



HHS Public Access

Author manuscript

J Psychiatr Res. Author manuscript; available in PMC 2018 January 01.

Published in final edited form as:

J Psychiatr Res. 2017 January ; 84: 128–136. doi:10.1016/j.jpsychires.2016.09.028.

Using administrative data to identify U.S. Army soldiers at high-risk of perpetrating minor violent crimes

Anthony J. Rosellini^a, John Monahan^b, Amy E. Street^{c,d}, Eric D. Hill^a, Maria Petukhova^a, Ben Y. Reis^e, Nancy A. Sampson^a, David M. Benedek^f, Paul Bliese^g, Murray B. Stein^{h,i}, Robert J. Ursano^f, and Ronald C. Kessler^a

^aDepartment of Health Care Policy, Harvard Medical School, Boston, Massachusetts, USA

^bSchool of Law, University of Virginia, Charlottesville, Virginia, USA

^cNational Center for PTSD, VA Boston Healthcare System, Boston, Massachusetts, USA

^dDepartment of Psychiatry, Boston University School of Medicine, Boston, Massachusetts, USA

^ePredictive Medicine Group, Boston Children's Hospital and Harvard Medical School, Boston, Massachusetts, USA

^fCenter for the Study of Traumatic Stress, Department of Psychiatry, Uniformed Services University School of Medicine, Bethesda, Maryland, USA

^gDarla Moore School of Business, University of South Carolina, Columbia, South Carolina, USA

^hDepartments of Psychiatry and Family Medicine & Public Health, University of California San Diego, La Jolla, California, USA

ⁱVA San Diego Healthcare System, San Diego, California, USA

Correspondence: Ronald C. Kessler, Ph.D., Department of Health Care Policy, Harvard Medical School, Boston, Massachusetts, USA. Kessler@hcp.med.harvard.edu; voice 617-432-3587; fax 617-432-3588.

Disclaimer

The views, opinions and/or findings contained in this research are those of the authors and do not necessarily reflect the views of the Department of the Army, Department of Defense, Department of Health and Human Services, or NIMH and should not be construed as an official DoD/Army position, policy or decision unless so designated by other documentation. No official endorsement should be made.

Contributors:

Dr. Kessler was the lead investigator of the project. Dr. Kessler, Dr. Stein, and Dr. Ursano were responsible for the design of the Army Study to Assess Risk and Resilience in Servicemembers (the source of the data used here). Dr. Rosellini, Dr. Monahan, Dr. Reis, Mrs. Sampson, and Dr. Kessler formulated the analysis plan. Mr. Hill and Dr. Petukhova were responsible for data analysis and provided critical comments on drafts of the paper. Dr. Rosellini and Dr. Kessler wrote the first draft of the paper and prepared all tables. Dr. Monahan, Dr. Street, Dr. Benedek, Dr. Bliese, Dr. Stein, and Dr. Ursano provided feedback on the first draft and critical revisions of the manuscript. All authors contributed to and have approved the final manuscript.

Conflict of interest

Dr. Monahan is a co-owner of the Classification of Violence Risk (COVR), Inc. Dr. Stein has been a consultant for Care Management Technologies, received payment for his editorial work from UpToDate and Depression and Anxiety, and had research support for pharmacological imaging studies from Janssen. In the past three years, Dr. Kessler has been a consultant for Hoffman-La Roche, Inc. and Johnson & Johnson Wellness and Prevention. Dr. Kessler has served on advisory boards for Mensante Corporation, Johnson & Johnson Services Inc. Lake Nona Life Project, and U.S. Preventive Medicine. Dr. Kessler is a co-owner of DataStat, Inc. The remaining authors declare nothing to disclose.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Abstract

Growing concerns exist about violent crimes perpetrated by U.S. military personnel. Although interventions exist to reduce violent crimes in high-risk populations, optimal implementation requires evidence-based targeting. The goal of the current study was to use machine learning methods (stepwise and penalized regression; random forests) to develop models to predict minor violent crime perpetration among U.S. Army soldiers. Predictors were abstracted from administrative data available for all 975,057 soldiers in the U.S. Army 2004–2009, among whom 25,966 men and 2,728 women committed a first founded minor violent crime (simple assault, blackmail-extortion-intimidation, rioting, harassment). Temporally prior administrative records measuring socio-demographic, Army career, criminal justice, medical/pharmacy, and contextual variables were used to build separate male and female prediction models that were then tested in an independent 2011–2013 sample. Final model predictors included young age, low education, early career stage, prior crime involvement, and outpatient treatment for diverse emotional and substance use problems. Area under the receiver operating characteristic curve was 0.79 (for men and women) in the 2004–2009 training sample and 0.74–0.82 (men-women) in the 2011–2013 test sample. 30.5–28.9% (men-women) of all administratively-recorded crimes in 2004–2009 were committed by the 5% of soldiers having highest predicted risk, with similar proportions (28.5–29.0%) when the 2004–2009 coefficients were applied to the 2011–2013 test sample. These results suggest that it may be possible to target soldiers at high-risk of violence perpetration for preventive interventions, although final decisions about such interventions would require weighing predicted effectiveness against intervention costs and competing risks.

Keywords

crime perpetration; military violence; prediction model; risk model; violence prediction

INTRODUCTION

Concerns about non-combat-related violence among U.S. military personnel (Department of the U.S. Army, 2012; Institute of Medicine, 2010) have led to *universal* prevention programs being implemented to train soldiers in violence reduction strategies (Department of Defense Instruction, 2014; Fort Lee, 2014). More intensive prevention programs exist (Naeem et al., 2009; Shea et al., 2013), but would be cost-effective only if *targeted* at soldiers with high risk of violence (Foster & Jones, 2006; Golubnitschaja & Costigliola, 2012). Tools have been developed to assess individual-level violence risk in forensic and inpatient settings (Whittington et al., 2013), but are labor-intensive (e.g., requiring in-depth one-on-one clinical evaluations), making them unrealistic to use in a large workforce such as the military.

A more practical approach for targeting would be to use routinely collected Army/DoD administrative data to develop actuarial prediction models of violence risk (Berk, 2008; Clarke et al., 2009). A model of this sort was recently developed to identify U.S. Army soldiers at high risk of *major physical* violent crime perpetration (e.g., homicide-manslaughter, aggravated assault) (Rosellini et al., 2016). The 5% of soldiers with highest predicted risk in that model accounted for roughly one-third of all major physical violent

crimes among soldiers. However, 85% of physical violent crimes in the Army are *minor* (e.g., simple assault, verbal aggression-harassment) and arguably lead to more distress-impairment in the population than less common major violence. A model to predict minor violence is likely to be different from one for major violence crimes because the predictors of the two are known to differ (Elbogen et al., 2014a; Elbogen et al., 2013; Elbogen et al., 2012; Gallaway et al., 2012; Sullivan & Elbogen, 2014). We consequently attempted to develop parallel models for Army men and women in to predict minor violent crimes.

