Award Number: W81XWH-12-2-0113

TITLE: Behavioral-Based Predictors of Workplace Violence in the Army STARRS

PRINCIPAL INVESTIGATOR: Ronald Kessler, Ph.D.

CONTRACTING ORGANIZATION: Harvard Medical School, Boston, MA 02115

REPORT DATE: October 2014

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE October 2014	**2. REPORT TYPE** Annual	**3. DATES COVERED** 30 Sep 2013 – 29 Sept 2014
4. TITLE AND SUBTITLE Behavioral-Based Predictors of Workplace Violence in the Army STARRS		**5a. CONTRACT NUMBER** W81XWH-12-2-0113
		5b. GRANT NUMBER W81XWH-12-2-0113
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Dr. Ronald Kessler E-Mail: Kessler@hcp.med.harvard.edu		**5d. PROJECT NUMBER**
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) HARVARD MEDICAL SCHOOL 25 SHATTUCK ST BOSTON MA 02115-6027		**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012		**10. SPONSOR/MONITOR'S ACRONYM(S)**
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT Over the past year we used data from the Army STARRS Historical Administrative Data System (HADS), New Soldiers Study (NSS), All-Army Study (AAS), and Pre-Post Deployment Study (PPDS) to (i) estimate standardized rates of several types of workplace violence perpetration and victimization, (ii) examine bivariate associations of administrative and survey predictors with workplace violence outcomes, and (iii) develop risk prediction equations using data mining. Standardized rates of perpetration and victimization across the HADS, NSS, AAS, and PPDS suggested that male soldiers have higher rates perpetrating workplace violence, while female soldiers have higher rates of victimization. Most bivariate associations with our outcomes were consistent with expectations. The most noteworthy findings from the past year pertain to our development of risk prediction equations. We used data mining methods to maximize the prediction of 15 perpetration and victimization outcomes in the HADS and NSS. Overall, these models performed quite well. Prediction accuracy was much better than would be expected by chance (AUCs = 0.7-0.9), and soldiers who were predicted to be high risk accounted for a large proportion of all workplace violence perpetrations and victimizations (18.4-72.9% of all first occurrences of perpetration/victimization occurred among soldiers in the top 5% of predicted risk).

15. SUBJECT TERMS
U.S. Army; new soldiers; workplace violence; physical violence; sexual violence; perpetration; victimization; prevalence; socio-demographics; risk factors; protective factors; administrative data; survey data; stepwise regression; elastic net; cross-validation; machine learning; data mining; risk algorithm; risk score

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** USAMRMC
a. REPORT U	**b. ABSTRACT** U	**c. THIS PAGE** U	UU	61	**19b. TELEPHONE NUMBER** *(include area code)*

Table of Contents

Page

INTRODUCTION

The primary objective of this project is to carry out secondary analyses of data on risk-protective factors for workplace violence perpetration-victimization in the Army Study to Assess Risk and Resilience in Servicemembers (A-STARRS), the largest epidemiological study of mental health risk and resilience ever conducted among US Army personnel. Although the primary focus of A-STARRS is suicide, much information also exists on other topics, one of them being violence. Using A-STARRS datasets, we will estimate the prevalence and predictors of workplace violence perpetration and victimization in order to develop risk prediction tools that can be used by the Army to target soldiers at high risk of workplace violence perpetration or victimization. A secondary objective is to use the results of these analyses to expand knowledge about modifiable risk and protective factors for workplace violence perpetration and victimization in the Army. The latter information could be of value to the Army to develop or modify preventive interventions for workplace violence perpetration and victimization.

KEYWORDS

U.S. Army; new soldiers; workplace violence; physical violence; sexual violence; perpetration; victimization; prevalence; socio-demographics; risk factors; protective factors; administrative data; survey data; stepwise logistic regression; elastic net penalized regression; cross-validation; machine learning; data mining; risk algorithm; risk score

OVERALL PROJECT SUMMARY

Due to the Year 1 delay in obtaining the Army Omnibus DUA, work and spending were also delayed. We continue to work to catch up to our original timeline. The current report therefore updates all Specific Aims for Years 1 and 2 because our work this year included much of the work planned for Year 1. Note that spending was also delayed until work began, causing a substantial carryforward from Year 1 to Year 2, and a slightly lower carryforward from Year 2 to Year 3. Our Current Objectives are described within the Specific Aims, where we also discuss in detail results from numerous analyses we have performed over the past year. Due to space constraints, all Main Tables are included in the appendices rather than embedded in text.

Specific Aim 1: Merge data on workplace violence across four administrative datasets and generate descriptive epidemiological data on prevalence/socio-demographics of workplace violence in the Historical Administrative Data System (HADS) 2004-2009. Distinguish violent acts in terms of content and as either investigated but not charged, charged but not founded, or founded.

Update: In Year 1, we generated a coding scheme to classify offense types and identify which offenses should be categorized as workplace violence by reviewing a number of alternative classification schemes. We finally settled on the Bureau of Justice Statistics National Corrections Reporting Program (NCRP) classification system (United States Department of Justice). We categorized offense types into violent and non-violent crimes and also distinguished between crimes related to family (e.g., domestic violence) versus. crimes not related to family. We consider all crimes that are both violent and non-familial as “workplace violence,” as these are the violent crimes that were committed by Regular Army personnel while on active duty.

In Year 2, after doing a review of the literature, we slightly reorganized the categorization of crimes that was presented in the Year 1 report. We moved simple assaults and "other physical violence" to a category of minor physical violence offenses. We also divided verbal violence into minor verbal physical offenses (e.g., intimidation and harassment) and verbal sexual violence (e.g., sexual harassment; referred to as "minor" sexual violence). Cross-tabulations were then used to estimate standardized rates of the new set of violence outcomes. See attached Main Table 1 for the rates of perpetration and victimization of violent crime per 1,000 person-years (PY) in the HADS 2004-2009 data using the new categorization. Whereas the Year 1 report only presented rates in the total sample, rates are now also presented separately for male and female soldiers. Note that the vast majority of all crimes are "founded." After looking carefully at these standardized rates, we decided to continue to focus on "founded" crimes for the perpetration outcomes of interest. For victimization, we decided to define our outcomes based on any reported victimization (i.e., since the offender may not always be known or charged). Aggravated assaults made up the vast majority of major physical violent crimes (1.6 "founded" perpetrations and 1.0 victimizations per 1,000 PY in the total sample), simple assaults were the primary form of minor physical violence (9.1 perpetrations and 7.6 victimizations per 1,000 PY), and rape/sodomy/sexual assaults (2.3 perpetrations and 1.8 victimizations per 1,000 PY in the total sample) was more common than verbal sexual violence (0.3 perpetrations and 0.1 victimizations per 1,000 PY in the total sample).

