a 20 item scale used to diagnose anxiety and to distinguish it from depressive syndromes. Each item is rated on a 4-point Likert scale ranging from Not at all (1) to 4 or more times a Very much so (4).

Pittsburgh Sleep Quality Index (PSQI) (Buysse, Reynolds, Monk, 1989). The PSQI is a 9 item scale used to measure the quality and patterns of sleep in adults. It differentiates “poor” from “good” sleep quality by measuring seven areas (components): subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleeping medications, and daytime dysfunction over the last month. Each item is rated on a 4-point Likert scale ranging from Not during the past month (0) to Three or more times a week (3).

The Swedish Demand–Control–Support Questionnaire (DCSQ) (SANNE, TORP, MYKLETUN, DAH, 2005). The DCSQ is a 17-item questionnaire, which covers psychological demands, decision latitude, and social support in the workplace. Questions are answered on a four- point scale from 1 Strongly Disagree to 4 Strongly Agree. All item scores except question ‘Do you have sufficient time for all your work tasks?’ are reversed before summation.

Scales to assess emotional regulation:

Mind-Wandering Questionnaire (MWQ) (Mrazek, Phillips, Franklin, 2013). The MWQ is a 5 item scale is intended to measure the frequency of mind-wandering, irrespective of whether mind-wandering is deliberate or spontaneous. Each item is rated on a 6-point Likert scale ranging from Almost Never (1) to Almost Always (6).

The Difficulties in Emotion Regulation Scale (DERS) (Kaufman, Xia, Fosco, 2015). The DERS is an 18 item scale that is based on a clinically-useful conceptualization of emotion regulation that was developed to be applicable to a wide variety of psychological difficulties and relevant to clinical applications and treatment development. Specifically, the conceptual definition of emotion regulation on which the DERS is based emphasizes the functionality of emotions and focuses on adaptive ways of responding to emotional distress, including the: (a) awareness, understanding, and acceptance of emotions; (b) ability to control behaviors when experiencing negative emotions; (c) flexible use of situationally-appropriate strategies to modulate the intensity and/or duration of emotional responses, rather than to eliminate emotions entirely; and (d) willingness to experience negative emotions as part of pursuing meaningful activities in life. Each item is rated on a 5-point Likert scale ranging from Almost Never (1) to Almost Always (5). Questions 1,4,6 are reverse coded: 1. I pay attention to how I feel. 4. I care about what I am feeling. 6. When I'm upset, I acknowledge my emotions.

The Experiences Questionnaire (EQ-Decentering) (Fresco, Moore, van Dulmen, 2007). The EQ-Decentering is a 12 item instrument that assessed decentering. Decentering is defined as the ability to observe one's thoughts and feelings in a detached manner. Items are rated on a 5-point Likert scale, where 1 = never, 5 = all the time.

Multidimensional Assessment of Interoceptive Awareness (MAIA) (Mehling, 2012). The MAIA is a 32 item self-report measure composed of the following 8 subscales: (i) Noticing: awareness of uncomfortable, comfortable and neutral bodily sensations; (ii) Not-Distracting: the tendency to not ignore or distract oneself from sensations of pain or discomfort; (iii) Not-Worrying: the tendency to not react with emotional distress or worry to sensations of pain or discomfort; (iv) Attention Regulation: the ability to sustain and control attention to bodily sensation; (v) Emotional Awareness: the awareness of the connection between bodily sensations and emotional states; (vi) Self-Regulation: the ability to regulate psychological distress by attention to bodily sensations; (vii) Body Listening: actively listening to the body for insight; and (viii) Trusting: experiencing one's body as safe and trustworthy. Individuals rate items based on a 6 point Likert scale from 0-5 with '0' indicating 'Never' and '5' indicating 'Always'. Total scores are obtained through reverse coding items 5, 6, 7, 8, 9 and summing all items.