MATERIALS AND METHODS

Sample

The training sample—The Historical Administrative Data System (HADS) of the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS) was used to develop the models (Ursano et al., 2014). The HADS combined de-identified data from 38 Army/DoD administrative sources (Supplemental Table 1) using a common ID code for the 975,057 Regular U.S. Army soldiers serving in 2004–2009 (Kessler et al., 2013). As detailed below, the HADS was analyzed using discrete-time survival analysis with person-month the unit of analysis (Willett & Singer, 1993). Each month in each soldier's career over the study period was treated as a separate observational record. Models were built separately for men and women based on evidence that violence risk factors differ by sex (Whittington et al., 2013). We focused on predicting first offenses because the vast majority of all soldiers who perpetrated minor violent crimes were first offenses (82% among men; 88% among women). We excluded familial violence because the vast majority of administratively-reported Army violence is non-familial, and predictors of familial violence are quite different from those of non-familial violence (Elbogen et al., 2010a; Marshall et al., 2005; Sullivan & Elbogen, 2014). Given the rarity of the outcome, we used the logic of case-control analysis to select a probability sample of control person-months weighted by the inverse of their probability of selection (Schlesselman, 1982).

The test sample—Model performance was tested by applying the coefficients estimated in the training sample to an independent test sample of 48,718 soldiers who participated in Army STARRS surveys in 2011–2012. We had administrative data for these survey respondents through December 2013. The STARRS survey samples, which are described in detail elsewhere (Kessler et al., 2013), consisted of probability samples of soldiers at all phases of service (basic training, non-deployed, deployed; roughly 1.3 million person-months).

Measures

Minor violence—Five administrative databases were used to obtain information about date, type, and judicial outcome of all reported crimes occurring over the study period (Supplemental Table 2). Crime types were coded according to the Bureau of Justice Statistics National Corrections Reporting Program (NCRP) classification system (U.S. Department of Justice, 2009). A total of 28,694 soldiers in the training sample and 747 in the test sample were classified as committing a non-familial minor violent crime (e.g., simple assault, blackmail-extortion-intimidation, harassment). Consistent with previous research on

administratively-recorded crime (Army Suicide Prevention Task Force & Chiarelli, 2010; Department of the U.S. Army, 2012; Skeem et al., 2015; Steadman et al., 2015), the outcome was any *founded* crime; that is, one for which the Army found sufficient evidence to warrant investigation of the soldier. This decision was based on the fact that founded crime records reflect actual violent behaviors much more closely than do conviction records, as the latter are strongly influenced by the vagaries of bureaucratic processing by the criminal justice system (e.g., charges being dropped because the arresting officer failed to read the suspect his/her rights before questioning or in conjunction with an agreement for a guilty plea to a non-violent offense).

Independent variables—A number of studies have examined predictors of minor violence perpetration in samples of active duty military personnel (Gallaway et al., 2012; Gallaway et al., 2013; MacManus et al., 2012a; MacManus et al., 2012b; MacManus et al., 2013) or veterans (Elbogen et al., 2014a; Elbogen et al., 2010a; Elbogen et al., 2013; Elbogen et al., 2012; Elbogen et al., 2014b; Hellmuth et al., 2012; Jakupcak et al., 2007; Sullivan & Elbogen, 2014). Four broad classes of significant predictors have been identified (Elbogen et al., 2010a): socio-demographic/dispositional (e.g., sex, race-ethnicity, personality); historical (e.g., childhood experiences, military career experiences, prior violence); clinical (e.g., mental and physical disorders, including PTSD and TBI); and contextual/environmental (e.g., access to weapons). As our analysis was carried out using administrative data collected for other purposes, we were not able to operationalize all predictors in these previous studies. However, we were able to identify 446 independent variables operationalizing previously-documented predictors: 21 socio-demographics, 104 historical (38 defining military career experiences and 66 representing prior crime perpetration-victimization), 282 clinical (measures of specific treated mental disorders and broader classes of mental and physical disorders; classes of filled prescriptions), and 39 contextual-environmental (unit characteristics; registered weapons) variables. These variables were defined based on administrative records as of the month *prior* to the target person-month to make sure these variables were not consequences of being arrested for the violent crime. A complete description of these variables is available online (Supplemental Tables 3–6).

Analysis methods

Data analysis was carried out remotely by Harvard Medical School analysts on a secure server at the University of Michigan Army STARRS Data Coordination Center. De-identified analysis was approved by the Human Subjects Committees of the Uniformed Services University of the Health Sciences for the Henry M. Jackson Foundation (the primary Army STARRS grantee), the University of Michigan, and Harvard Medical School. The governing Institutional Review Boards did not require informed consent because HADS data were de-identified.

Analysis began by using cross-tabulations to calculate outcome incidence (expressed as number of founded crimes per 1,000 person-years). Model-building was then based on discrete-time person-month survival analysis rather than incidence analysis (Willett & Singer, 1993). This is an important distinction because examination of risk factors based on

incidence analysis can yield inaccurate results (Kraemer, 2009). Our models examined predictors of first occurrence of a founded minor violent crime in each month of the career of each soldier in the Army between January 2004 and December 2009. The models allowed for time-varying independent variables, as the vast majority of variables had values that changed over time (e.g., rank, time in service, history of prior health care visits, etc.).

The major challenge in developing these models was that use of such a large number of independent variables introduces the possibility of over-fitting. Machine learning methods were used to minimize this problem by searching for stable data patterns. A six-step model building approach was used (see Supplemental Table 7 for additional details):

- i. Bivariate associations of temporally prior independent variables with the outcome were examined, controlling for historical time (season, year), using SAS Version 9.3 *proc logistic* (SAS Institute Inc., 2010). Functional forms of significant non-dichotomous predictors were transformed to capture substantively plausible nonlinearities.
- ii. We then estimated multivariate models that included all significant bivariate predictors but, as expected, model coefficients were highly unstable due to strong inter-correlations among predictors.
- iii. Ten-fold cross-validated forward stepwise regression (Anderssen et al., 2006; Kohavi, 1995) was used to identify the optimal number of independent variables to maximize sensitivity (i.e., the proportion of all observed crimes detected) among the 5% of soldiers with highest predicted risk.
- iv. A search for interactions among all significant bivariate predictors was carried out using the R-package *RandomForests* (RF) (Liaw & Wiener, 2002). The incremental improvement in fit achieved by using RF was determined by adding a composite variable representing the RF predicted probability to the optimal regression model estimated in the previous step and evaluating incremental increase in sensitivity among soldiers in the top 5% of cross-validated predicted risk.
- v. In order to compensate for selection of suboptimal predictors in stepwise models, we estimated elastic net penalized regression models (Zou & Hastie, 2005) using the R-package *glmnet* (Friedman et al., 2010) with the number of independent variables fixed to approximate the optimal number in the stepwise models. Given the active debate about identifying high-risk individuals using information about race-ethnicity (Berk, 2009), this step was carried out both with and without race-ethnicity among the independent variables.
- vi. Conventional (unpenalized) discrete-time survival models were estimated using the best set of independent variables selected in the elastic net models in order to compare unpenalized with penalized model coefficients. Coefficients from both models were then used to calculate predicted probabilities of the outcome for each person-month. Person-

months were ranked by predicted probability and grouped into 20 categories of equal size (ventiles; with the highest 5% of predicted risk being the “top-ventile”) and the proportion of observed crimes in each predicted-risk ventile (i.e., sensitivity) was calculated. The coefficients were then applied to the test sample to calculate sensitivity and positive predictive value (i.e., number of crimes per 1,000 person-years) among the 5% of soldiers with the highest predicted risk.

