Big Data
Volume 3 Number 2, 2015

DOI: 10.1089/big.2014.0055

CASE STUDY

The U.S. Army Person-Event Data Environment: A Military–Civilian Big Data Enterprise

Loryana L. Vie,[1,2,*] Lawrence M. Scheier,[1,2] Paul B. Lester,[2] Tiffany E. Ho,[1,2]
Darwin R. Labarthe,[3] and Martin E.P. Seligman[1]

Abstract

This report describes a groundbreaking military–civilian collaboration that benefits from an Army and Department of Defense (DoD) big data business intelligence platform called the Person-Event Data Environment (PDE). The PDE is a consolidated data repository that contains unclassified but sensitive manpower, training, financial, health, and medical records covering U.S. Army personnel (Active Duty, Reserve, and National Guard), civilian contractors, and military dependents. These unique data assets provide a veridical timeline capturing each soldier's military experience from entry to separation from the armed forces. The PDE was designed to afford unprecedented cost-efficiencies by bringing researchers and military scientists to a single computerized repository rather than porting vast data resources to individual laboratories. With funding from the Robert Wood Johnson Foundation, researchers from the University of Pennsylvania Positive Psychology Center joined forces with the U.S. Army Research Facilitation Laboratory, forming the scientific backbone of the military–civilian collaboration. This unparalleled opportunity was necessitated by a growing need to learn more about relations between psychological and health assets and health outcomes, including healthcare utilization and costs—issues of major importance for both military and civilian population health. The PDE represents more than 100 times the population size and many times the number of linked variables covered by the nation's leading sources of population health data (e.g., the National Health and Nutrition Examination Survey). Following extensive Army vetting procedures, civilian researchers can mine the PDE's trove of information using a suite of statistical packages made available in a Citrix Virtual Desktop. A SharePoint collaboration and governance management environment ensures user compliance with federal and DoD regulations concerning human subjects' protections and also provides a secure portal for multisite collaborations. Taking similarities and differences between military and civilian populations into account, PDE studies can provide much more detailed insight into health-related questions of broad societal concern. Finding ways to make the rich repository of digitized information in the PDE available through military–civilian collaboration can help solve critical medical and behavioral issues affecting the health and well-being of our nations' military and civilian populations.

Key words: electronic medical records; epidemiology; military; Person-Event Data Environment; physical health; population health; soldier; well-being

Introduction to the Person-Event Data Environment

The U.S. military has compiled a vast, secure computer repository of digitized information that documents the full breadth of a service member's military experience. Modern record keeping of military experience—in computerized formats easily usable today—began in the late 1990s. Since then, the U.S. military has collected extensive service member information obtained from entrance exams (e.g., personality and aptitude), required annual physicals, pre- and post-deployment health assessments, medical and hospitalization treatment records, and periodic assessments of psychological functioning, job

[1]*Positive Psychology Center, University of Pennsylvania, Philadelphia, Pennsylvania.*
[2]*Research Facilitation Laboratory, Army Analytics Group, Monterey, California.*
[3]*Feinberg School of Medicine, Northwestern University, Evanston, Illinois.*

*Address correspondence to: *Loryana L. Vie, Research Facilitation Laboratory, 20 Ryan Ranch Road, Suite 170, Monterey, CA 93940,* E-mail: lvie@sas.upenn.edu

performance, and military service qualification tests. The Person-Event Data Environment (PDE) business intelligence platform is a cloud-based virtual data repository for housing this digitized information. The PDE was established by the U.S. Department of the Army to facilitate research and analysis of issues and policies that affect the military workforce—Active Duty, Reserve, National Guard, civilians, and contractors. Though research and analysis projects from all services are emerging in the PDE now, the system is primarily used by the Army, and therefore we will limit the current discussion to that service. In its current form, the PDE informs senior military and civilian government leadership and policy makers about a variety of human resource-related issues, including the health of the force, training program efficacy, and return on investment, selection, and attrition. Furthermore, the PDE addresses a critical need as the Army modernizes and engages novel behavioral health interventions to improve Army military service. Researchers can use the PDE to conduct program evaluation, cost-effectively pooling military data assets across numerous facets of a study. Findings of such studies can be highly informative for policy development and evaluation regarding health and healthcare issues in the civilian population as well.

Functionally, the PDE serves two central purposes: (1) acquire, integrate, and securely store data for Army-approved research projects, and (2) provide a secure, virtual workspace where approved researchers can access "sensitive" although unclassified Army military service, performance, manpower, and health data. Figure 1 depicts a computer screenshot of the PDE SharePoint environment, which serves as both a governance portal and a collaborative environment for researchers (both within and across projects).

The PDE was initiated in 2006 as a business intelligence platform with an initial goal emphasizing command workforce, critical skill resource assessment, and outcome studies. Other government organizations soon recognized the power of a collaborative "commons" offered by the PDE technology. Specifically, the Defense Manpower Data Center began contributing data and computational resources to the PDE, and other armed forces branches began approaching the Army for PDE access. Following its rapid expansion and utilization, the Army Human Research Protections Office (AHRPO) and the U.S. Army Public Health Command were recruited to help configure standardized governance procedures to ensure human subjects' protection and regulatory oversight. The Army has applied additional resources through the Research Facilitation Laboratory, a behavioral science unit created to help the commons repurpose PDE data for operational studies and to promote scientific advancement.

Opportunity

In 2011, the Robert Wood Johnson Foundation (RWJF) funded the University of Pennsylvania's Positive Psychology Center to create a sustainable military–civilian collaboration accessing PDE data assets. This *proof-of-concept* research project examines the role that psychological health assets (e.g., optimism and positive affect)

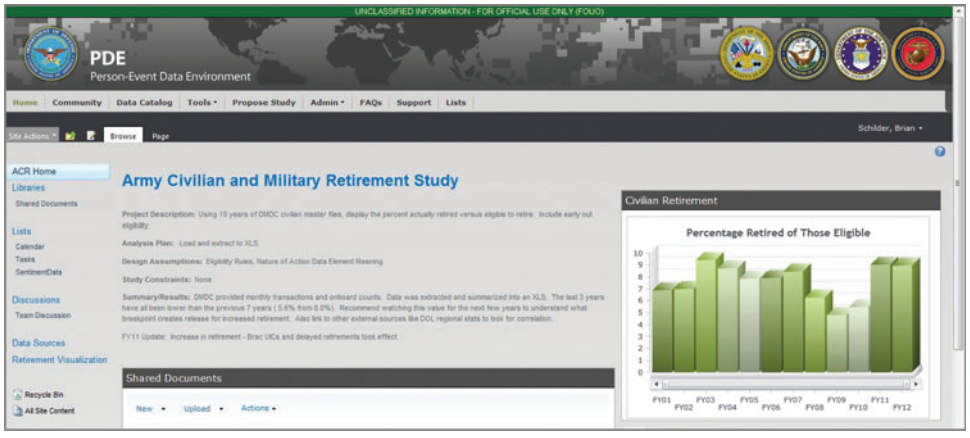

FIG. 1. Screenshot of the PDE SharePoint environment. Source: Reproduced with permission from the U.S. Department of the Army, Army Analytics Group, 2014.