Based on these standardized rates, we prioritized which workplace violence (i.e., non-familial) outcomes for which we would develop risk prediction tools. Specifically, we identified outcomes that were frequent enough to produce stable prediction equations among Regular Army soldiers: major physical violence perpetration and victimization (i.e., excluding sexual violence) among males, females, and in the total population; perpetration of major (e.g., rape, sexual assault) sexual violence among males; victimization of sexual violence among males, females and in the total population; and perpetration and victimization of minor physical violence among males, females and in the total population. We decided not to include sexual verbal violence because the frequency of perpetration/victimization was too low to obtain a stable prediction model.

Specific Aim 2: Analyze longitudinal profiles of recurrence of administratively-recorded workplace violence perpetration and victimization in the HADS.

Update: Our Year 1 report presented a table of recurrence of administratively recorded perpetration of workplace violence. We have now updated this table to correspond to the final set of outcomes described in Specific Aim 1. Disaggregation of the rates from Main Table 1 show that the percent of active duty Regular Army soldiers in service at any time between 2004-2009 who ever during this period were accused of perpetration of major physical, minor physical, or major sexual workplace (i.e., non-familial) violence were 0.6%, 2.9%, and 0.6%, respectively (see attached Main Table 2). Roughly one-quarter of perpetrations of major physical violence are recurrences rather than first offenses, and roughly one-third of perpetrations of both minor violence and major sexual violence are recurrences. Similar patterns of perpetration recurrence are seen only among male soldiers. In contrast, female soldiers had lower rates of perpetration recurrence (i.e., 15-20% of all major physical, minor physical, and major sexual perpetrations by females were recurrences). The percent of active duty Regular Army Soldiers in service who were ever victims of major physical, minor physical, or major sexual violence were

0.4%, 1.7% and 0.5% respectively. 6.3% of all soldiers who were victims of major physical violence were victimized more than once, 11.3% were repeated victims of minor physical violence, and 16.8% were repeated victims of major sexual violence. Similar victimization estimates were observed separately in male and female soldiers. Given that rates of single (i.e., first) perpetrations and victimizations were much higher than recurrences, we decided to prioritize the development of risk algorithms that predict the first occurrences of perpetration or victimization of our workplace violence outcomes of interest. However, we also plan to eventually develop risk prediction tools for reoccurrences of perpetration and victimization.

Specific Aim 3: Compare violence perpetration-victimization rates with consenting soldier administrative data for soldiers in the A-STARRS survey sample 2011-2013.

Update: In Year 1 we were delayed in receiving the updated, corrected consenting soldier administrative data for the A-STARRS survey samples. In Year 2 we have focused our analysis for this aim on survey data from the New Soldier Study (NSS). The NSS interviewed new soldiers recently arriving to Basic Combat Training occurred between April 2011 and December 2012, and we have person-month outcome data up until December 2013. Thus, the maximum number of months of follow-up for NSS respondents is 32 months (for soldiers interviewed in April 2011) while the minimum number of months follow-up is 12 months (for soldiers interviewed in December 2012). We consequently developed a weight for the NSS sample so that we could compare NSS rates of perpetration and victimization to rates in the 2004-2009 HADS. This weight made adjustments so that survey responders who completed the NSS and gave permission to link to their administrative data were representative of all soldiers who took the NSS survey regardless of whether they completed or gave administrative data. The weight also made adjustments so that the NSS sample would be representative of the total Army population 2011-2013. However, it is important to note that the Army provided us with a limited number of covariates to develop our weight. For example, we were given no information on perpetration and victimization of crime in the total population from 2011-2013. Thus, it is possible there is bias in the NSS rates presented here. Nonetheless, we used the best-available weight to compare rates of perpetration and victimization of workplace violence in the 2011-2013 NSS sample to the 2004-2009 HADS sample.

Main Table 3 (attached) shows the rate of first occurrences of perpetration and victimization of major physical, minor physical, and major sexual workplace (i.e., non-familial) violence per 1,000 PY in the NSS (April 2011- Dec 2012) compared to soldiers from the 2004-2009 HADS whose length of time in the Regular Army was 32 months or fewer (i.e., the maximum number of NSS follow-up months). The rates of major sexual violence perpetration and victimization are comparable in the NSS (perpetration = 2.7/1,000 PY; victimization = 4.1/1,000 PY) and HADS (perpetration = 2.7/1,000 PY; victimization = 3.9/1,000 PY). However, rates of physical violence perpetration and victimization (both major and minor) are significantly lower in the NSS (perpetration = 1.8/1,000 PY to 6.8/1,000 PY; victimization = 0.8/1,000 PY to 4.5/1,000 PY) than the 2004-2009 HADS (perpetration = 2.9/1,000 PY to 13.1/1,000 PY; victimization = 2.2/1,000 PY to 7.8/1,000 PY). Similar patterns were observed separately among male and female soldiers.

Specific Aim 4: Generate descriptive epidemiological data on prevalence/socio-demographic correlates of self-reported workplace violence perpetration/victimization in the A-STARRS surveys.

Update: We have created the outcomes for self-reported workplace violence perpetration and victimization in the All-Army Study (AAS), which uses a sample of active duty Regular Army soldiers not deployed, and in the post-deployment phase of the Pre-Post Deployment Study (PPDS) of three Brigade Combat Teams assessed just before deployment and then again just after returning from deployment to Afghanistan. In Year 1, we reported the prevalence of self-reported workplace violence from the AAS and PPDS using a preliminary weighted AAS dataset from April 2011-December 2011 (only the first 3 quarters of the AAS) that had not yet been consolidated with the AAS In-Theater data or National Guard/Army Reserve data. We also used a preliminary PPDS file that hadn't been updated with the final cohort of consenting soldiers (at the time of the Year 1 report we were still collecting follow-up data for our final time point).

The final AAS and PPDS datasets were created this past summer and we now report the updated rates of self-reported workplace violence in these surveys (see attached Main Table 4). Both surveys asked respondents to report how often they were verbally violent in the past 30 days (i.e., either yelled, insulted, swore, or threatened someone; had a heated argument; or got into a loud argument in a public place). 17.2% of respondents in the AAS and 13.8% of those in the PPDS reported perpetrating at least one form of verbal violence over this time period. The survey also included a question about "physical confrontation" during an argument. A total of 3.0% of AAS respondents and also 3.1% of PPDS respondents reported this kind of experience occurring in the 30 days before interview. Respondents were also asked, unrestricted by time, whether they had ever hit another person to the point of bruising and/or healthcare needs. 2.4% of respondents in the AAS and 6.2% of those in the PPDS reported such experiences. As expected, male soldiers virtually always had higher rates of these self-reported perpetration behaviors than female soldiers.

The AAS and PPDS also asked a series of questions about victimization (see Main Table 4). These included questions about experiences of physical assault and rape over the past 12 months as well as ever during a deployment. Past year physical assault victimization was reported by 1.1-1.2% of respondents (across the two surveys) and past year sexual assault victimization by 0.2-0.5%. In regard to experiences during deployment, 1.3-2.1% reported physical assault victimization during deployment (across the two surveys), 0.3-1.0% reported sexual assault victimization, and 3.8-4.1% reported being bullied by unit members. Whereas self-reported rates of physical assault victimization were generally higher among men, rates of sexual assault victimization and bullying were consistently higher among women.