RESULTS

Incidence by sex, time-in-service, and deployment status

Incidence was significantly higher among men than women (10.0/1,000 person-years versus 6.4/1,000 person-years; $\chi^2_1=630.5$, $p<.001$) and inversely related to time-in-service ($\chi^2_7=776.9-6,061.0$, $p<.001$). (Table 1) Incidence was lower among currently-deployed (3.4–2.4/1,000 person-years) than never-deployed (12.5–7.5/1,000 person-years) or previously-deployed (11.6–6.5/1,000 person-years; $\chi^2_1=175.4-4,167.8$, $p<.001$) soldiers and generally declined with time-in-service (Supplemental Table 8).

Building the models

The majority of independent variables had significant (.05 level, two-sided tests) bivariate associations with the outcome among men (88.1%) and women (77.3%) (Supplemental Table 9–22). Sensitivity among soldiers in the top 5% of cross-validated predicted risk was optimized in stepwise models using roughly two dozen predictors. As sensitivity improved by less than 1% when the RF variables were added to the models (Supplemental Table 23), RF was excluded from the elastic net models. The latter models selected 27 predictors as optimal for men and 24 for women. The 5% of men with highest predicted risk in the optimal model accounted for 30.5% of all crimes. The comparable percentage was 28.9% among women. Area under the receiver operating characteristic curve (AUC) was 0.79 for both men and women. Incidence in the top-ventile of predicted risk (i.e., positive predictive value) was 60.8/1,000 person-years among men and 36.9/1,000 among women. Model performance was slightly lower when excluding race-ethnicity and allowing the elastic net models to replace race-ethnicity with other independent variables (Supplemental Table 23).

Final model predictors

Four socio-demographic variables were in the final elastic net models (both sexes): younger age, Non-Hispanic Black race-ethnicity, less than high school education, and high school graduate with no college education.(Table 2) Six Army career variables were in the models (both sexes): non-deployed status, Area-based Component Commands (i.e., Commands responsible for Army operations in specific regions of the world), low Armed Forces Qualifications Tests score, past 12-month demotion, E1–E4 rank, and E5–E6 rank. Less than 10 years of service and having a recent positive drug test were associated with increased risk only among men.

Six indicators of past 12–24 month crime perpetration and victimization, all associated with increased risk were in the final models: number of different types of perpetration in the past

12 months, any major violence perpetration in the past 12 months, any crime perpetration in the past 12 and past 24 months, and any major violence victimization in the past 12 and 24 months. Three additional crime variables were associated with elevated risk only among men (number of different types of perpetration in the past 24 months; perpetration of major violence in the past 24 months, number of different victimization types in the past 12 months) and another three only among women (any crime victimization in the past 12 and 24 months, number of different victimization types in the past 24 months).

Finally, whereas three clinical variables were associated with increased risk in both sexes (number of outpatient visits for any mental disorder in the past three months, any outpatient visits for marital problems in the past 12 months, any outpatient visits for stressors/adversities in the past 12 months), another three were significant only among men (number of outpatient visits for any mental disorder in the past 12 months, any outpatient visits for an alcohol disorder in the past 12 months, any outpatient visits for other substance disorder in the past 12 months) and two others only among women (inpatient treatment for any mental disorder in the past 3 and 12 months). No contextual-environmental variables were selected by elastic net.

Model accuracy was virtually identical when using the elastic net coefficients or conventional (unpenalized) discrete-time survival coefficients (Figure 1 and Supplemental Table 23), but six of the variables selected for men and four for women had variance inflation factors > 5.0 (indicative of multicollinearity) (Belsley, 1991; Stine, 1995) in the unpenalized models. (Table 2) We consequently used the predicted probabilities generated from the elastic net models in subsequent investigations of model performance.

Comparability with the models for major violent crime

Major violent crime perpetration was a significant predictor in the final models (both sexes). In addition, roughly three-fourths of the predictors in the optimal models for minor violent crime were similar to predictors in our previously-published models for major violent crime (Rosellini et al., 2016). Based on these results, a question could be raised whether the previously-developed models for major violent crime might be equally useful in predicting minor violent crime. We evaluated this question by using the coefficients from our previously-developed major physical violent crime models to predict minor violent crime. Sensitivity in the top-ventile of risk among men decreased from 30.5% to 25.8% and among women from 28.9% to 21.9%.

Stability over longer time periods and different points in the military career

The models were designed to predict crime perpetrated in the *next month*. It is not clear how well prediction accuracy holds up over longer time periods. We addressed this question by estimating sensitivity in the top-ventile of predicted risk for all possible 1-month, 6-month, and 12-month follow-up periods from January 2004 through January 2009 (12-months of follow-up data were unavailable after January 2009) and also by dividing the period between January 2004 and January 2009 in half and thirds. (Table 3) Results were quite consistent over years, with sensitivity in the top-ventile of predicted risk highest over 1-month time

periods averaging 28.5% for men and 25.3% for women and remaining elevated over 6-month (22.7–20.9% men-women) and 12-month (18.4–17.8% men-women).

Although several indicators of early career stage were significant predictors, the failure of RF to improve model performance suggests that there were no substantial interactions between career stage and other independent variables. Nevertheless, sensitivity in the top-ventile of predicted risk varied inversely with time-in-service among both men and women ($\chi^2_{7-6} = 1014.1-89.8$, $p < .001$) (Table 4). However, when cut-points were recalibrated to focus on the 5% of soldiers at highest predicted risk *within* each time-in-service subsample, the association between time-in-service and sensitivity in the top-ventile of predicted risk became non-significant (25.2–28.1% among men, $\chi^2_7 = 5.4$, $p = .61$; 20.4–25.6% among women, $\chi^2_7 = 3.5$, $p = .84$). In contrast, the association of time-in-service with incidence in the top-ventile of predicted risk *increased* when ventiles were defined within time-in-service subsamples. When the top-ventile of predicted risk was defined in the total male sample, for example, incidence was highest in the first year of service (71.2/1,000 person-years) and lowest in the second decade of service (39.7/1,000 person-years; $\chi^2_7 = 109.5$, $p < .001$). When using the within time-in-service top-ventile cut-points, incidence was even higher in the second year-of-service (82.2/1,000 person-years) and even lower in the third decade of service (11.9/1,000 person-years).

Test sample performance

The penalized coefficients estimated in the 2004–2009 models were applied to the test sample through the end of 2013. Sensitivity in the top-ventile of predicted risk was very similar in this test sample as in the training sample: 28.5% for men and 29.0% for women. AUC was 0.74 for men and 0.82 for women.

DISCUSSION

Although numerous studies have examined risk factors for soldier-veteran violence (Elbogen et al., 2014a; Elbogen et al., 2013; Elbogen et al., 2012; Elbogen et al., 2014b; Elbogen et al., 2010b; Gallaway et al., 2012; Gallaway et al., 2013; Hellmuth et al., 2012; Jakupcak et al., 2007; MacManus et al., 2012a; MacManus et al., 2012b; MacManus et al., 2013; Sullivan & Elbogen, 2014), no attempts were made to develop individual-level risk scores prior to our recent work predicting major violence (Rosellini et al., 2016). The goal of the current study was to develop comparable models for minor violent crime. We found that such models could be developed and that these models had AUCs equal to or higher than those of widely-used violence risk tools developed for forensic and inpatient settings (Whittington et al., 2013).