play in various health outcomes and healthcare utilization.[1–3] The project research team consists of a steering committee involving both Army and academic scholars, a cadre of senior scientists, and project management on-site accessing data through secure Army portals. Collaboration between all these components has resulted in eight fundamental areas of scholarly inquiry: (1) mental health, (2) physical health outcomes (e.g., cardiovascular functioning), (3) attrition, (4) substance use, (5) deployment-related events (i.e., trauma and concussion-related traumatic brain injuries), (6) post-traumatic stress disorder, (7) criminal behavior, and (8) healthcare costs.

In addition to cost-efficiencies, the PDE provides useful alternatives to the many pressing challenges faced by behavioral and medical researchers. For one thing, national and longitudinal research studies are logistically cumbersome, relatively expensive, and time-consuming. Soldiers provide a vast amount of behavioral and medical data on a routine basis as part of regular monitoring, as well as ongoing personnel and manpower data for practical force considerations. The Army can use this information to detect mission-critical issues, including the preparedness of its forces, their suitability for combat, and the effects of Army service on soldier functioning, including manifestations of prolonged deployment.

The ability to analyze Army data on soldier performance under one computational "roof" is attractive for several other reasons. Cohorts are often limited to a particular geographic region or specific occupational group (e.g., nurses), limiting their overall generalizability. Soldiers, on the other hand, ostensibly come from all walks of life, from geographically diverse neighborhoods, represent a heterogeneous sociodemographic profile, and present for medical care with different behavioral and health maladies influenced by many factors, including family history, gender, race, age, occupation, and training regimen, to name a few. This variety alone provides a veritable-rich environment for studying human behavior and aggregates this information at the population level. A good deal of the knowledge gained from using the PDE will find application in comparable populations and, as a result, the collaborations will provide a cost-effective means through which scientists can create lasting solutions to new and continuing health problems.

Other opportunities arise from the continued monitoring of soldier performance, health, and psychosocial functioning. Disease registries typically recruit participants after a medical condition or ailment of interest is present, obscuring prospective markers with etiologic importance. Also, stressful or traumatic events are difficult to anticipate and, as a result, are often studied only retrospectively, resulting in inconsistent data collection efforts. Data stored in the PDE provide a plethora of information from initial accession through eventual separation from the Army, providing a means to monitor soldier functioning from emerging adulthood through later life for career soldiers. A bulk of the Army is relatively young (17–30 is the largest single demographic group and in 2014 constituted 64% of Active Duty soldiers) providing a unique trove of data that can inform social policy for years to come. From a life course perspective, this is a critical age group experiencing transition from adolescent to adult health, a point at which trends of unfavorable or continued favorable trajectories of adult health diverge. Policies for health promotion and disease prevention are therefore especially likely to be informed by new knowledge generated by investigations based on this unique data resource.

How Big Is Big?: Data Complexity and Survey Frequency

One pressing question is whether the Army PDE is really managing big data in the usual sense. The census of Army strength on an annual basis hover around 1.2 million soldiers including all three major components (Active Duty, Reserve, and National Guard). The PDE, which maintains an ongoing cumulative record of military personnel, is several times this size due to retention of some and constant recruitment of new members to the population. By any stretch of the imagination, and when compared to federal government data bases (e.g., Treasury Department, Social Security, and Labor Statistics), social networking sites (e.g., Twitter, Google, and Facebook), or commercial data processors of financial information (e.g., Heartland Payment Systems or Global Payment Systems), the PDE is not necessarily large. However, additional considerations perhaps qualify Army data as "big." For one thing, soldiers' data in the PDE is longitudinal following their military service from point of entry to discharge or normal termination (the latter including injury or death). Owing to the digitization of Army records from 1990 forward, this means that several million soldiers can be examined longitudinally in the PDE. Second, as we outline below, soldiers provide a wealth of data on many different facets of their training, health, and functioning. The health records alone include a vast array of information pertaining to doctor's visits, hospitalization, medication dispensed, and associated insurance data recording, among other items, the medical

reason for the visit, the diagnosis reported for billing purposes, and the cost of the visit to the medical treatment facility. Thus, the PDE can house upward of 10,000 or more variables collected on a single soldier accumulated through the duration of military service. This contrasts with the more limited financial and address information that commercial data processors possess on any particular individual. Finally, soldiers operate as individuals embedded in units as small as squads and as large as brigades or divisions. This "unit of observation" results in nested or hierarchically clustered data, owing to the similarity of soldier behavior within as opposed to across different intact social units. Nested data provide an opportunity to examine social influence processes and also whether certain behavioral training programs or leadership styles are more or less beneficial based on aggregate soldier profiles.

Another data complexity arises given the great variability when soldiers enter and leave the Army, deploy, and receive routine and nonroutine health assessments. To illustrate, PDE data assets consist of assessments gathered routinely (e.g., weapons qualifications), on a periodic basis (e.g., annual health assessments), as well as event-based (e.g., pre- and post-deployment health assessments). Although periodic health assessments (PHAs; annual physicals) are typically evenly spaced, routine assessments can be delayed based on the timing, location, and duration of a soldier's deployment. The lack of equal time intervals in the spacing between assessments can be methodologically challenging, particularly for researchers invested in modeling developmental change using time-structured assessments, although these challenges can be addressed statistically using alternative random coefficient modeling procedures.

Imposition of Sample Selection and Gating Criteria

The tremendous variability that exists in the timing of military health and behavioral assessments necessitates careful consideration of how to manage such large amounts of complex data. One solution has been to create replicable "gating criteria" that delimit samples. These take shape as Oracle Structured Query Language (SQL) queries to establish precisely which soldiers meet the time (during which military-related events may transpire) and assessment windows (which assessments are mandatory during a specified time frame). There are also service criteria, for instance, the precise component of service (Active Duty, Reserve, or National Guard) and whether a soldier deployed to Iraq or Afghanistan or some other region or combat zone. The ultimate objective of gating is to create a uniquely definable "cohort." These cohorts can then be examined using traditional longitudinal data mining approaches, including survival analysis, growth modeling, fixed-effect structural equation modeling, and ordinary least squares regression to ascertain effects over time. The gating criteria offer a way of managing the variability that characterizes soldier service records, while at the same time offering a rigorous means to detect periodicity and regularity often encountered with longitudinal data.