We then looked at socio-demographic and career history correlates of self-reported perpetration and victimization (see attached Main Table 5). We had to collapse responses across certain outcome items so that the coefficients would be stable (e.g., perpetration of any verbal violence; perpetration of any physical violence; victim of any violence). These results are presented in Main Table 5. Overall, significant associations were fairly consistent across the AAS and PPDS. Most were also consistent with expectations, thus we do not discuss them all in detail here. For example, females had significantly lower odds of reporting violence perpetration (ORs = 0.5-0.7) but significantly higher odds of reporting victimization (ORs = 1.6-2.5). Soldiers over 30 years old had significantly lower odds of perpetrating violence (ORs = 0.1-0.5) but higher odds of victimization (ORs = 3.5-14.1). A similar pattern was found for months in active and non-active service; soldiers with 86+ months in service were at significantly lower odds of self-reported perpetration (ORs = 0.5-0.7) but higher odds of being a victim (ORs = 3.6-5.3). The only significant association for race-ethnicity was for self-reported verbal violence in the AAS; soldiers identifying as “other” race were at lower odds of verbal violence perpetration than Non-Hispanic White soldiers. Interestingly, these findings for race contrast those found in our HADS and NSS models that defined perpetration outcomes using administrative crime records (see Specific Aims 7 and 8 below). This suggests possible bias either in the likelihood of arrest and/or the likelihood of self-reporting violence perpetration. We plan to further examine these potential biases over the next year as we pursue Specific Aim 5.

Specific Aim 5: To study patterns-predictors of under-reporting consenting soldier administrative data system reports of workplace violence victimization in the AAS and PPDS samples by comparing self-reported with administratively-recorded victimization. Develop correction procedures and a computer program for Army leadership to use in adjusting for under-reporting in future analyses of Army administrative databases.

Update: Given the delays in obtaining administrative data in Year 1, we have not yet been able to link AAS and PPDS self-reported responses to administrative criminal records. Thus, no progress has been made on this aim.

Specific Aim 6: Analyze longitudinal profiles of recurrence of administratively-recorded and self-reported workplace violence in the AAS and PPDS surveys.

Update: Given the delays we experienced obtaining administrative data in Year 1, we have not yet been able to link AAS and PPDS responses to administrative criminal records. Thus, no progress has been made on this aim.

Specific Aim 7: Use data mining methods to develop prediction equations for HADS predictors of administratively-recorded workplace violence perpetration-victimization in the HADS dataset. Parallel prediction equations will be estimated for a number of different types of violence to determine if the significant predictors vary depending on the severity, persistence, or character of the violent acts. Cross-validate final prediction equations. Develop a computer program for Army leadership to generate risk scores for individual soldiers based on HADS profiles.

Update: We spent the majority of Year 2 working on Specific Aims 7 and 8 and have made substantial progress. In regard to Aim 7, we prioritized the development of HADS algorithms to predict the first occurrence of 10 outcomes: major physical workplace violence perpetration (separately among male and female soldiers and also in the total sample), major physical workplace violence victimization (males; females; total sample), major workplace sexual violence victimization (males; females; total sample), and major workplace sexual violence perpetration (among males only). We are still in the process of analyzing the models for first perpetration and victimization of minor violence (men; women; and total sample) and these results are not included in this report. We also plan on developing parallel models in the future that predict the *recurrence* of workplace violence.

Developing the prediction equations has been a long and arduous task. It has involved consultation with many statisticians and experts in the field of data mining to develop a methodology that would identify not only the optimal area under the receiver operating characteristic curve (AUC) and concentration of risk (CR, i.e., the percent of perpetrations/victimizations occurring among soldiers with the highest predicted risk probabilities) but also provide a list of the interpretable predictors that could later be used to "dig down" into the theoretical implications of the models. It is also noteworthy that the data mining methods that were ultimately used required substantial computing power (e.g., cross-validation models could take as long as one week to converge).

HADS analyses began by defining the administrative outcomes in the 2004-2009 data, generating person-year samples for each of the 10 outcomes, and reviewing bivariate associations with four types of potentially important administrative predictors of workplace violence perpetration and victimization (i.e., socio-demographic factors, military career experiences, prior crime experiences, and health/stress [mental and physical]). Several hundred significant predictors were found in these bivariate models, many of which were consistent with study hypotheses. Accordingly, below we review some, but not all, of the significant bivariate associations. Tables for these bivariate results are also available upon request.

HADS bivariate models. In the total sample, perpetration of major physical violence was significantly inversely related to: age at enlistment, current deployment, rank, time in service, AFQT score, age, education, and number of dependents. It was also positively associated with being male, in the infantry, unmarried, Non-Hispanic Black or Hispanic, having prior perpetration or victimization of other violent crimes, having a positive drug test in the past 3 months, having healthcare visits for impulse control disorders, taking medications for anxiety, chemical dependency, or cognitive disorders, having registered guns, percentage of combat deaths and suicide deaths in one's unit over the past year, and number of different duty units over the past year. Bivariate associations with victimization of major physical violence were very similar to those of perpetration. Inverse associations were found for age of enlistment, rank, time in service, AFQT score, current age, education, and number of dependents. Positive associations were found for being never married, Non-Hispanic Black, and female, prior victimization of minor physical, major sexual, and sexual verbal violence, prior perpetration of minor crime and drunkenness/disorderly conduct, having a positive drug tests in the past 3 months, having healthcare visits for impulse control disorders, and using medications for anti-anxiety, chemical dependency, and bipolar disorder.

We also found similar significant bivariate associations with perpetration and victimization of major sexual violence. Among male soldiers, perpetration of major sexual violence among males was significantly inversely related to age at enlistment, current deployment, rank, time in service, AFQT score, age, education, number of dependents, being never being married, and Non-Hispanic Black and Hispanic. Perpetration of sexual violence was also positively associated with being in the infantry, prior perpetration and victimization of violent crime, being prescribed sedative-hypnotics or medications for chemical dependency, healthcare visits for impulse control disorder, hospitalizations for suicidal behaviors in the past 12 month, and number of different duty units over the past 12 months. In regard to bivariate associations with major sexual violence victimization in the total sample, significant positive associations were found for being female, Non-Hispanic White, prior victimization of physical violence, prior perpetration of sexual violence, prior drunkenness/disorderly conduct, having a positive drug test in the past 3 months, and treatment for impulse control disorders and chemical dependency.