Despite the fact that penalized regression methods are designed to maximize overall model performance at the expense of individual coefficient accuracy, several observations about the predictors in our models are noteworthy. First, we found that young age and indicators of disadvantaged socio-demographic and career status were the strongest predictors of violence. This pattern is consistent with many previous studies (Gallaway et al., 2012; MacManus et al., 2012a; MacManus et al., 2013), although we were unable to evaluate any of the numerous hypotheses advanced to account for these associations (Fear et al., 2009;

Hariri et al., 2000; Harman et al., 2001; Hourani et al., 2006). One noteworthy exception, though, was that being unmarried, which is typically part of the constellation of variables related to young age and low social status predicting violent crime (Blokland & Nieuwbeerta, 2005; Sampson et al., 2006), was absent from our final models. This is part of a larger pattern of the protective effects of marriage being weaker in the U.S. Army than the general population (Gilman et al., 2014). It is also noteworthy in this regard that treatment for marital problems was in our final models for both sexes even though previous research has suggested that marital problems are associated only with intimate partner violence, not non-familial violence (Elbogen et al., 2010a).

Second, not only prior crime perpetration but also several measures of prior crime victimization were selected in the models for both sexes. This is consistent with previous research showing that recent victimization is associated with subsequent crime perpetration (Sadeh et al., 2014), a pattern typically interpreted as due to subcultural factors related to the use of violence as a means of dispute resolution and to reciprocal processes of interpersonal provocation and retaliation (Silver et al., 2011).

Third, diagnosed-treated mental disorders accounted for roughly one-fifth of all the predictors in our final models. This is broadly consistent with previous research finding mental disorders associated with elevated violence rates (Swanson et al., 2015a; Swanson et al., 2015b), although it is noteworthy that neither PTSD nor TBI, two of the most commonly studied risk factors for violence perpetration among soldiers (Grafman et al., 1996; Gallaway et al., 2012; Hellmuth et al., 2012; Elbogen et al., 2014b; Sullivan & Elbogen, 2014), was among the final model predictors.

Even though positive predictive value was roughly six times as high in the highest risk ventile of our models as the total-population, minor violent crimes are still relatively rare even among highest-risk soldiers, with only 6% of men and 3% of women in the highest-risk ventile perpetrating minor violence during the subsequent month (although the stability of predictions over time means that an additional 5% of men in the highest-risk ventile committed a minor violent crime in the following month, an additional 4% in the third month, etc.). This observation raises the question whether minor violent crime prevalence is sufficiently high to warrant using prediction models to target soldiers for high-risk preventive interventions. The only principled way to answer this question is to carry out a systematic cost-effectiveness analysis taking into consideration both the benefits of available interventions in reducing violent crime to victims, perpetrators, and the Army and intervention costs (including competing risk, such as risks to soldiers labeled as being “high risk,” the majority of whom would, in fact, not commit a crime in the absence of an intervention). Such an analysis is beyond the scope of this report.

The analysis was limited by the administrative data not including all significant predictors of violence found in previous research (e.g., information on witnessing family violence as a child and other pre-accession violent behaviors; Elbogen et al., 2013; Elbogen et al., 2014b). Furthermore, we did not investigate the causal influences of modifiable risk factors to help guide the design of preventive interventions. However, we achieved our more modest goal of demonstrating that useful prediction models can be developed from existing administrative

data. Given the availability of administrative variables for all soldiers in the Army, it would be relatively easy to generate predicted risk scores for each soldier and to update these scores over time to monitor changes for purposes of guiding targeted preventive interventions. Whether cost-effectiveness considerations judge this to be something that would have a positive net value, though, is a matter that requires future analysis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Role of the funding source

The data analyzed in this report were collected as part of the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). Army STARRS was sponsored by the Department of the Army and funded under cooperative agreement number U01MH087981 with the U.S. Department of Health and Human Services, National Institutes of Health, National Institute of Mental Health (NIH/NIMH). This research was conducted by Harvard Medical School and is funded by the Department of Defense, Office of the Assistant Secretary for Defense for Health Affairs, Defense Health Program (OASD/HA), awarded and administered by the U.S. Army Medical Research & Materiel Command (USAMRMC), at Fort Detrick, MD, under Contract Number: (Award # W81XWH-12-2-0113). Although a draft of this manuscript was submitted to the Army for review and comment before submission, this was with the understanding that comments would be advisory.

References

- Anderssen E, Dyrstad K, Westad F, Martens H. Reducing over-optimism in variable selection by cross-model validation. *Chemometr Intell Lab Syst*. 2006; 84:69–74.
- Chiarelli, PW. Army Suicide Prevention Task Force. Army health promotion, risk reduction, suicide prevention report 2010. U.S. Army; Washington, D.C: 2010.
- Belsley, DA. Conditioning Diagnostics: Collinearity and Weak Data in Regression. John Wiley & Sons; New York: 1991.
- Berk, RA. Statistical Learning from a Regression Perspective. Springer; New York: 2008.
- Berk RA. The role of race in forecasts of violent crime. *Race Soc Probl*. 2009; 1:231–242.
- Blokland AJ, Nieuwbeerta P. The effects of life circumstances on longitudinal trajectories of offending. *Criminology*. 2005; 43:1203–1240.
- Clarke, B.; Fokoue, E.; Zhang, HH. Principles and Theory for Machine Learning and Data Mining. Springer; New York: 2009.
- Department of Defense Instruction. DoD workplace violence prevention and response policy. Department of Defense; Washington D.C: 2014.
- Department of the U.S. Army. Army 2020: generating health & discipline in the force ahead of the strategic reset. US Army; Washington, DC: 2012.
- Elbogen EB, Cueva M, Wagner HR, Sreenivasan S, Brancu M, Beckham JC, Van Male L. Screening for violence risk in military veterans: predictive validity of a brief clinical tool. *Am J Psychiatry*. 2014a; 171:749–757. [PubMed: 24832765]
- Elbogen EB, Fuller S, Johnson SC, Brooks S, Kinneer P, Calhoun PS, Beckham JC. Improving risk assessment of violence among military veterans: an evidence-based approach for clinical decision-making. *Clin Psychol Rev*. 2010a; 30:595–607. [PubMed: 20627387]
- Elbogen EB, Johnson SC, Newton VM, Fuller S, Wagner HR, Beckham JC. Self-report and longitudinal predictors of violence in Iraq and Afghanistan war era veterans. *J Nerv Ment Dis*. 2013; 201:872–876. [PubMed: 24080674]
- Elbogen EB, Johnson SC, Wagner HR, Newton VM, Timko C, Vasterling JJ, Beckham JC. Protective factors and risk modification of violence in Iraq and Afghanistan War veterans. *J Clin Psychiatry*. 2012; 73:e767–773. [PubMed: 22795217]