To illustrate, a recent study employed SQL queries to elucidate the contribution of deployment combat stress to alcohol consumption, medical symptoms, and a positive screen for posttraumatic stress disorder (PTSD). These are important areas for scientific inquiry as well as major costs drivers for the Army. Understanding these cost drivers can be explored through intelligent querying of the database, which involves specifying selective gating criteria that reveal interesting data patterns related to health and cost for such criteria.

As an example, the following gating criteria were applied sequentially to the database: (1) deployed between September 2012 and December 2014 (n = 154,340), (2) completed the required pre- and post-deployment health assessments (n = 46,176), (3) had taken a self-report survey assessing psychosocial functioning roughly a year prior to deployment (n = 14,294), and (4) were between the ages of 18 and 30 (n = 10,058). This netted a cohort of soldiers that had deployed during a specific time frame, were young adults, and had completed the required psychosocial and health assessments. The study was framed by learned helplessness theory, which states that depression and related mental illnesses result from perceived absence of control over the outcome of a situation and an exaggerated sense of helplessness or negativity.* In this study, an attributional tendency termed "negative explanatory style" was assessed using a self-report measure of "catastrophic thinking," which can be thought of as ruminating about worst case outcomes.†

To test linkages between combat stress and poor health outcomes among the panel soldiers we assessed the relative contribution of deployment combat-related stress, catastrophic thinking, and three demographic control measures (age, gender, and rank). Soldiers

*http://en.wikipedia.org/wiki/Learned_helplessness
†www.psychologytoday.com/blog/in-the-face-adversity/201103/catastrophic-thinking

who reported any one of seven combat-related stressors (e.g., "encountered dead bodies," "discharged a weapon," and experienced a blast") were coded accordingly. A measure of catastrophic thinking was obtained from a self-awareness tool soldiers take annually. Sample items included "when bad things happen to be, I expect more bad things to happen" and "when bad things happen to me, I cannot stop thinking about how much worse things will get," which were scored on a five-point Likert-type scale.

Outcome assessments were based on a 17-item PTSD symptom index[4] (cut-point of 30 or more indicated 'at risk'), a three-item alcohol misuse screener[5] (ranging from 0–12), and a 29-item index of general health concerns (ranging from 0–29). Controlling for age, rank, and gender the findings indicated that combat-related stress (trauma) was significantly related to a positive screen for PTSD (odds ratio [OR] = 3.11, 95% confidence interval [CI] = 2.37–4.08, $p < 0.0001$), alcohol misuse ($\beta = 0.07$, $p < 0.0001$), and more health concerns ($\beta = 0.21$, $p < 0.0001$), respectively. Catastrophic thinking also placed soldiers at risk for all three outcomes (PTSD: OR = 1.18, 95% CI = 1.02–1.37, $p < .05$; alcohol misuse: $\beta = 0.03$, $p < .01$; and health concerns: $\beta = .07$, $p < .0001$). Structured queries like the one used here: 1) reinforce the necessity of modeling heterogeneous measures of military experiences; and 2) highlight the potential need for tailoring current health promotion programs to address possible subgroup differences in military-specific outcomes.

Measures

From their initial point of entry into the Army (accession), soldiers provide continued (semiroutine) information on their health, psychological functioning, vocational aptitude, personality, fitness, and training qualification. There are ancillary data sources that track mandatory officer evaluations, military and civilian education, and soldiers who seek alternative training through special operations training and aviation schools. This wealth of data can provide a composite picture of a soldier's life and be used for operational studies or research purposes. Table 1 includes an overview of several key Army health data assets that are fundamental to the RWJF military–civilian collaboration. The table illustrates the basic content for each asset, the primary source of the health data (e.g., self-rated reports), and the administration sequence for each assessment (periodic or event-based).

As an illustration of how resourceful the PDE can be, we now present a detailed overview of one of the health data assets, the PHA. Active component and select Reserve personnel are required to receive an annual PHA. The PHA is a standardized preventive screening tool designed to improve the reporting and visibility of the individual medical readiness describing each soldier's physical ability to deploy. Specifically, this assessment consists of three integrated steps: (1) an online Health Risk Assessment (e.g., family history, medical conditions, and current medication use) with referrals made for laboratory studies and immunizations, (2) support staff review the personal medical information (e.g., height, weight, and medications), and (3) a healthcare provider reviews each soldier's statement of health, evaluates any required laboratory results, performs a medical symptom-focused exam, rates body system functioning using the medical physical profile serial qualification system, and provides referrals for additional medical services as indicated.

Over 600 data elements are collected during this three-step assessment process, approximately 107 of which contain administrative (e.g., "date PHA form approved") or personal information (e.g., soldier's telephone number). As we discuss in the following sections, personally identifying information can never be examined for research purposes. The health data in the PHA database can be categorized as follows: allergy information, 43 variables (e.g., reports allergy to iodine); behavioral health information, 43 variables (e.g., reports feeling down); clinical evaluation, 119 variables (e.g., diastolic blood pressure); overall health, 87 variables (e.g., soldier has chronic pain); family history, 115 variables (e.g., father had cancer); medications, 18 variables (e.g., "class of drug"); preventive health, 74 variables (e.g., frequency of alcohol use); and functional capacity, 18 variables (e.g., score for physical capacity or stamina). Across a five-year career, this represents over 2,500 soldier health data elements that can be gathered and made available to researchers as de-identified data. Even more impressive is the fact that these data elements can be merged with other longitudinal databases housed in the PDE, enabling researchers to examine contextual factors (e.g., psychosocial strengths, deployments, years of service, and job performance, to name a few) that may relate to health at specific points in time.