Using data mining to develop the HADS risk equations. Importantly, there were strong inter-correlations among the 400+ HADS predictors, making it impossible to include all the individually significant predictors in a single logistic regression equation for each workplace violence outcome. Our data mining approach consequently used (i) cross-validated stepwise logistic regressions to identify a final parsimonious set of predictors for each outcome, (ii) cross-validated random forests to search for interactions among predictors, and (iii) elastic net penalized regressions to stabilize predictor coefficients (i.e., due to multicollinearity). This data mining approach is very similar to what we have been using to develop suicide risk algorithms in A-STARRS (see Kessler et al., in press attached as an Appendix). We started by running cross-validated stepwise logistic regressions to determine the best number of predictors for each outcome. In addition to running cross-validated models that did not restrict the number of predictors allowed to "step in," we also specified several increasingly restrictive models to reduce the number of predictors to a more manageable number (e.g., 10, 15, 20, 25, 50, or 75 predictors). The AUCs and CRs derived from cross-validated stepwise regression models were then evaluated in order to select the final set of predictors for each outcome that resulted in the highest classification accuracy and optimal CR. This was done for all 10 workplace violence outcomes, and results suggest that the models restricted to allow 25 predictors (i.e., for major physical violence perpetration among males and in the total sample) or 20 predictors (i.e., for major physical violence perpetration among females; major physical violence victimization among males, females, and in the total sample; major sexual violence perpetration among males; and major sexual violence victimization among males, females, and in the total sample) had the highest classification accuracy (AUCs = 0.7 to 0.9) and largest proportions of the particular outcome occurring among those in the top 5% of predicted risk (CRs = 21.3% to 72.9%).

After selecting the best cross-validated stepwise logistic regression for each outcome we then ran a non-cross-validated model restricted to the optimal number of predictors (i.e., 20 or 25 predictors depending on the outcome) and output the odds ratios (ORs). We did this because of our interest in the theoretical implications of the models (see Appendix Tables 1-10 for the list of final predictors for each HADS outcome). We also looked for interactions among predictors by running random forest models for each outcome. Specifically, we generated predicted probabilities for each outcome using cross-validated random forest models on the final sets of predictors for each outcome. To determine the importance of any predictor interactions, we subsequently estimated 3 logistic regression models for all 10 outcomes using (i) the stepwise regression predicted probability alone, (ii) the random forest predicted probability alone and (iii) both the stepwise and random forest predicted probabilities. We compared the AUCs and CRs for these models to determine the incremental predictive validity of the random forests models. For all outcomes in the HADS 2004-2009, the predicted probabilities from the random forest models did not improve AUC or CR over and above the corresponding stepwise regression models. We consequently excluded the random forest predicted probabilities from all subsequent models.

Next, we ran cross-validated elastic net penalized regression models with various mixing parameter penalties (i.e., setting alpha = 0.1, 0.3, 0.5, 0.7, 0.9, and 0.99) using the best set of predictor variables from the non-cross-validated stepwise regression. As mentioned earlier, we used elastic net as a way of correcting for multicollinearity among the final sets of predictors. We evaluated the AUC and CR for each specified mixing parameter penalty and selected the best final elastic net model for each outcome. For all 10 outcomes, the alpha = 0.5 models resulted in highest classification accuracy and CR. We subsequently ran non-cross-validated elastic nets with alpha = 0.5 and compared the AUC, CR, and size of the coefficients of the best predictors to the results from the non-cross-validated stepwise regressions. The final results for perpetration and victimization of non-familial major physical workplace violence can be found in Main Table 6 (attached), whereas the final results for perpetration and victimization of non-familial major sexual workplace violence can be found in Main Table 7 (attached).

In regard to the classification accuracy of these final models, performance of the non-cross-validated stepwise logistic regressions were virtually identical to the performance of the non-cross-validated elastic nets; both approaches resulted in AUCs of 0.7 (for sexual violence perpetration among males), 0.8 (for major physical violence perpetration and victimization among males, females, and in the total sample; major sexual violence victimization among males and females), or 0.9 (for major sexual violence victimization in the total sample). As an AUC of 0.5 suggests no better prediction accuracy than would be expected by chance, the AUCs achieved in our analyses collectively suggests that our final models performed quite well at predicting the 10 workplace violence outcomes. CRs were also similar across the final stepwise and elastic net models for each outcome but did vary slightly depending on the outcome. For all major physical violence perpetration and major sexual violence victimization models (6 of the 10 outcomes), over 30% of all first occurrences of the outcomes occurred among soldiers in the top 5% of predicted risk derived (33.1-72.9% of these 6 outcomes occurred among soldiers in the top 5% of predicted risk). In comparison, the prediction models for major physical violence victimization and major sexual violence perpetration (i.e., the remaining 4 outcomes) had slightly lower CRs (i.e., 21.9-28.8% of these 4 outcomes occurred among soldiers with the top 5% of predicted risk). The strongest major physical violence model (Main Table 6) was for major physical violence perpetration in the total sample – 37.2% (elastic net) to 37.4% (stepwise) of perpetrations occurred among soldiers in the top 5% of predicted risk. The weakest physical violence model was for major violence victimization in the total sample, where 24.5% (elastic net) to 24.7% of all victimizations occurred among soldiers in the top 5% of predicted risk. The strongest sexual violence model (Main Table 7) was for major sexual violence victimization in the total sample, resulting in by far the highest CR (71.7% [elastic net] to 72.9% [stepwise] of all victimizations occurring among soldiers in the top 5% of predicted risk). This occurred primarily because female soldiers have much higher rates of sexual victimization than male soldiers, and gender stepped-in as one of the best 20 predictors for this outcome in the total sample. In contrast, the weakest sexual perpetration/victimization model was for predicting major sexual violence perpetration among males; 21.9% of all occurrences were perpetrated by male soldiers with the top 5% of predicted risk.

In order to further interpret the performance of our final models, we also calculated the standardized rates of each outcome per 1,000 PY within six discretized CR groups (i.e., top 1%, 2-5%, 5-10%, 10-45%, 45-100%). Similar to the results reported above for CR, these rates were similar but varied slightly across the final non-cross-validated stepwise logistic regression and elastic net models for each outcome. The highest rates of physical violence perpetration/victimization (Main Table 6) were found in the model predicting major physical violence perpetration among male soldiers; there were 29.2 (elastic net) to 29.4 (stepwise) first occurrences of major physical violence perpetration per 1,000 PY among soldiers in the top 1% of predicted risk. In comparison, there were 0.5 first occurrences per 1,000 PY among male soldiers in the bottom 45-100% of predicted risk (for both elastic net and stepwise models) and 3.2 first occurrences of major physical violence perpetration per 1,000 PY among all male soldiers in the sample (Main Table 3). The highest rates of sexual violence perpetration/victimization (Main Table 7) were found in the model predicting major sexual violence victimization among female soldiers; there were 119.3 (elastic net) to 120.4 (stepwise regression) first occurrences of sexual violence victimization per 1,000 PY among female soldiers in the top 1% of predicted risk. In comparison, there were 1.8 first occurrences per 1,000 PY among male soldiers in the bottom 45-100% of predicted risk (for both elastic net and stepwise models) and 22.5 first occurrences of major physical violence perpetration per 1,000 PY among all female soldiers in the sample (Main Table 3).