- Elbogen EB, Johnson SC, Wagner HR, Sullivan C, Taft CT, Beckham JC. Violent behaviour and post-traumatic stress disorder in US Iraq and Afghanistan veterans. *Br J Psychiatry*. 2014b; 204:368–375. [PubMed: 24578444]
- Elbogen EB, Wagner HR, Fuller SR, Calhoun PS, Kinneer PM, Beckham JC. Correlates of anger and hostility in Iraq and Afghanistan war veterans. *Am J Psychiatry*. 2010b; 167:1051–1058. [PubMed: 20551162]
- Fear NT, Rubin GJ, Hatch S, Hull L, Jones M, Hotopf M, Wessely S, Rona RJ. Job strain, rank, and mental health in the UK Armed Forces. *Int J Occup Environ Health*. 2009; 15:291–298. [PubMed: 19650584]
- Lee, Fort. [accessed 25.02.16] Prevention of workplace violence program. 2014. <http://www.lee.army.mil/hrd/prevention.of.workplace.violence.program.aspx>
- Foster EM, Jones D. Can a costly intervention be cost-effective?: an analysis of violence prevention. *Arch Gen Psychiatry*. 2006; 63:1284–1291. [PubMed: 17088509]
- Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010; 33:1–22. [PubMed: 20808728]
- Gallaway MS, Fink DS, Millikan AM, Bell MR. Factors associated with physical aggression among US Army soldiers. *Aggress Behav*. 2012; 38:357–367. [PubMed: 22898873]
- Gallaway MS, Fink DS, Millikan AM, Mitchell MM, Bell MR. The association between combat exposure and negative behavioral and psychiatric conditions. *J Nerv Ment Dis*. 2013; 201:572–578. [PubMed: 23817154]
- Gilman SE, Bromet EJ, Cox KL, Colpe LJ, Fullerton CS, Gruber MJ, Heeringa SG, Lewandowski-Romps L, Millikan-Bell AM, Naifeh JA, Nock MK, Petukhova MV, Sampson NA, Schoenbaum M, Stein MB, Ursano RJ, Wessely S, Zaslavsky AM, Kessler RC. Sociodemographic and career history predictors of suicide mortality in the United States Army 2004–2009. *Psychol Med*. 2014; 44:2579–2592. [PubMed: 25055175]
- Golubnitschaja O, Costigliola V. General report & recommendations in predictive, preventive and personalised medicine 2012: white paper of the European Association for Predictive, Preventive and Personalised Medicine. *Epma J*. 2012; 3:14. [PubMed: 23116135]
- Grafman J, Schwab K, Warden D, Pridgen A, Brown HR, Salazar AM. Frontal lobe injuries, violence, and aggression: a report of the Vietnam Head Injury Study. *Neurology*. 1996; 46:1231–8. [PubMed: 8628458]
- Hariri AR, Bookheimer SY, Mazziotta JC. Modulating emotional responses: effects of a neocortical network on the limbic system. *Neuroreport*. 2000; 11:43–48. [PubMed: 10683827]
- Harman SM, Metter EJ, Tobin JD, Pearson J, Blackman MR. Longitudinal effects of aging on serum total and free testosterone levels in healthy men. Baltimore Longitudinal Study of Aging. *J Clin Endocrinol Metab*. 2001; 86:724–731. [PubMed: 11158037]
- Hellmuth JC, Stappenbeck CA, Hoerster KD, Jakupcak M. Modeling PTSD symptom clusters, alcohol misuse, anger, and depression as they relate to aggression and suicidality in returning U.S. veterans. *J Trauma Stress*. 2012; 25:527–534. [PubMed: 23073972]
- Hourani LL, Williams TV, Kress AM. Stress, mental health, and job performance among active duty military personnel: findings from the 2002 Department of Defense Health-Related Behaviors Survey. *Mil Med*. 2006; 171:849–856. [PubMed: 17036605]
- Institute of Medicine. *Returning Home from Iraq and Afghanistan: Preliminary Assessment of Readjustment Needs of Veterans, Service Members, and their Families*. The National Academies Press; Washington, DC: 2010.
- Jakupcak M, Conybeare D, Phelps L, Hunt S, Holmes HA, Felker B, Klevens M, McFall ME. Anger, hostility, and aggression among Iraq and Afghanistan War veterans reporting PTSD and subthreshold PTSD. *J Trauma Stress*. 2007; 20:945–954. [PubMed: 18157891]
- Kessler RC, Colpe LJ, Fullerton CS, Gebler N, Naifeh JA, Nock MK, Sampson NA, Schoenbaum M, Zaslavsky AM, Stein MB, Ursano RJ, Heeringa SG. Design of the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *Int J Methods Psychiatr Res*. 2013; 22:267–275. [PubMed: 24318217]

- Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. Proceedings of the 14th International Joint Conference on Artificial Intelligence; Montreal, Quebec, Canada. Morgan Kaufmann Publishers Inc; 1995. p. 1137-1143.
- Kraemer HC. Events per person-time (incidence rate): a misleading statistic? *Stat Med*. 2009; 28:1028–1039. [PubMed: 19152356]
- Liaw A, Wiener M. Classification and regression by randomForest. *R News*. 2002; 2:18–22.
- MacManus D, Dean K, Al Bakir M, Iversen AC, Hull L, Fahy T, Wessely S, Fear NT. Violent behaviour in U.K. military personnel returning home after deployment. *Psychol Med*. 2012a; 42:1663–1673. [PubMed: 22115074]
- MacManus D, Dean K, Iversen AC, Hull L, Jones N, Fahy T, Wessely S, Fear NT. Impact of pre-enlistment antisocial behaviour on behavioural outcomes among U.K. military personnel. *Soc Psychiatry Psychiatr Epidemiol*. 2012b; 47:1353–1358. [PubMed: 22038568]
- MacManus D, Dean K, Jones M, Rona RJ, Greenberg N, Hull L, Fahy T, Wessely S, Fear NT. Violent offending by UK military personnel deployed to Iraq and Afghanistan: a data linkage cohort study. *Lancet*. 2013; 381:907–917. [PubMed: 23499041]
- Marshall AD, Panuzio J, Taft CT. Intimate partner violence among military veterans and active duty servicemen. *Clin Psychol Rev*. 2005; 25:862–876. [PubMed: 16006025]
- Naeem F, Clarke I, Kingdon D. A randomized controlled trial to assess an anger management group programme. *Cogn Behav Therapist*. 2009; 2:20–31.
- Rosellini AJ, Monahan J, Street AE, Heeringa SG, Hill ED, Petukhova M, Reis BY, Sampson NA, Bliese P, Schoenbaum M, Stein MB, Ursano RJ, Kessler RC. Predicting non-familial major physical violent crime perpetration in the US Army from administrative data. *Psychol Med*. 2016; 46:303–316. [PubMed: 26436603]
- Sadeh N, Binder RL, McNiel DE. Recent victimization increases risk for violence in justice-involved persons with mental illness. *Law Hum Behav*. 2014; 38:119–125. [PubMed: 23855324]
- Sampson RJ, Laub JH, Wimer C. Does marriage reduce crime? a counterfactual approach to within-individual causal effects. *Criminology*. 2006; 44:465–508.
- SAS Institute Inc. SAS/STAT Software. SAS Institute Inc; Cary, NC: 2010. Version 9.3 for Unix
- Schlesselman, JJ. Case-control Studies: Design, Conduct, Analysis. Oxford University Press; New York: 1982.
- Shea MT, Lambert J, Reddy MK. A randomized pilot study of anger treatment for Iraq and Afghanistan veterans. *Behav Res Ther*. 2013; 51:607–613. [PubMed: 23916629]
- Silver E, Piquero AR, Jennings WG, Piquero NL, Leiber M. Assessing the violent offending and violent victimization overlap among discharged psychiatric patients. *Law Hum Behav*. 2011; 35:49–59. [PubMed: 20145985]
- Skeem JL, Kennealy P, Monahan J, Peterson J, Appelbaum P. Psychosis uncommonly and inconsistently precedes violence among high-risk individuals. *Clin Psychol Sci*. 2015; 4:40–49.
- Steadman HJ, Monahan J, Pinals DA, Vesselinov R, Robbins PC. Gun violence and victimization of strangers by persons with a mental illness: data from the MacArthur Violence Risk Assessment Study. *Psychiatr Serv*. 2015; 66:1238–1241. [PubMed: 26073414]
- Stine RA. Graphical interpretation of variance inflation factors. *Am Stat*. 1995; 49:53–56.
- Sullivan CP, Elbogen EB. PTSD symptoms and family versus stranger violence in Iraq and Afghanistan veterans. *Law Hum Behav*. 2014; 38:1–9. [PubMed: 23646917]
- Swanson JW, McGinty EE, Fazel S, Mays VM. Mental illness and reduction of gun violence and suicide: bringing epidemiologic research to policy. *Ann Epidemiol*. 2015a; 25:366–376. [PubMed: 24861430]
- Swanson JW, Sampson NA, Petukhova MV, Zaslavsky AM, Appelbaum PS, Swartz MS, Kessler RC. Guns, impulsive angry behavior, and mental disorders: results from the National Comorbidity Survey Replication (NCS-R). *Behav Sci Law*. 2015b; 33:199–212. [PubMed: 25850688]
- U.S. Department of Justice. [accessed 24.02.16] National Corrections Reporting Program, 2009 (ICPSR 30799). 2009. <http://www.icpsr.umich.edu/icpsrweb/NACJD/studies/30799?archive=NACJD&permit%5B0%5D=AVAILABLE&q=30799&x=0&y=0>