Utility of the PDE

Current use of the PDE generally falls into three categories: novel research, organizational analysis, and program evaluation. In terms of novel research, the PDE is

Table 1. Person-Event Data Environment health data assets

Database	*Description of contents*	*Primary source*	*Admin.*	*Pop.*
Deployment Health Assessments	Health before and after deployment (e.g., self-rated health, alcohol and tobacco use, PTSD, depression, combat exposure, injury and concussion risk, health concerns, major life stressors, Rx use, environmental and exposure concerns, suicide ideation, violence or potential for self-harm)	Self-rated & objective	Event-based	S
Digital Training Management System	Comprehensive training records (e.g., marksmanship training, predeployment training), physical fitness metrics (e.g., push-ups, sit-ups, two-mile run, participation in a weight control program)	Objective	Event-based & periodic	S
Drug & Alcohol Management Information System	Positive drug and alcohol screens (e.g., urinalysis, breathalyzer), referrals and enrollment in treatment, patient follow-up, and progress	Objective	Event-based	S
Electronic Physical Evaluation Board	Physical Evaluation Board key dates (e.g., date started, referral date, approval date, date placed on TDRL), disposition of the board, overall percentage of disability, description of condition	Objective	Event-based	S
Medical Data Repository	Electronic health records (e.g., appointment dates, Rx medications, procedures and surgeries, vitals [e.g., blood pressure], healthcare costs, pathology laboratory results)	Objective	Event-based	S, D
Periodic Health Assessment	Yearly physical assessments (e.g., overall health, clinical evaluation, medications, family history, behavioral health, preventive health, physical profile, deployability)	Self-rated & objective	Periodic	S
Social Security Admin. Death File	Death date, death verification code, last residence (state)	Objective	Event-based	S
ArmyFit—Global Assessment Tool	Psychological strengths (e.g., adaptability, positive/negative coping, catastrophizing, social engagement, optimism, meaning, character, depression, positive & negative affect, family satisfaction, family support, work engagement, friendship, loneliness, organizational trust), health, health-related behaviors (e.g., cigarette smoking)	Self-rated	Periodic	S, D, C
Theater Medical Data Store	Electronic health records during deployment (e.g., appointment dates, medications, procedures, injury type, illness diagnostic categories, symptoms, blood pressure, pulse rate, temperature)	Objective	Event-based	S

Admin., administration; C, DoD civilian; D, dependent; Pop., population; PTSD, posttraumatic stress disorder; Rx, prescription medication; S, soldier; TDRL, temporary disability retired list.

being used for a wide variety or research projects examining the health- and work-related behaviors of members of the military. For example, a recently published study using PDE data assets analyzed data from the Army's Global Assessment Tool (GAT) and found that high-performing soldiers tended to report relatively higher GAT scores on measures of psychosocial functioning (i.e., optimism), whereas soldiers with behavioral problems tended to have relatively lower GAT scores[6]; this study reinforced the practical utility of the GAT as a psychometric instrument. Other research studies in the PDE are underway now, including research designed to model military family resilience, feasibility studies for the use of social media data for suicide prevention, the development of risk algorithms for a range of behavioral problems, and research examining how leadership behaviors can influence follower psychological health.

There are also questions about population cardiovascular health and the different approaches to investigating these important medical considerations that offer another example of the utility and significance of the PDE for conducting novel research. As previously stated, the Army is a large, racially and ethnically heterogeneous population with diverse age groups from as young as 17 through later middle life (ages 60–70). As such, the extensive inventory of personal-level data on cardiovascular health status, health behaviors, psychological factors, and social determinants of health provides an exceptionally rich existing data set

regarding relations among these health factors and potential strategies for cardiovascular disease prevention. Record linkage for individuals throughout their military careers provides opportunities for longitudinal as well as cross-sectional examination of these relationships. Health-related policy changes within the Army can be proposed, developed, implemented, and evaluated through ongoing investigations with these data. The applicability of findings to the civilian population and the opportunity to make direct statistical comparisons to the civilian population add further value to this exceptional health data asset.

For example, one line of effort has recently examined PHA data in the PDE to contrast measures of health among Army Active Duty, Reserve, and National Guard soldiers, and civilians from the National Health and Nutrition Examination Survey. Appendix I provides a case study that showcases comparisons on several health metrics that are major cost drivers for both the Army and general population.

Additionally, the PDE is used to perform a range of organizational analyses to answer questions posed by the Department of Defense's (DoD's) senior leadership as well as members of the U.S. Congress. The PDE was recently used to respond to inquiries from Congress about demographic characteristics of service members within the DoD. In addition, the PDE has great utility for the development and pilot testing of new standardized assessment forms and behavioral instruments. Researchers can follow standard psychometric procedures and examine different forms of validity, including concurrent and factorial (using existing data), predictive (with prospective data), and convergent and discriminant (with other PDE data assets).

Finally, the PDE is used to perform large-scale program evaluation on a range of DoD-related military programs. For example, there have been calls in the literature for more rigorous evaluation of the Army's Comprehensive Soldier & Family Fitness (CSF2) program—a psychological health and resilience training program.[7,8] The PDE has made it possible to securely bring together data on a range of soldier outcomes, begin evaluating the CSF2 Master Resilience Training (MRT) program, and respond to these clarion calls with evidence-based findings. Specifically, researchers have examined longitudinally participation in the MRT program and subsequent ratings of psychosocial strengths and assets.[9] Follow-up work examined associations between participation in the MRT program and its effect on the prevalence of diagnoses for mental health or substance abuse problems.[10] Subsequent research will extend to other outcomes, including health ratings, job performance, and healthcare costs, to name a few. The PDE has also been used to evaluate other Army programs, including the Army Surgeon General's Performance Triad—a program designed to promote healthy sleep, physical activity, and nutrition behaviors in soldiers. For researchers in the PDE evaluating military programs using observational studies or quasi-experimental designs, use of the propensity score method (which assesses the likelihood of being assigned to the treatment group based solely on one's demographic or covariate information) strengthens the ability to make causal inferences even in the absence of a randomized control group.[11]

Compliance with Human Subjects' Protections

Extensive measures are taken to protect the confidentiality and personal identity of soldiers whose information is part of the digitized resources housed in the PDE. As part of protecting personally identifying information (PII), social security numbers undergo a two-step transformation and encoding process, which results in the assignment of a random 12-character alphanumeric "key" to each soldier. Data in the PDE can then be merged and linked via the randomly generated "keys" in order to create linked files from multiple databases and time points in support of Army- and DoD-approved research. The governance process that creates identification keys relies on physically and logically separate computer systems with secure Army and DoD firewalls using a VPN connection. In addition, personnel responsible for the extraction, transfer, and load of de-identified data are federally approved contractors who undergo extensive Health Insurance Portability and Accountability Act (HIPAA) training and also work in a secure environment.

Additional transformations for limited data sets containing protected health information (PHI) include truncating birth dates so that only year or month and year are available. A soldier's unit identification code, rank, and pay grade are also transformed for PDE research studies. As outlined below, each transformation of PII is designed to reduce the risk of a soldier being re-identified by a researcher, while maintaining enough information for standard aggregate statistical analysis and longitudinal record linkage.