Batteries to assess cognition:

Executive Abilities: Measures and Instruments for Neurobehavioral Evaluation and Research

(EXAMINER) (Kramer, 2014). EXAMINER assesses a series of tasks targeting working memory, inhibition, set shifting, fluency, insight, planning, social cognition and behavior. The battery includes both computer-administered and paper-and-pencil tasks, and measures inhibition, set shifting, working memory, fluency, planning, error monitoring, insight, and social function.

The Cambridge Neuropsychological Test Automated Battery (CANTAB). The CANTAB includes highly sensitive, precise and objective measures of cognitive function, correlated to neural networks. CANTAB tests have demonstrated sensitivity to detecting changes in neuropsychological performance¹ and include tests of working memory, learning and executive function; visual, verbal and episodic memory; attention, information processing and reaction time; social and emotion recognition, decision making and response control.

1. Reaction Time (RTI). Reaction Time provides assessments of motor and mental response speeds, as well as measures of movement time, reaction time, response accuracy and impulsivity. The participant must select and hold a button at the bottom of the screen. Circles are presented above (one for the simple mode, and five for the five-choice mode.) In each case, a yellow dot will appear in one of the circles, and the participant must react as soon as possible, releasing the button at the bottom of the screen, and selecting the circle in which the dot appeared. Outcome measures are divided into reaction time and movement time for both the simple and five-choice variants.

2. Rapid Visual Information Processing (RVP). Rapid Visual Information Processing is a measure of sustained attention. A white box is shown in the center of the screen, inside which digits from 2 to 9 appear in a pseudo-random order, at the rate of 100 digits per minute. Participants are requested to detect target sequences of digits (for example, 2-4-6, 3-5-7, 4-6-8). When the participant sees the target sequence they must respond by selecting the button in the center of the screen as quickly as possible. The level of difficulty varies with either one- or three-target sequences that the participant must watch for at the same time. Outcome measures cover latency (speed of response), probability of false alarms and sensitivity.

3. Paired Associates Learning (PAL). Paired Associates Learning assesses visual memory and new learning. Boxes are displayed on the screen and are “opened” in a randomised order. One or more of them will contain a pattern. The patterns are then displayed in the middle of the screen, one at a time and the participant must select the box in which the pattern was originally located. If the participant makes an error, the boxes are opened in sequence again to remind the participant of the locations of the patterns. Increased difficulty levels can be used to test high-functioning, healthy individuals. Outcome measures include the errors

made by the participant, the number of trials required to locate the pattern(s) correctly, memory scores and stages completed.

Tests to assess performance:

The Test of Performance Strategies-2 (TOPS-2) (Thomas, 1999). The TOPS was designed to assess eight psychological strategies used in competition (i.e. activation, automaticity, emotional control, goal- setting, imagery, negative thinking, relaxation and self-talk,) and eight used in practice (the same strategies except negative thinking is replaced by attentional control).

NASA Task Load Index (NASA-TLX) (Human Performance Research Group NASA Ames Research Center Moffett Field. California). The NASA-TLX is a multi-dimensional rating procedure that provides an overall workload score based on a weighted average of ratings on six subscales: Mental Demands, Physical Demands, Temporal Demands, Own Performance, Effort, and Frustration.

Statistical Methods:

To test the feasibility and efficacy of our Enhanced Stress Resilience Training curriculum in faculty surgeons and anesthesiologists, we designed a stepped-wedge trial (SWT) with randomized sequential intervention at each of three hospital campuses within the UCSF system. SWTs are a type of cluster randomized trial that is not equivalent to the gold-standard RCT, but is a reasonable alternative when a true control condition is not possible, as in a pragmatic trial setting. Based on published work in similar populations (Krasner, 2009), investigators would expect achieve 80% power to detect an effect size of 0.465 to 0.474 in 70 participants, examining a composite survey of well-being-related scales. Thus, based on preliminary needs assessments which showed close to 100 interested faculty, investigators expect to have adequate power for all analyses. Investigators will use linear mixed effects modeling for repeated measures in this longitudinal setting (Cognition, well-being and performance measures).