- Ursano RJ, Colpe LJ, Heeringa SG, Kessler RC, Schoenbaum M, Stein MB. The Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *Psychiatry*. 2014; 77:107–119. [PubMed: 24865195]
- Whittington R, Hockenhull JC, McGuire J, Leitner M, Barr W, Cherry MG, Flentje R, Quinn B, Dundar Y, Dickson R. A systematic review of risk assessment strategies for populations at high risk of engaging in violent behaviour: update 2002–8. *Health Technol Assess*. 2013; 17:i–xiv. 1–128.
- Willett JB, Singer JD. Investigating onset, cessation, relapse, and recovery: why you should, and how you can, use discrete-time survival analysis to examine event occurrence. *J Consult Clin Psychol*. 1993; 61:952–965. [PubMed: 8113496]
- Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Series B Stat Methodol*. 2005; 67:301–320.

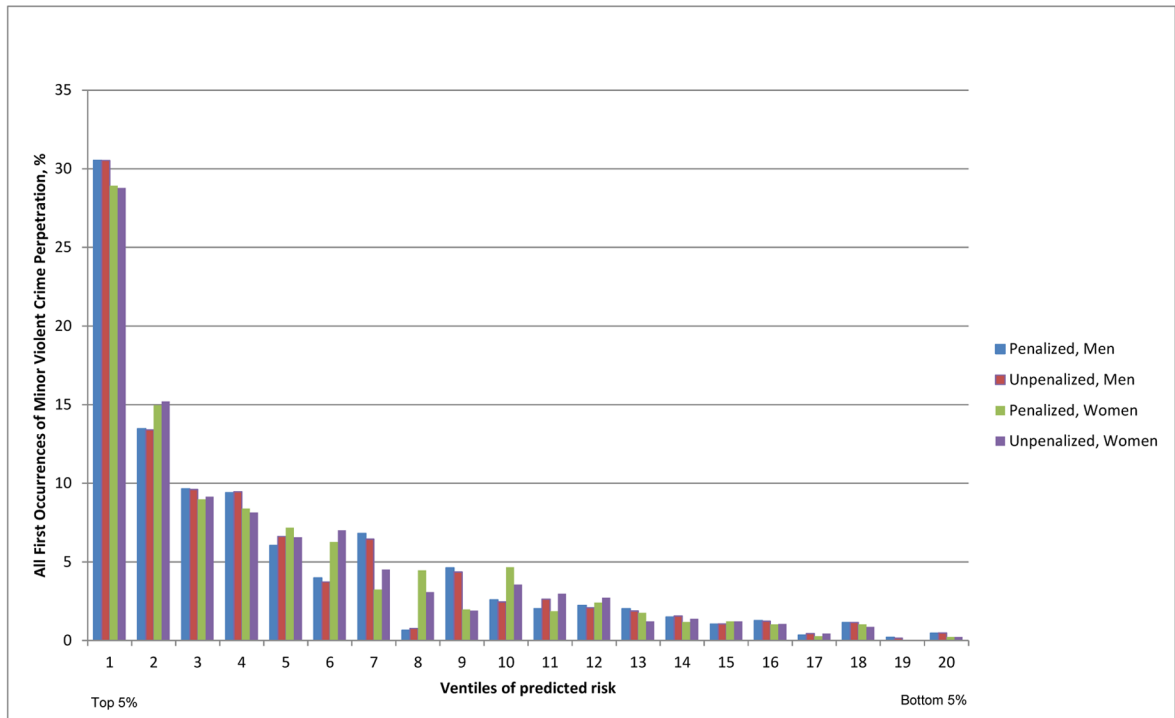


Figure 1. Proportion of observed crimes committed by Ventile of Predicted Risk Based on the Final Discrete-Time Survival Models of Men (27 predictors) and Women (24 predictors)
^aVentiles are 20 groups of person-months of equal frequency dividing the total sample of person-months into equally sized groups defined by level of predicted perpetration risk

Table 1

Incidence/1,000 person-years of first founded non-familial minor violent crime perpetration by time-in-service and sex among Regular Army soldiers in the Army STARRS 2004–2009 Historical Administrative Data Systems (HADS)^a

Years-in-service	Men						Women					
	Incidence/1,000 person-years		Distribution of the crimes		Population distribution ^c		Incidence/1,000 person-years		Distribution of the crimes		Population distribution ^c	
	Est	(se)	(n) ^b	%	(se)	% ^c	Est	(se)	(n) ^b	%	(se)	% ^c
0–1	12.5	0.2	3,998	15.4	0.2	12.3	7.8	0.4	466	17.1	0.7	14.0
1–2	15.7	0.2	4,555	17.5	0.2	11.1	10.6	0.5	555	20.3	0.8	12.3
2–3	14.4	0.2	3,933	15.2	0.2	10.5	9.6	0.5	436	16.0	0.7	10.6
3–4	13.0	0.2	2,958	11.4	0.2	8.8	8.8	0.5	348	12.8	0.6	9.3
4–5	11.6	0.3	1,926	7.4	0.2	6.4	7.2	0.5	221	8.1	0.5	7.2
5–10	9.8	0.1	5,257	20.2	0.2	20.6	5.2	0.2	484	17.7	0.7	21.7
10–20	4.8	0.1	2,997	11.5	0.2	24.2	2.2	0.2	199	7.3	0.5	20.8
20+	2.1	0.1	342	1.3	0.1	6.2	1.1	0.2	19	0.7	0.2	4.2
Total	10.0	0.1	25,966	100.0	--	100.0	6.4	0.1	2,728	100.0	--	100.0
χ^2_7	6,061.0*						776.9*					

Abbreviations: Est, estimate; se, standard error.