Accessing Medical and Health Data

Access to medical and health data is covered under the HIPAA (45 CFR Subpart 46, PL 104-191) and the

Privacy Act of 1974 (Pub. L. No. 93-579, and its subsequent amendments USC Sec. 552a, Title 5, Part I, Chapter 5, Subchapter II). Both statutes carefully delineate the safeguarding of PHI and the manner in which "limited data set" identifiers can be disclosed (e.g., birthdates are transformed to MMYYYY).

All preparation of personal health and medical data in the PDE must comport with the Standards of De-Identification of Protected Health Information (Section 164.514[b][2] of the 1974 Privacy Act). The latter requirement involves establishing compliance with either the safe harbor or the expert determination method. Both methods ensure compliance with federal standards that essentially mitigate privacy risks pertaining to sharing PHI between covered entities (i.e., health insurers) and outside parties. The former procedure requires removal of 18 limited data set identifiers (e.g., name, address, e-mail, driver license, or other unique identifiers) from PHI in order to reduce the potential of "re-identification." The latter method relies on scientifically valid statistical audit procedures designed to evaluate the potential risk for re-identification (disclosure) given the proposed de-identification procedures.

Projects deemed research involving human subjects must also undergo an external scientific review using an Institutional Review Board and vetting by the AHRPO, which provides regulatory oversight for human subjects' protection of soldiers. All of these assurances and regulatory requirements are detailed in DoD Information (3216.02 "Protection of Human Subjects and Adherence to Ethical Standards in DoD-Supported Research") and Department of the Army (e.g., AR 70-25 "Use of Volunteers as Subjects of Research") guidelines. Applications for AHRPO human subjects and regulatory approval rely on the Force Health Protection and Readiness IRBNet portal, which is part of the Defense Medical Research Network. The Medical Research and Materiel Command website provides documentation of the Army procedures and applicable DoD regulatory requirements for human subjects' protection.

Conclusions

The RWJF military–civilian collaboration paves the way to methodically and incrementally open the PDE access aperture over time, thereby melding the enormous data assets of the Army with top research scientists from private commercial ventures and university-based settings. This scale-up requires blending the needs of researchers with the operational features of the PDE, all the while ensuring the protection and confidentiality of personal and health information obtained from the individuals tasked with defending our country. The PDE offers an unprecedented resource to the scientific community, and it is quickly becoming the most extensive collection of digitized information on this important population or any other population we know of, given its tremendous breadth and depth.

In addition, the PDE is also moving in the direction of creating metadata resources to document the various DoD and Army data assets. This will include archived institutional history that describes the evolution of soldier assessment forms, version and content changes in surveys, data management concerns (e.g., variable coding and transformation), and details on data collection methods and parameters describing test administration. Future SharePoint capabilities will enable members of the PDE research community to record comments on data assets and their elements (e.g., indicate whether data fields are incomplete or have different values than expected). This information will prove quite valuable to subsequent researchers and will help build a more efficient "commons" research process. The PDE can also draw upon resources provided by a DoD Metadata Registry, managed by the Defense Information Systems Agency, a web-based repository that promotes interoperability and reuse of computer technology (e.g., data models, symbologies, transformations, and schemas) among military department and defense agencies.

What the Future Holds

Given its relatively recent inception, the PDE has yet to reach full operational capability. Rapid growth of the PDE requiring greater bandwidth, procurement of sufficient "seat" licenses for commercial statistical packages, and the computing power required to manage and analyze large complex data structures currently limit the number of users the PDE can host. Resolving these bandwidth and related operational limitations will be part of getting the PDE to full operational capability. There is also an effort underway to "automate" much of the governance of the PDE, including security procedures for vetting end users, conducting background checks, and ensuring that study research goals are compliant with data use agreements. Every study in the PDE has to maintain current documentation of individual researchers' human subjects' protections compliance, adherence to PDE governance, and DoD Information Assurance certification. As essential as this process is for operational logistics, it can also

be cumbersome. In addition, data assets are not released unless checked manually to ensure release comports with Army Data Use Agreements and, in the case of medical data, meets the requirements specified in Data Sharing Agreements. This process is crucial to maintaining soldier confidentiality as well as regulatory compliance. Furthermore, the Army is also undergoing rapid changes in the types of platforms used to gather soldier data. As an example, a new online platform "ArmyFit" is now gathering data on soldier (and spouse) fitness, including nutritional information, sleep, and physical exercise. This is part of the Army Resilience Directorate's mandate to include emotional, social, spiritual, family, and physical fitness dimensions as part of routine assessments of soldier functioning. The advent of new web-accessible platforms collecting routine information on almost 50,000 soldiers each month means new soldier data (with unique coding formats) are constantly streaming into the PDE, broadening the capability of researchers to track emerging epidemiological trends.

We specifically note that the military–civilian collaboration will reap untold opportunities for researchers, who will gain access to unique and very extensive, prospective data on a very large population of Army soldiers. This will enable them to examine population-based trends in a wide range of health-related behaviors and conditions with important implications for society at large. Likewise, the Army will benefit from the expertise of leading behavioral and medical scientists interested in measuring and improving soldier performance and health, with insights of great potential value for population health more generally.

Acknowledgments

We would like to thank the numerous Army, Department of Defense, and University of Pennsylvania entities that have worked tirelessly to ensure the success of the military–civilian collaboration. It is hoped that the fruits of this initial effort can have positive and lasting effects on soldiers' lives and eventually find ways to inform Army leadership of critical health issues. In particular, Jenny L. English, health statistician, U.S. Army/PASBA-MEDCOM, and Audrey L. Luken, MEDPROS program manager, G3-7 Medical Readiness Division, U.S. Army Medical Command, provided critical information on the data assets described in this profile.

Funding

The PDE is funded by the Departments of Defense, Army, Navy, and Air Force. Support for this publication is provided in part by the Robert Wood Johnson Foundation through a grant to the Positive Psychology Center of the University of Pennsylvania, Martin E.P. Seligman, principal investigator. Paul B. Lester receives support from the Research Facilitation Laboratory, Army Analytics Group, Office of the Deputy Under Secretary of the Army.

Author Disclosure Statement

The authors had no conflicts of interest with respect to the authorship or the publication of this article. The University of Pennsylvania has a proprietary interest in the Master Resilience Training (MRT) program. Loryana L. Vie, Lawrence M. Scheier, Tiffany E. Ho, and Martin E.P. Seligman did not analyze MRT data in the PDE or take part in discussions of intervention data related to the MRT program. Paul B. Lester is the Federal Government Sponsor for research activities in the PDE. He provides governance oversight as well as technical assistance obtaining various U.S. Army data assets, including the GAT data from the ArmyFit platform. The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of the Army, Department of Defense, or the U.S. government.