Caution is needed in interpreting the final sets of 20 or 25 predictors for each outcome as the data mining methods employed maximized overall prediction accuracy rather than individual coefficient accuracy. Although we do not review the full final predictor sets for each outcome here, it is nonetheless noteworthy that (i) several predictors were identical across outcomes and (ii) all four classes of administrative predictors (socio-demographic factors, military career experiences, prior crime experiences, and health/stress) were important. See Appendix Tables 1-10 for the list of final predictors for each model and the associated ORs from the non-cross-validated stepwise logistic regression and elastic net models. For example, the final set of predictors for the 4 major physical/sexual perpetration outcomes (see Appendix Tables 1-3 and 7) each included Non-Hispanic Black (stepwise ORs = 1.8-3.7; elastic net ORs = 1.4-2.0), current deployment (stepwise ORs = 0.2-0.4; elastic net ORs = 0.3-0.4), being demoted in the past 12 months (stepwise ORs = 1.5-2.1; elastic net ORs = 1.5-1.6), and perpetrating any crime in the prior 24 months (stepwise ORs =1.6-2.1; elastic net ORs = 1.4-2.0). Likewise, all 3 of the final major physical violence victimization models (see Appendix Tables 4-6) included current deployment (stepwise ORs = 0.6; elastic net ORs = 0.6), having a rank of E5 or higher (stepwise ORs = 0.6-0.7; elastic net ORs = 0.6-0.7) or E7 or higher (stepwise ORs = 0.4; elastic net ORs = 0.4), having an AFQT score below 50 (stepwise ORs = 1.3-1.4; elastic net ORs = 1.3-1.4), perpetrating any crime in the prior 12 months (stepwise ORs = 1.2-1.6; elastic net ORs = 1.2-1.6), being the victim of any crime in the prior 12 months (stepwise ORs = 1.6-3.0; elastic net ORs = 1.6-2.9), and average percent of combat deaths over the past year (stepwise ORs = 1.6-4.0; elastic net ORs = 1.6-3.7). Further, all 3 of the final major sexual violence victimization models (see Appendix Tables 8-10) included being over 23 years old (stepwise ORs = 0.6; elastic net ORs = 0.6-0.7), having a rank of E5 or higher (stepwise ORs = 0.3-0.5; elastic net ORs = 0.5-0.6), having 25+ total months in service (stepwise ORs = 0.6; elastic net ORs = 0.6), total number of days in the past 12 months with outpatient visits for a mental health diagnosis (stepwise ORs = 1.2-1.3; elastic net ORs = 1.1-1.3), and number of different duty units over the past 12 months (stepwise ORs = 1.2-1.3; elastic net ORs = 1.1-1.2).

Specific Aim 8: Use data mining methods to develop prediction equations for survey, neurocognitive, genetic, and limited IADF predictors of administratively-recorded workplace violence perpetration and victimization in the NSS. Cross-validate final prediction equations. Develop a computer program for Army leadership to generate risk scores for individual soldiers based on IADF profiles.

Update: Main Table 8 (attached) shows the standardized rates per 1,000 PY for major physical, minor physical, and major sexual workplace violence perpetration and victimization in the NSS 2011-2013. Because the vast majority of all first occurrences of non-familial major physical and major sexual violent crimes were perpetrated by males, we restricted analysis of perpetration of these 2 outcomes to males only (i.e., females accounted for only 3 occurrences of major physical violence perpetration and 3 occurrences of sexual violence perpetration). We restricted the analysis of first victimization of major sexual violence to females for the same reason. For perpetration and victimization of minor physical crime we focused on the total sample. In comparison, there were too few victimizations of major physical violence to produce stable results and thus we will need to return to this outcome when we have more months of follow-up.

Similar to the HADS, we built NSS person-month discrete-time survival analysis data files to predict first onsets of the workplace violence outcomes from temporally primary survey data. These person-month files start at the month of interview (in the NSS this would be the month of arriving to Basic Combat Training) and follow soldiers until December 2013. As mentioned earlier, the interviews were conducted from April 2011 – December 2012. In other words, we had between 12 and 32 months of follow-up for new soldiers depending on when they completed the NSS survey. Before building the person-month files, we constructed hundreds of predictors from the NSS survey. There are many advantages of survey predictors over the administrative predictors used in HADS. For example, we had validated self-report measures of lifetime and 30-day mental disorders (regardless of treatment), whereas our administrative predictors of mental disorder were based solely on records of outpatient and inpatient treatment visits (the HADS could not be used to operationalize mental disorders among soldiers who did not have such visits). In addition, the surveys have a much richer array of predictors, including childhood adversities, traumatic events, stressors, lifetime and 30-day mental disorders, personality scales, and social network information. However, a significant limitation of the current NSS sample is that we have thus far only been able to follow soldiers for 12-32 months (through December 2013). We therefore have only a small number of outcomes to predict using the survey data. For instance, the sample for perpetration of major physical violence in the NSS had 366,263 person-months and only 56 first occurrences of major physical violence. In comparison, we had over 30 million person-months and 5,300 first occurrences of major physical violence perpetration in the 2004-2009 HADS models. As such, we have many fewer significant predictors even though the predictors are substantively richer than those in HADS.

NSS bivariate models. Consistent with our approach in the HADS, we began by examining bivariate associations between each of the survey predictors with the 5 NSS outcomes: major physical perpetration and major sexual perpetration among males, major sexual victimization among females, and minor physical perpetration and victimization in the total sample. Here we review some, but not all, of the significant bivariate associations. Most significant associations were consistent with study hypotheses. Tables for these bivariate results are also available upon request.

The strongest bivariate associations with first perpetration of major physical violence among males included positive associations with being Non-Hispanic Black, having any medical failure at accession, having had 4 or more years with anger attacks, 1 or more years with problem of conduct/behavior, 3 or more years with alcohol or drug problems, having 5 or more sexual partners in the past 12 months, and having severe or very severe stress in the past 12 months. Perpetration of major physical violence among males also had inverse associations with being beyond the first year of service, current symptoms of a major depressive episode, number of years with depression and anxiety, number of lifetime panic attacks, being previously bullied, having one or more parents with a history of depression or anxiety, and having an introverted personality. Perpetration of sexual violence among males had significant positive bivariate associations with being Non-Hispanic Black and Hispanic, E1 rank, having an AFQT scores below the median, a lifetime history of oppositional defiant disorder, being sent to a juvenile detention center as a child, being physically abused as a child, and religiosity, and significant inverse associations with levels of introversion, parent history of mania, having relatives with drinking problems, and having a history of self-harm behaviors. Perpetration of minor physical violence in the total sample was associated with many of the same predictors described above, in addition to several other predictors (due to their being many more cases of minor violence perpetration compared to the other outcomes). Significant positive associations with minor physical violence perpetration were found for being Non-Hispanic Black, currently or previously married, having a lifetime history of intermittent explosive disorder, substance use disorders, and conduct disorder, total count of lifetime disorders, number of years with insomnia, number of years with anger attacks, number of years with problem conduct/behaviors, being previously physically assaulted, being in a foster home or juvenile detention as a child, anger/irritability, antisocial personality, social anhedonia, and having severe or very severe stress in the past 12 months. Minor physical violence perpetration had significant inverse associations with soldier's education, parent education, and introverted personality.