* Significant at the .05 level, two-sided test.

^a 28,694 Regular Army soldiers had first founded non-familial minor violent crime perpetrations between January 1, 2004 and December 31, 2009 out of the 975,057 Regular Army soldiers (821,807 men; 153,250 women) in active duty service over that time period.

^b n = number of soldiers with first founded non-familial minor violent crime perpetrations in the time interval represented by the row.

^c Percent of the total population person-months in the time interval represented by the row. Men had a total of 31,249,056 population person-months and women had a total of 5,129,943 population person-months.

Table 2

Coefficients (odds-ratios) from the final penalized and unpenalized survival models for first founded non-familial minor violent crime perpetration^a

	Men				Women							
	Penalized ^b		Unpenalized		Penalized ^b		Unpenalized					
	%	(se)	OR	(95% CI)	VIF ^c	%	(se)	OR	(95% CI)	VIF ^c		
I. Socio-demographics												
Age - 17-22	26.0	0.1	1.2	1.2*	(1.2-1.2)	1.5	28.3	0.3	1.3	1.3*	(1.2-1.5)	1.6
Race/ethnicity - Non-Hispanic Black	18.1	0.1	2.0	2.0*	(2.0-2.1)	1.1	37.0	0.3	2.0	2.0*	(1.8-2.1)	1.1
Education - Less than high school	11.2	0.1	2.5	2.6*	(2.5-2.8)	2.0	6.7	0.1	2.8	2.8*	(2.3-3.4)	1.6
Education - Completed high school but no college	63.9	0.1	1.7	1.7*	(1.6-1.8)	2.5	61.8	0.3	1.7	1.7*	(1.5-2.1)	2.4
II. Army career												
Previously or never deployed (not currently)	76.3	0.1	3.4	3.6*	(3.4-3.7)	1.0	84.2	0.2	2.9	3.0*	(2.5-3.5)	1.0
Area-based Component Command	18.0	0.1	1.7	1.7*	(1.7-1.8)	1.0	16.8	0.2	1.7	1.7*	(1.6-1.9)	1.0
Rank junior enlisted (E1-E4)	44.9	0.1	2.3	2.4*	(2.3-2.6)	4.4	47.8	0.3	3.9	4.4*	(3.5-5.6)	3.4
Rank intermediate enlisted (E5-E6)	27.7	0.1	1.8	1.9*	(1.8-2.0)	2.7	25.8	0.3	2.2	2.6*	(2.0-3.3)	2.3
Years in service 10 or less	69.6	0.1	1.4	1.4*	(1.3-1.5)	1.8	--	--	--	--	--	--
AFQT 49 th percentile or less	28.6	0.1	1.3	1.3*	(1.3-1.4)	1.1	33.4	0.3	1.2	1.3*	(1.2-1.4)	1.2
Demotion in the past 12 months	2.8	0.0	1.5	1.5*	(1.4-1.5)	1.1	2.6	0.1	1.6	1.6*	(1.4-1.9)	1.1
Positive drug test in past 3 months	0.4	0.0	1.3	1.3*	(1.2-1.4)	1.1	--	--	--	--	--	--
III. Prior crime												
a. Perpetration of....												
Any crime in past 12 months (Count) ^d	5.1	0.0	1.1	1.0	(0.9-1.1)	19.4	3.6	0.1	1.1	1.1	(0.9-1.4)	10.8
Any crime in past 24 months (Count) ^d	7.8	0.0	1.1	1.1*	(1.1-1.2)	18.6	--	--	--	--	--	--
Major violence in past 12 months	0.4	0.0	1.3	1.3*	(1.1-1.5)	3.2	0.2	0.0	1.5	1.5*	(1.2-2.0)	1.1
Any crime in past 12 months	5.1	0.0	1.3	1.3*	(1.2-1.5)	17.2	3.6	0.1	1.0	1.0	(0.7-1.4)	12.2
Major violence in past 24 months	0.7	0.0	1.4	1.3*	(1.1-1.5)	3.2	--	--	--	--	--	--
Any crime in past 24 months	7.8	0.0	1.4	1.4*	(1.3-1.6)	16.4	5.6	0.1	1.9	1.9*	(1.6-2.3)	2.8
b. Victim of....												
Any crime in past 12 months (Count) ^e	2.6	0.0	1.3	1.4*	(1.3-1.4)	1.2	--	--	--	--	--	--
Major violence in past 12 months	0.3	0.0	1.0	1.0	(0.7-1.0)	2.8	0.7	0.0	1.1	1.1	(0.7-1.7)	3.0

	Men						Women					
	Penalized ^b			Unpenalized			Penalized ^b			Unpenalized		
	%	(se)	OR	OR	(95% CI)	VIF ^c	%	(se)	OR	OR	(95% CI)	VIF ^c
Major violence in past 24 months	0.4	0.0	1.3	1.5*	(1.2–1.7)	2.7	1.1	0.1	1.4	1.4*	(1.0–2.1)	3.1
Any crime in past 24 months (Count) ^e	--	--	--	--	--	--	7.1	0.1	1.1	1.1	(1.0–1.3)	7.9
Any crime in past 12 months	--	--	--	--	--	--	4.4	0.1	1.1	1.1	(0.9–1.4)	3.0
Any crime in past 24 months	--	--	--	--	--	--	7.1	0.1	1.3	1.3	(1.0–1.6)	8.5
IV. Clinical factors												
a. Outpatient treatment of...												
Any mental disorder in past 3 months (Categorical) ^f	13.4	0.1	1.0	1.0*	(1.0–1.1)	2.2	21.1	0.2	1.1	1.1*	(1.1–1.2)	1.3
Any mental disorder in past 12 months (Categorical) ^f	28.0	0.1	1.1	1.1*	(1.1–1.1)	2.8	--	--	--	--	--	--
Marital problems in past 12 months	2.4	0.0	1.5	1.5*	(1.4–1.6)	1.4	3.2	0.1	1.5	1.5*	(1.3–1.8)	1.4
Stressor/adversity in past 12 months	8.9	0.1	1.4	1.4*	(1.3–1.5)	1.8	13.0	0.2	1.3	1.3*	(1.2–1.5)	1.5
Alcohol disorder in past 12 months	2.3	0.0	1.4	1.4*	(1.3–1.6)	5.8	--	--	--	--	--	--
Any substance disorder in past 12 months	2.7	0.0	1.1	1.0	(0.9–1.1)	6.1	--	--	--	--	--	--
b. Inpatient treatment of...												
Any mental disorder in past 3 months	--	--	--	--	--	--	0.3	0.0	1.3	1.3	(0.9–1.9)	1.6
Any mental disorder in past 12 months	--	--	--	--	--	--	0.9	0.1	1.7	1.8*	(1.3–2.3)	1.6

Abbreviations: OR, odds ratio; 95% CI, 95% confidence interval; se, standard error; VIF, variance inflation factor; MPP, mixing parameter penalty; MOS, military occupational specialty; TRADOC, training and doctrine command; FORSCOM, forces command; ODD, oppositional defiant disorder; WO, warrant officer; CO, commissioned officer; NCO, non-commissioned officer.