References

1. Seligman MEP, Csikszentmihalyi M. Positive psychology. An introduction. Am Psychol 2000; 55:5–14.
2. Seligman MEP. Positive health. Appl Psychol Int Rev 2008; 57:3–18.
3. Peterson C, Bossio LM. Optimism and physical well-being. In: Chang EC, editor. Optimism & Pessimism: Implications for Theory, Research, and Practice. Washington, DC: American Psychological Association, 2001, pp. 127–145.
4. Blanchard EB, Jones-Alexander J, Buckley TC, Forneris CA. Psychometric properties of the PTSD checklist (PCL). Behavioral Research & Therapy 1996; 34:669–673.
5. Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions (AUDIT-C): An effective brief screening test for problem drinking. Archives of Internal Medicine 1998; 158:1789–1795.
6. Lester PB, Harms PD, Herian MN, et al. A force of change: Chris Peterson and the US Army's Global Assessment Tool. J Posit Psychol 2014; 10: 7–16.
7. Smith SL. Could Comprehensive Soldier Fitness have iatrogenic consequences? A commentary. J Behav Health Ser R 2013; 40:242–246.
8. Steenkamp MM, Nash WP, Litz BT. Post-traumatic stress disorder: Review of the Comprehensive Soldier Fitness program. Am J Prev Med 2013; 44:507–512.
9. Lester PB, Harms PD, Herian MN, et al. The Comprehensive Soldier Fitness Program Evaluation Report #3: Longitudinal Analysis of the Impact of Master Resilience Training on Self-Reported Resilience and Psychological Health Data. United States Army Comprehensive Soldier Fitness Program. Washington, DC: U.S. Government Printing Office, 2011.
10. Harms PD, Herian MN, Krasikova DV, et al. The Comprehensive Soldier and Family Fitness Program Evaluation Report #4: Evaluation of Resilience Training and Mental and Behavioral Health Outcomes. United States

Army Comprehensive Soldier Fitness Program. Washington, DC: U.S. Government Printing Office, 2013.
11. Scheier LM. Methods for approximating random assignment: Regression discontinuity and propensity scores. In: Baker E, Peterson PP, McGaw B, editors. International Encyclopedia of Education (3rd ed.). London: Elsevier, 2010, pp. 104–110.

Cite this article as: Vie LL, Scheier LM, Lester PB, Ho TE, Labarthe DR, Seligman MEP (2015) The U.S. Army Person-Event Data Environment: a military–civilian big data enterprise. *Big Data* 3:2, 67–79, DOI: 10.1089/big.2014.0055.

Abbreviations Used

AHRPO = Army Human Research Protections Office
CSF2 = Comprehensive Soldier & Family Fitness
DoD = Department of Defense
GAT = Global Assessment Tool
HIPAA = Health Insurance Portability and Accountability Act
MRT = Master Resilience Training
PDE = Person-Event Data Environment
PHA = Periodic Health Assessment
PHI = protected health information
PII = personally identifying information
RWJF = Robert Wood Johnson Foundation
SQL = Structured Query Language

(*Appendix follows*→)

Appendix I: Case Study

Below we present one example of how Person-Event Data Environment (PDE) data can be used to generate a "report card" on several major cost drivers for the military as well as the U.S. civilian population. In this project, researchers from the Robert Wood Johnson Foundation military–civilian collaboration contrasted five metrics obtained from Active Duty and Reserve/National Guard soldiers with data from the 2012 National Health and Nutrition Examination Survey (NHANES).[A1] The NHANES is one of several nationally representative general population studies that provide valid and reliable measures of health and psychosocial functioning in the United States, and it represents the largest ongoing individual-level health examination survey (other examples include the Behavioral Risk Factor Surveillance System[A2] or the Midlife in the United States Study[A3]).

The five metrics selected for illustration are heavy cigarette use (transforming number of cigarettes in the past 30 days into the equivalent of a pack or more: "21 or more cigarettes per day"); heavy alcohol consumption (three AUDIT-C[A4] items assessing alcohol frequency ["How often did you have a drink with alcohol?"], intensity ["How many drinks did you have on a typical day when you were drinking?"] and binge drinking ["How often did you have 6 or more drinks on one occasion?"]); depression (using the 9-item PHQ-9,[A5] a general population depression screener assessing depressed mood or irritability, decreased interest or pleasure, significant weight change or change in appetite, change in sleep, change in activity, fatigue or loss of energy, feelings of guilt or worthlessness, diminished concentration, and suicidal tendencies); physician care (seeing a medical practitioner over the past year); and hospitalization (whether the respondent had been hospitalized within the past year or, for soldiers, since their last annual Army physical).

Table A1 shows the demographic comparison between the Army and civilian samples. The sample size in the

Table A1. Demographic characteristics of military personnel and civilians (2012)

Characteristic	*Active Duty soldiers* (n = *265,525*) n *(%)*	*Reserve/National Guard soldiers*[a] (n = *398,240*) n *(%)*	*Civilians*[b] (n = *4,854*) n *(%)*
Gender			
Male	224,767 (84.65)	327,074 (77.51)	2,403 (49.51)
Female	40,758 (15.35)	71,166 (17.87)	2,451 (50.49)
Age group, years			
17–29	150,700 (56.76)	209,745 (52.67)	1,441 (29.69)
30–39	78,125 (29.42)	95,467 (23.97)	963 (19.84)
40–49	33,312 (12.55)	70,044 (17.59)	899 (18.52)
50–65	3,388 (1.28)	22,989 (5.77)	1,551 (31.95)
Race/ethnicity			
Hispanic	28,553 (10.75)	40,239 (10.10)	1,062 (21.88)
White	157,773 (59.42)	268,861 (67.51)	1,566 (32.26)
Black	58,111 (21.89)	65,742 (16.51)	1,331 (27.42)
Asian	13,687 (5.15)	14,257 (3.58)	742 (15.29)
Other, including multiracial	7,401 (2.79)	9,133 (2.29)	153 (3.15)
Education			
No high school diploma	1,245 (0.47)	10,371 (2.63)	1,160 (23.90)
High school diploma or equivalent	195,187 (74.04)	286,588 (72.66)	1,013 (20.87)
Some college	11,941 (4.53)	17,671 (4.48)	1,491 (30.72)
College degree and higher	55,249 (20.96)	79,802 (20.23)	1,190 (24.52)
Marital status			
Never married	89,403 (33.68)	186,554 (46.88)	2,515 (57.12)
Married	159,616 (60.14)	182,438 (45.85)	1,136 (25.80)
Separated/divorced/widowed	16,397 (6.18)	28,932 (7.27)	752 (17.08)
Length of service (years)			
0–3	90,348 (34.03)	97,493 (24.48)	N/A
4–8	64,907 (24.44)	115,742 (29.06)	N/A
9–15	57,496 (21.65)	74,841 (18.79)	N/A
≥16	52,774 (19.88)	110,169 (27.66)	N/A