Major sexual violence victimization among females was significantly inversely related to rank, age at accession, AFQT score, being Non-Hispanic Black, Christian, and having one or more dependents. Several more variables were positively associated with this outcome: having a lifetime PTSD or oppositional defiant disorder diagnosis, number of lifetime physical and sexual assaults, total number of lifetime traumatic events, being bullied ten or more times in one's life, parental history of psychopathology and violence, being sexually and physically abused as a child, sexual harassment as a child, living in a foster home, the total number of childhood adversities, social anhedonia, affectivity lability, sensation seeking traits, impulsivity, lifetime suicide ideation, and number of lifetime occurrences of self-harm behaviors. In regard to minor physical violence victimization in the total sample, positive associations were found for: being female, being born in the U.S., having an E1 rank, having a lifetime history of mania, total number of lifetime mental disorders, having 11 or more lifetime panic attacks, 6 or more years with depression, number of lifetime sexual assaults, total number of lifetime traumas, several childhood adversities (e.g., being in foster care, childhood sexual assault, childhood sexual harassment), having 2 or more sexual partners in the past 12 months, and having severe or very severe stress in the past 12 months. Inverse associations with minor physical violence victimization were found for AFQT scores and personality traits of agreeableness, impulsivity, and dispositional optimism.

Using data mining to develop the NSS risk equations. After evaluating the bivariate associations described above, we applied the same data mining techniques described in Specific Aim 7 (i.e., cross-validated stepwise logistic regression to identify the final set of predictors, non-cross-validated stepwise regression and elastic net models to determine final model performance for each outcome), with the only difference being that the predictors were from the NSS survey rather than from HADS administrative data. Thus, we do not re-report the order of operations in detail here. There was, however, one noteworthy difference in the NSS data mining analyses. A number of NSS predictors had to be excluded from the models predicting major physical and major sexual violence perpetration because the low rates of these outcomes resulted in highly unstable coefficients. As a result, fewer predictors (compared to the HADS models) stepped in to the cross-validated stepwise regression models that were used to determine the final set of predictors for each of the five outcomes (e.g., unrestricted models started with between 38 and 132 predictors, with a maximum of 8 to 22 unrestricted predictors ultimately stepping in).

AUCs and CR from the cross-validated stepwise regressions results suggest that the models restricted to allow 5 (i.e., for perpetration of major physical violence among males and major sexual violence victimization among females), 10 (i.e., for perpetration of major sexual violence among males and minor violence victimization in the total sample), or 15 (i.e., for minor violence perpetration in the total sample) predictors had the highest classification accuracy (AUCs = 0.7-0.8) and the largest proportions of the outcomes occurring among soldiers with the top 5% of predicted risk (CRs = 12.0% to 26.4%). After selecting the best cross-validated stepwise regression for each outcome, we then ran non-cross-validated models restricting to the optimal number of predictors (i.e., 5, 10, or 15) to obtain ORs (see Appendix Tables 11-15 for the list of final predictors for each NSS outcome). We also ran cross-validated random forests models for each outcome in effort to identify interactions among the predictors, but similar to the HADS models the random forests predicted probabilities did not meaningfully improve AUC or CR and were thus excluded from subsequent models.

The final predictor sets for each outcome were then evaluated using cross-validated elastic net models using various mixing parameter penalties to correct for multicollinearity among predictors (i.e., setting alpha = alpha = 0.1, 0.3, 0.5, 0.7, 0.9, and 0.99). Similar to the 10 HADS models, the 5 NSS outcomes performed best at an alpha of 0.5. We then ran non-cross-validated elastic nets with alpha = 0.5 and compared the AUC, CR, and size of the coefficients of the best predictors to the non-cross-validated stepwise logistic regression models. The final performance of these final models can be found in Main Table 9 (attached). For all 5 outcomes, performance of the stepwise regression and elastic net models were virtually identical in regard to overall classification accuracy (stepwise regression AUCs = 0.7 to 0.8; elastic net AUCs = 0.7 to 0.8). Although CRs were also similar across stepwise regression and elastic net models, it is noteworthy that the elastic net performed somewhat better at predicting major sexual violence perpetration among males (stepwise CR = 20.6% versus elastic net CR = 26.4% of all first perpetrations of sexual violence occurring among male soldiers in the top 5% of predicted risk).

The stepwise regression and elastic net models predicting perpetration of major physical violence among males had the highest CR in the top 5% (28.3% [stepwise regression] to 30.2% [elastic net] of all first occurrences of major physical violence perpetration occurring among male soldiers in the top 5% of predicted risk), whereas the models for minor physical violence victimization in the total sample had the lowest CR in the top 5% (18.4% [for both stepwise regression and elastic net] of first victimizations of minor physical violence occurring among those in the top 5% of predicted risk). In comparison to the HADS models described in Specific Aim 7, the NSS models had similar AUCs but consistently lower CRs. This is not surprising as the NSS models used (i) much smaller samples (i.e., fewer occurrences of the defined outcomes), and (ii) fewer predictors (i.e., some predictors had to be omitted from analysis due to coefficients instability). Importantly, we expect the performance of these NSS models to improve once we obtained an additional 12-18 months of follow-up outcome data.

We also calculated the standardized rates of each outcome per 1,000 PY within each of the discretized CR groups (top 1%, 2-5%, 5-10%, 10-45%, 45-100%). The highest rates of physical violence perpetration/victimization were found in the models predicting minor physical violence perpetration in the total sample; there were 57.2 (stepwise regression) to 62.1 (elastic net) first occurrences of minor physical violence perpetration per 1,000 PY among new soldiers in the top 1% of predicted risk (Main Table 9). In comparison, there were 2.5 (stepwise regression) to 2.9 (elastic net) first occurrences per 1,000 PY among all soldiers in the bottom 45-100% of predicted risk and 6.3 first occurrences of minor physical violence perpetration per 1,000 PY among all new soldiers in the sample (Main Table 8). The highest rates of sexual violence perpetration/victimization were found in the model predicting major sexual violence victimization among new female soldiers; there were 112.6 (elastic net) to 157.3 (stepwise regression) first occurrences of sexual violence victimization per 1,000 PY among female soldiers in the top 1% of predicted risk. In comparison, there were 13.6 (elastic net) to 14.4 (stepwise regression) first occurrences per 1,000 PY among all soldiers in the bottom 45-100% of predicted risk and 22.0 first occurrences of minor physical violence perpetration per 1,000 PY among all new soldiers in the sample (Main Table 8).