^aThe analysis sample included all person-months with the outcome plus a probability sample of all other person-months in the population stratified by sex and marital status (total case-control sample of 1,001,092 person-months; 910,688 for men; 90,404 for women). All records in the control sample were weighted by the inverse of probability of selection. Variables were coded dichotomously (yes/no) unless otherwise indicated.

^bThe optimal penalized (elastic net) models had a mixing parameter of $\alpha=0.7$ for men and $\alpha=0.9$ for women. These are weighted more in the direction of the lasso penalty ($\alpha=1.0$), which favors deletion of all but one predictor in each highly correlated predictor set, than the ridge penalty ($\alpha=0.0$), which favors the inclusion of all predictors even when they are highly correlated and adjustment for multicollinearity by coefficient shrinkage.

^cVariance Inflation Factor (VIF) for the coefficient associated with independent variable Xi in the above equation equals $1/(1-R^2_i)$, where R^2_i is the coefficient of determination of a regression equation in which Xi is the dependent variable and all the other independent variables in the model are included as predictors of Xi. VIF 5.0 is typically considered an indicator of meaningful multicollinearity (Belsey, 1991; Stine, 1995).

^dThis was a continuous variable coded 0–9 to represent the number of different types of founded perpetrations. The percentages reported reflect that 5.1% of men and 3.6% of women perpetrated one or more crimes in the past 12 months, and 7.8% of men perpetrated one or more crimes in the past 24 months.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

This was a continuous variable coded 0–5 to represent the number of different types of victimizations. The percentages reported reflect that 2.6% of men had been victimized one or more times in the past 12 months and 7.1% of women had been victimized one or more times in the past 24 months.

This was a categorical variable coded 0–4 (0=0 visits; 1=1–2 visits; 2=3–5 visits; 3=6–10 visits; 4=11+ visits). The percentages reported reflects that 13.4% of men and 21.1% of women had one or more days with outpatient visits for any mental disorder in the past 3 months, and 28.0% of men had one or more such visit in the past 12 months.

Table 3

Proportion of observed crimes committed by those in the top-ventile of predicted risk over 1-month, 6-month, and 12-month periods and across time-intervals^a

	Men			Women		
	1-month	6-month	12-month	1-month	6-month	12-month
	Est (se)	Est (se)	Est (se)	Est (se)	Est (se)	Est (se)
1/04–1/09	28.5 0.3	22.7 0.1	18.4 0.1	25.3 0.9	20.9 0.3	17.8 0.2
1/04–8/05	28.1 0.5	22.6 0.2	18.4 0.1	24.1 1.6	20.2 0.6	17.2 0.4
9/05–4/07	28.4 0.5	22.2 0.2	18.3 0.1	26.5 1.7	21.1 0.6	17.9 0.4
5/07–1/09	28.8 0.5	23.1 0.2	18.6 0.1	25.4 1.5	21.2 0.6	18.1 0.4
1/04–6/06	28.3 0.4	22.6 0.2	18.5 0.1	24.3 1.3	20.2 0.5	17.2 0.3
7/06–1/09	28.6 0.4	22.7 0.2	18.3 0.1	26.2 1.3	21.5 0.5	18.2 0.3

Abbreviations: Est, estimate; se, standard error.

^aEstimates are based on the predicted probabilities from the optimal penalized (elastic net) models. The optimal penalized models had a mixing parameter of $\alpha=0.7$ for men and $\alpha=0.9$ for women. These are weighted more in the direction of the lasso penalty ($\alpha=1.0$), which favors deletion of all but one predictor in each highly correlated predictor set, than the ridge penalty ($\alpha=0.0$), which favors the inclusion of all predictors even when they are highly correlated and adjustment for multicollinearity by coefficient shrinkage. February–December 2009 were excluded because we did not have 12-months of follow-up data after these months

Table 4

Incidence/1,000 person-years and the proportion of observed crimes committed by those in the top-ventile of predicted risk within time-in-service subsamples by sex and time-in-service among Regular Army soldiers in the Army STARRS 2004–2009 Historical Administrative Data Systems (HADS)^a

Years-in- service	Overall top-ventile of predicted risk						Within-time-in-service top-ventile of predicted risk ^b					
	Incidence/1,000 person-years (PPV)			Proportion of person-months in top-ventile			Incidence/1,000 person-years (PPV)			Within-row % of observed crimes		
	Est	(se)		Est	(se)	%	Est	(se)		Est	(se)	
I. Men												
0–1	71.2	3.1	20.7	0.6	0.1	3.6	62.9	2.4	25.2	0.7		
1–2	65.6	2.0	37.4	0.7	0.2	9.0	82.2	3.1	26.1	0.7		
2–3	58.3	1.9	35.7	0.8	0.2	8.9	75.6	3.0	26.3	0.7		
3–4	54.4	2.0	35.1	0.9	0.2	8.4	66.3	3.0	25.6	0.8		
4–5	54.4	2.6	33.0	1.1	0.2	7.0	59.4	3.3	25.6	1.0		
5–10	51.3	1.7	25.5	0.6	0.1	4.9	50.9	1.6	26.1	0.6		
10–20	39.7	2.2	14.3	0.6	0.0	1.7	26.0	1.0	27.3	0.8		
20+	40.3	8.4	9.1	1.6	0.1	0.5	11.9	1.3	28.1	2.4		
Total	60.8	0.8	30.5	0.3	0.0	5.0	51.9	0.8	26.1	0.3		
χ^2_7		109.5*		1,014.1*		7,341.4*		973.1*		5.4		
II. Women												
0–1	38.9	5.1	19.1	1.8	0.3	3.8	33.1	4.0	21.2	1.9		
1–2	36.9	3.4	32.1	2.0	0.5	9.2	45.3	5.3	21.4	1.7		
2–3	31.9	3.1	34.2	2.3	0.5	10.3	39.2	5.2	20.4	1.9		
3–4	29.4	3.5	28.4	2.4	0.5	8.5	36.2	5.2	20.7	2.2		
4–5	34.3	5.3	27.6	3.0	0.5	5.8	33.0	5.6	23.1	2.8		
5–10	26.5	3.1	21.1	1.9	0.2	4.2	24.5	2.7	23.6	1.9		
10–20	28.4	7.7	9.5	2.1	0.1	0.8	11.4	1.7	25.6	3.1		
20+	--	--	--	--	--	--	4.2	2.1	21.1	9.4		
Total	36.9	1.5	28.9	0.8	0.1	5.0	27.9	1.3	22.0	0.8		
χ^2_7		8.0		89.8*		844.4*		103.2*		3.5		

Abbreviations: PPV, positive predictive value, Est, estimate; se, standard error.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

* Significant at the .05 level, two-sided test.

^aEstimates are based on the predicted probabilities from the optimal penalized (elastic net) models. The optimal penalized models had a mixing parameter of $\alpha=0.7$ for men and $\alpha=0.9$ for women. These are weighted more in the direction of the lasso penalty ($\alpha=1.0$), which favors deletion of all but one predictor in each highly correlated predictor set, than the ridge penalty ($\alpha=0.0$), which favors the inclusion of all predictors even when they are highly correlated and adjustment for multicollinearity by coefficient shrinkage.

^bVentiles were re-classified independently within each time in service group so the top-ventile of predicted risk includes 5% of the person-months within each time in service category.