[a]Army Reserve and National Guard soldiers differ only in the source of their pay. (The National Guard receives pay compensation from the federal budget, but they are organized and run by the individual states. Army Reservists receive compensation directly from the federal budget.) Otherwise their standards of performance and required training programs are identical and both service branches can deploy if needed. Therefore, for the purpose of this article, these two groups were combined.

[b]Civilians are composed of a nationally representative sample from the NHANES (2011–2012).

Table A2. Adjusted and weighted* means and period prevalence for health metrics among military personnel and civilians (2012)

	Heavy cigarette use[†] *(% yes)*			*Heavy alcohol consumption (AUDIT-C score)*			*Depression severity (PHQ-9 score)*			*Visited healthcare provider*[††] *(% yes)*			*Hospitalized*[‡] *(% yes)*		
	AD	*Res./NG*	*Civ.*	*AD*	*Res./NG*	*Civ.*	*AD*	*Res./NG*	*Civ.*	*AD*	*Res./NG*	*Civ.*	*AD*	*Res./NG*	*Civ.*
Age															
17–29	3.25[a]	3.57[b]	0.40[c]	2.11[a]	1.94[a]	2.96[b]	1.38[a]	0.95[b]	3.33[c]	29.88[a]	25.35[b]	80.35[c]	11.64[a]	8.21[b]	8.37[c]
30–39	4.82[a]	5.81[b]	1.88[c]	2.32[a]	2.14[a]	2.97[b]	2.41[a]	1.90[a]	4.03[b]	48.34[a]	37.25[b]	74.63[c]	16.60[a]	10.30[b]	10.01[c]
40–49	7.08[a]	8.64[a]	4.06[b]	1.79[a,b]	1.66[a]	2.54[b]	3.21[a,b]	2.52[a]	4.10[b]	61.96[a]	48.48[b]	83.14[c]	19.27[a]	12.06[b]	9.84[c]
50–65	7.77[a]	9.62[a]	3.04[b]	1.10[a]	1.11[a]	1.81[a]	3.08[a]	2.13[a]	3.78[a]	71.84[a]	56.36[b]	87.08[c]	21.64[a]	14.71[b]	10.39[c]
Gender															
Male	4.22[a]	5.34[a]	3.44[b]	2.49[a]	2.33[a]	3.40[b]	1.50[a]	0.99[b]	2.96[c]	37.26[a]	30.79[b]	75.54[c]	12.96[a]	8.37[b]	6.71[c]
Female	1.88[a]	2.67[a]	1.24[b]	1.60[a]	1.52[a]	1.86[a]	2.05[a]	1.39[a]	4.47 [b]	54.23[a]	49.07[b]	88.25[c]	20.90[a]	16.18[b]	12.50[c]
Race/ethnicity															
Hispanic	1.84[a]	1.88[b]	0.33[c]	1.63[a]	1.62[a]	2.87[b]	2.01[a]	1.57[a]	3.81[b]	35.32[a]	28.91[b]	70.74[c]	13.49[a]	8.99[b]	10.34[c]
White	5.06[a]	5.92[b]	3.17[c]	2.38[a]	2.33[a]	3.21[b]	1.99[a]	1.42[a]	3.67[b]	40.64[a]	35.12[b]	84.83[c]	14.28[a]	9.97[b]	9.15[c]
Black	1.71[a]	1.60[a]	0.66[b]	1.93[a]	1.93[a]	2.72[b]	2.27[a]	1.87[a]	3.67[b]	40.46[a]	34.43[b]	83.58[c]	14.61[a]	9.92[b]	12.85[c]
Asian	1.93[a]	2.62[b]	0.14[c]	2.29[a]	2.15[a]	1.83[a]	1.27[a]	0.76[a]	2.50[b]	36.74[a]	26.85[b]	78.99[c]	12.64[a]	7.33[b]	5.39[c]
Other	2.43[a]	2.80[a]	5.69[b]	3.13[a]	2.74[a]	3.59[a]	3.46[a]	2.69[a]	5.95[b]	42.15[a]	34.30[b]	80.66[c]	14.32[a]	9.85[a]	11.51[b]
Education															
No HSD	5.73[a]	4.20[b]	5.62[c]	1.83[a]	1.85[a]	2.84[a]	3.86[a]	5.56[a]	2.47[a]	34.38[a]	26.69[b]	72.76[c]	14.25[a]	8.10[b]	12.97[c]
HSD or equiv.	4.04[a]	5.05[b]	1.70[c]	1.92[a]	1.76[a]	2.64[b]	2.27[a]	1.50[b]	4.07[c]	36.18[a]	30.91[b]	78.39[c]	13.88[a]	9.44[b]	8.83[c]
Some college	3.77[a]	5.08[a]	2.41[b]	1.91[a]	1.63[a]	2.67[a]	1.90[a]	1.33[a]	3.43[a]	51.52[a]	42.75[b]	82.55[c]	16.66[a]	11.56[b]	9.39[c]
College degree or greater	2.63[a]	3.95[a]	0.77[b]	1.98[a]	1.88[a]	2.73[b]	1.21[a]	0.92[a]	2.37[b]	50.56[a]	44.40[b]	89.00[c]	14.74[a]	10.77[b]	8.62[c]
Marital status															
Never married	4.58[a]	6.28[a]	2.73[b]	1.82[a]	1.67[a]	2.40[b]	1.64[a]	1.35[a]	2.85[b]	45.76[a]	41.05[b]	82.78[c]	16.37[a]	11.56[b]	10.08[c]
Married	2.80[a]	3.57[b]	1.04[c]	1.45[a]	1.41[a]	2.91[b]	1.60[a]	1.32[a]	4.57[b]	27.01[a]	25.61[b]	79.22[c]	9.39[a]	7.60[b]	8.28[c]
Sep./Div./Widow	5.04[a]	7.12[a]	3.34[b]	2.23[a]	1.89[a]	2.51[a]	2.35[a,b]	1.95[a]	4.04[b]	52.59[a]	44.50[b]	82.47[c]	19.03[a]	12.40[b]	11.13[c]

*Means are adjusted for all variables in the table. Means and period prevalence rates are weighted for NHANES complex sampling design and nonresponse. Comparisons were conducted using ANCOVA and logistic regression with linear combinations.