As mentioned is Specific Aim 7, although caution is needed in interpreting the final set of predictors for each outcome, it is noteworthy that some predictors were identical across NSS perpetration outcomes (see Appendix Tables 11-15 for the list of final predictors for each model and the associated ORs from the non-cross validated stepwise regression and elastic net models). For example, the final set of predictors for the 3 perpetration outcomes each included predictors for Non-Hispanic Black race (stepwise ORs = 2.0-5.0; elastic net ORs = 2.0-4.3) and introverted personality (stepwise ORs = 0.8-0.9; elastic net ORs = 0.8-1.0). Notably, the predictor for Non-Hispanic Black race was also included as one of the final predictors in all 6 of the HADS perpetration models. In comparison, although there was no exact overlap among the final set of predictors for the 2 NSS victimization models, variables representing certain personality traits (e.g., thoughtfulness, impulsivity) and past trauma/childhood adversities were consistently included among the final set of predictors for both models (see Appendix Tables 14-15).

Specific Aim 9: Use data mining methods to develop prediction equations for retrospective IADF and self-report survey predictors of administratively-recorded workplace violence perpetration-victimization in the AAS and PPDS datasets. Cross-validate final prediction equations. Develop a computer program for Army leadership to generate risk scores for individual Soldiers based on IADF profiles.

Update: As noted above, in Year 2 we focused on analysis of the consenting soldier administrative data system reports of workplace violence victimization in the NSS and in the 2004-2009 HADS. No progress has been made on this aim.

KEY RESEARCH ACCOMPLISHMENTS

- We generated a coding scheme to classify offense types and identify which offenses should be categorized as workplace violence by reviewing a number of alternative classification schemes. We finally settled on the Bureau of Justice Statistics National Corrections Reporting Program (NCRP) classification system (United States Department of Justice). We categorized offense types into violent and non-violent crimes and also distinguished between offense types related to family versus crimes not related to family. All crimes that are both violent and not related to family are classed as "workplace violence," as these are the violent crimes that were committed by Regular Army personnel while on active duty. We used this coding scheme to classify all of the reported occurrences of criminal activity in the following administrative datasets: Centralized Operations Police Suite / Military Police Reporting System (COPS/MPRS), Criminal Investigation Division Information Management System / Automated Criminal Investigation/Criminal Intelligence (CIMS/ACI2), Criminal Investigation Division Information Management System / Automated System Crime Record Center (CIMS/ASCRC), Army Court Martial Information System (ACMIS), Sexual Assault Data Management System (SADMS).
- Samples were drawn to analyze longitudinal profiles of first occurrence and recurrence of administratively-recorded perpetration and victimization of workplace (non-familial) major physical (murder, homicide, manslaughter, aggravated arson, aggravated assault, kidnapping, and robbery), minor physical (simple assault and verbal violence such as harassment, blackmail, extortion and intimidation), and major sexual violence (rape, sodomy and sexual assault) in the HADS 2004-2009.
- We created administrative recorded predictors of workplace violence. The HADS was created by merging data from 38 different Army and DoD data systems for each soldier on active duty in the Army over the study period (we had information on predictors from 2000-2009). Separate observational records were created for each of the over 37 million person-months of active duty service of the over 975,000 Regular Army soldiers in the HADS, each record containing complete historical administrative information for the soldier in question. We coded 5 sets of administrative predictors from this data including socio-demographics, career history variables, prior criminal activity, health (mental and physical) and stress, and unit-level variables
- Using the same coding scheme developed in bullet point #1, we classified offenses from the NSS consenting soldier administrative data 2011-2013. We generated a person-month data files starting with the month of interview and following the soldier prospectively from 12-32 months coding each occurrence of major physical, minor physical and major sexual workplace violence across the time period (soldiers interviewed in April 2011 were followed for 32 months while soldiers interviewed in December 2012 were followed 12 months). Separate observational records were created for each of the 400,000+ person-months of active duty service among 20,000+ Regular Army soldiers in the NSS.

- We developed 10 sets of survey predictor and limited administrative predictor variables for the NSS samples including: socio-demographic, accession variables, self-reported mental disorders, traumatic events, childhood adversities, personality scales, social support variables, self-reported suicidal behavior (including ideation, plan, attempt, and self-harm behaviors), injuries, and self-reported treatment for mental disorders.
- Discrete-time survival analysis with time-varying covariates was used to study bivariate predictors of workplace violence in both the HADS 2004-2009 and the NSS consenting soldier administrative data 2011-2013.
- A state-of-the-art ensembling machine learning methodology was used to develop optimal risk prediction equations identifying soldiers at highest risk of perpetration/victimization of workplace violence in both the HADS 2004-2009 and the NSS 2011-2013. Thus far we have developed preliminary algorithms for 10 outcomes in the HADS and 5 outcomes in the NSS.
- We are working with the Army Analytics Group (AAG) to create a continuously-updating version of the HADS for real-time implementation in targeting and matching.

CONCLUSION

In summary, we have made substantial progress on many of our specific aims over the past year. First, we have produced final standardized rates (per 1,000 PY) for several different forms of workplace violence perpetration and victimization (e.g., Main Tables 1-4; Main Table 8). Collectively, these rates confirm that a high proportion of soldiers in the HADS, NSS, AAS, and PPDS samples have perpetrated and/or been a victim of workplace violence. In general, male soldiers had higher rates perpetrating workplace violence (particularly major physical and major sexual violence), while female soldiers had higher rates of being victims of violence (particularly minor physical violence and major sexual violence). These findings underscore the overall importance of our research. Second, we have developed several preliminary risk algorithms to predict the first occurrence of workplace violence perpetration and victimization in the HADS and NSS. The data mining methods used to develop the risk equations are cutting-edge. The models are useful in that that can identify soldiers at the highest risk of workplace violence perpetration and victimization with accuracy much better than would be expected by chance (i.e., based on AUCs). Soldiers determined to be at highest risk (i.e., in the top 5% of predicted risk) also accounted for a large proportion of first occurrences of workplace violence perpetration and victimization. The coefficients from the bivariate analyses and final (i.e., best model) predictor sets corresponding to each outcome provide important information about specific risk and resilience factors of workplace violence perpetration and victimization that may eventually be integrated into targeted preventative interventions. The ultimate clinical utility of our risk algorithms is in their ability to identify soldiers with high perpetration/victimization risk. As it is both infeasible and impractical to offer preventive interventions for workplace violence perpetration and victimization to all soldiers, our algorithms can be used to prioritize who should be offered such programs (e.g., requiring soldiers in the top 1% or top 5% of predicted risk to attend perpetration or victimization prevention programs).