[†]Heavy cigarette smoking was assessed for those who currently smoke and is characterized by smoking at least 21 cigarettes a day.

[††]Soldiers were asked if they had seen a healthcare provider since their last military examination. Participants in NHANES were asked if they had seen a healthcare provider in the last year.

[‡]Soldiers were asked if they had been hospitalized or had surgery since their last military examination (55% had conducted their prior periodic health assessment within a 15-month window, 30% between up to 2 years, 13% between 2 and 3 years, and 2% beyond this window). Participants in NHANES were asked if they had been a patient in a hospital overnight in the past year.

[a,b,c,d]Raised superscript letters that are different indicate statistically significant differences ($p < 0.05$) of adjusted and weighted means. Same letters indicate no statistically significant differences in means. Tukey's honest significance test method was used to adjust for multiple comparisons and the increased probability of making false-positive type I errors.

AD, Active Duty soldiers; Civ., Civilian (NHANES); HSD, high school diploma; NHANES, National Health and Nutrition Examination Survey; Res./NG, Reserve and National Guard soldiers; Sep./Div./Widow, Separated/Divorced/Widowed.

Table A3. Multinomial logistic regression predicting group membership from demographic and health metrics

	Active Duty vs. civilian[a]		*Reserve/National Guard vs. civilian*		*Overall likelihood-ratio test*[b]
Metric	*OR*	*CI*	*OR*	*CI*	p
Age	0.918	(0.918–0.918)	0.948	(0.948–0.948)	< 0.0001
Gender[c]	5.475	(5.412–5.539)	3.302	(3.273–3.331)	< 0.0001
Race[c]	1.722	(1.708–1.737)	2.521	(2.504–2.539)	< 0.0001
Education[c]	0.223	(0.221–0.225)	0.199	(0.198–0.201)	< 0.0001
Marital status[c]	2.222	(2.203–2.242)	0.898	(0.891–0.904)	< 0.0001
Cigarette use[d]	0.840	(0.835–0.844)	0.825	(0.821–0.829)	< 0.0001
Heavy alcohol consumption	0.808	(0.807–0.810)	0.818	(0.817–0.819)	< 0.0001
Depression severity	0.824	(0.822–0.826)	0.747	(0.745–0.749)	< 0.0001
Seen healthcare provider[c]	0.228	(0.226–0.230)	0.175	(0.174–0.176)	< 0.0001
Hospitalized[c]	4.768	(4.709–4.827)	3.150	(3.114–3.186)	< 0.0001

[a]The multinomial regression simultaneously tests two logit models comparing one of the three groups against the reference category, which is the Civilian-NHANES group.

[b]The overall-likelihood test is analogous to testing whether or not all two separate ORs are significantly different from an OR of 1.0, which shows that the predictor does not efficiently discriminate group membership.

[c]Reference categories: female; non-white; high school diploma or less; non-married; did not see healthcare provider; not hospitalized.

[d]Variable transformed for this analysis to include nonsmokers (0 = 0 cigarettes/day, 1 = ≤ 10, 2 = 11–20, 3 = 21–30, 4 = 31 +).

OR, odds ratio; CI, 95% confidence interval.

PDE is more than 100 times that of NHANES, most conspicuously so for race-ethnic groups other than white. The military population is younger, less highly educated, and more often married than the NHANES population sample. Notably, the Active Duty and Reserve/National Guard Soldier populations are generally quite similar demographically, supporting their potential pooling for many research purposes. Table A2 contains the results of group mean comparisons for the five metrics outlined above. A careful inspection shows that, adjusted for all other covariates in the model, soldiers consistently reported higher rates of heavy cigarette smoking compared to civilians, and this pattern held for gender, race, education, and marital status subgroups. Civilians reported heavier alcohol consumption, and this pattern held with few exceptions across the different subgroups (Asian and other race groups were not different). Up through age 40–49, higher rates of depression were reported by the civilian population. This pattern held across most of the demographic subgroups with the exception of less educated civilians who reported fewer symptoms. Active Duty and Reserve/National Guard soldiers were less likely to have visited a healthcare provider during the study period compared to civilians; Active Duty and Reserve/National Guard soldiers were more likely to have had surgery or been hospitalized in the last year compared to civilians; the latter may reflect the occupational risks experienced by this group.

Table A3 contains odds ratios and confidence intervals obtained from a multinomial logistic regression model. This model provides information on the relative efficiency of the demographic and health-related predictors to differentiate group membership (reference group is the NHANES civilian population). Odds ratios less than 1.0 indicate higher likelihood of being a member of the civilian population. These results, adjusted for the other health behaviors and demographics, show a fairly consistent pattern reinforcing the lower rates of heavy alcohol consumption, depressive symptoms, and utilization of healthcare providers, and the much higher rates of being hospitalized among all soldiers compared to civilians. Overall, this illustration demonstrates the applicability of the PDE to questions of population health important for both the military and civilians. Analysis of epidemiologic patterns within the military can clearly inform health issues, such as those shown, for military health policy. Findings from studies of military samples can have tremendous bearing on the knowledge of civilian health, particularly when variable definitions are closely aligned and variables can be compared directly.

Appendix References

A1. Ferketich AK, Schwartzbaum JA, Frid DJ, Moeschberger ML. Depression as an antecedent to heart disease among women and men in the NHANES I study. Arch Intern Med 2000; 160:1261–1268.

A2. Pierannunzi C, Hu SS, Balluz L. A systematic review of publications assessing reliability and validity of the Behavioral Risk Factor Surveillance System (BRFSS), 2004–2011. BMC Med Res Methodol 2013; 13:49.

A3. Mroczek DK, Kolarz CM. The effect of age on positive and negative affect: A developmental perspective on happiness. J Pers Soc Psychol 1998; 75:1333–1349.

A4. Bush K, Kivlahan DR, McDonell MZB, et al. The AUDIT alcohol consumption questions (AUDIT-C): An effective brief screening test for problem drinking. Arch Intern Med 1998; 158:1789–1795.

A5. Kroenki K, Spitzer RL, Williams JBW. The PHQ-9: Validity of a brief depression severity measure. JGIM 2001; 16:606–613.