Moving forward, we have several plans to allow us to further accomplish our project goals and objectives. First, we are currently in the processes of conducting analyses in the HADS to determine if we can identify theoretically-guided high risk “subtypes” of soldiers who perpetrate or are the victim of violence. Specifically, we are using cluster analysis to see if we can find soldiers who share similar characteristics (i.e., predictors) and may thus benefit from preventive interventions tailored to those specific characteristics. Second, we plan to develop parallel risk equations that predict the *recurrence* of perpetration and victimization of workplace violence in the HADS. Third, we plan to re-estimate our data mining models in the NSS as we obtain additional (i.e., longer) follow-up outcome data. This will allow us to develop additional NSS perpetration and victimization risk algorithms (e.g., for major physical violence victimization) as well as likely improve the performance of the models presented in this report. Fourth, we will merge administrative perpetration/victimization outcome data with the AAS and PPDS survey data in order to complete Specific Aims 5 and 6. Fifth, we plan to use the data mining methods employed in Year 2 to create parallel perpetration-victimization risk equations in the PPDS and AAS (i.e., for both first occurrences and reoccurrences of workplace violence perpetration/victimization). As the AAS and PPDS surveys include detailed self-report assessments of experiences occurring during prior deployments (compared to the coarse deployment-related predictors in the HADS), analysis of these data will allow us to answer important questions about the bivariate and multivariate importance of deployment experiences in predicting workplace violence perpetration and victimization. Sixth, we will continue to work with the AAG to create a continuously-updating version of the HADS algorithms for real-time implementation of our risk equations. We will also continue to have discussions with the Army to determine the best methods of implementing our survey-based algorithms.

PUBLICATIONS, ABSTRACTS, AND PRESENTATIONS

We have prioritized the publication of several of the HADS data mining prediction equations. We currently have two manuscripts that are nearly ready for submission, one focusing on major physical violence perpetration (among men, women, and in the total sample), and a second focusing on major sexual violence victimization (among men, women, and in the total sample). We plan on submitting these manuscripts within the next few months. We otherwise have no publications, abstracts, or presentations to report.

INVENTIONS, PATENTS AND LICENSES

We have no inventions, patents, or licenses to report.

REPORTABLE OUTCOMES

- We have generated the rates per 1,000 PY for perpetration and victimization of major and minor physical and major sexual workplace violence in the 2004-2009 HADS (all Regular Army soldiers on active duty during this period) and in the 2011-2013 NSS (new soldiers).
- We have generated the rates per 1,000 PY of self-reported major physical violence, minor physical violence, and major sexual violence in the consolidated A-STARRS AAS and PPDS survey samples
- We have examined numerous bivariate associations of predictors (e.g., socio-demographics, career history variables, mental and physical health constructs, stress, prior perpetration/victimization, childhood adversities, traumatic events, social networks) with the administrative record defined perpetration/victimization outcomes in the HADS and NSS.
- We have examined socio-demographic and career history bivariate associations with self-reported violence in the AAS and PPDS.
- We used cutting-edge data mining methods to develop preliminary risk equations in the HADS and NSS. These predictive models can identify high risk soldiers that should be targeted with prevention and intervention programs.

- Final models to identify high risk soldiers are in progress, and we have started working with AAG to discuss continuously-updated versions our risk equations for real-time implementation in identifying soldiers who may benefit from workplace violence preventative interventions.

OTHER ACHIEVEMENTS

We have no other achievements to report.

REFERENCES

United States Department of Justice. Office of Justice Programs. Bureau of Justice Statistics. National Corrections Reporting Program, 2009. ICPSR30799-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2013-02-11. doi:10.3886/ICPSR30799.v2

Kessler, R.C., Warner, C.H., Ivany, C., Petukhova, M.V., Rose, R. Bromet, E.J., Brown III, M., Cai, T., Colpe, L.J., Cox, K.L., Fullerton, C.S., Gilman, S.E., Gruber, M.J., Heeringa, S.G., Lewandowski-Romps, L., Li, J., Millikan-Bell, A.M., Naifeh, J.A., Nock, M.K., Rosellini, A.J., Sampson, N.A., Schoenbaum, M., Stein, M.B., Wessely, S., Zaslavsky, A.M., Ursano, R.J. (In Press.) Predicting U.S. Army suicides after hospitalizations with psychiatric diagnoses in the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *JAMA Psychiatry*.

APPENDICES

- Appendix A: Main Tables. Main Tables 1-9 that correspond to the main results presented in our Overall Project Summary (e.g., standardized rates of workplace violence, AUCs, CRs; labeled Tables 1-9).
- Appendix B: Appendix Tables. Appendix Tables 1-15 that list out the predictors and their coefficients from the final models for each of the 10 HADS and 5 NSS workplace violence outcomes.
- Appendix C: Predicting Suicides After Psychiatric Hospitalization in US Army Soldiers. Example Army STARRS paper that used similar data mining methods to develop a risk equation for predicting post-inpatient hospitalization suicide (this paper is currently in press in JAMA Psychiatry).
- Appendix D: Quad Chart.

Appendix A: Main Tables

Main Table 1. Rates per 1,000 person-years of workplace (non-familial) violence perpetration and victimization in the Historical Administrative Data System (HADS) 2004-2009 Regular Army (n=975,051)

	Major Physical						Minor Physical						Major Sexual	Minor Sexual
	Murder/ Homicide/ Manslaughter	Kidnapping	Aggravated Arson	Aggravated Assault	Robbery	Any Major Physical	Simple Assault	Blackmail/ Extortion/ Intimidation	Rioting	Harassment	Other Physical Violence	Any Minor Physical	Rape/ Sodomy/ Sexual Assault	Sexual Verbal
Total														
Perpetrators														
Accused	0.2	0.2	0.0	1.6	0.2	2.1	9.3	1.2	0.0	0.6	0.2	10.6	2.8	0.3
Founded	0.2	0.2	0.0	1.6	0.1	2.0	9.1	1.1	0.0	0.6	0.2	10.4	2.3	0.3
Judicial Guilty	0.1	0.0	0.0	0.2	0.0	0.3	0.1	0.1	0.0	0.0	0.0	0.1	0.4	0.0
NJP guilty	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.0
AAT	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0
Any Guilty	0.1	0.0	0.0	0.2	0.0	0.3	0.1	0.1	0.0	0.0	0.1	0.3	0.7	0.0
Victims	0.2	0.1	0.0	1.0	0.1	1.4	4.5	0.8	0.0	0.3	0.2	5.5	1.8	0.1
Males														
Perpetrators														
Accused	0.2	0.2	0.0	1.7	0.2	2.2	9.8	1.3	0.0	0.7	0.2	11.2	3.3	0.4
Founded	0.2	0.2	0.0	1.7	0.2	2.2	9.5	1.2	0.0	0.6	0.2	11.0	2.6	0.3
Judicial Guilty	0.1	0.0	0.0	0.2	0.0	0.3	0.1	0.1	0.0	0.0	0.0	0.2	0.4	0.0
NJP guilty	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.0
AAT	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0
Any Guilty	0.1	0.0	0.0	0.3	0.0	0.4	0.2	0.1	0.0	0.0	0.1	0.3	0.8	0.1
Victims	0.2	0.0	0.0	1.1	0.1	1.4	4.0	0.7	0.0	0.3	0.1	4.8	0.3	0.1
Females														
Perpetrators														
Accused	0.1	0.0	0.0	0.9	0.0	1.0	6.3	0.5	0.0	0.3	0.0	6.9	0.3	0.1
Founded	0.1	0.0	0.0	0.9	0.0	1.0	6.2	0.5	0.0	0.3	0.0	6.8	0.3	0.1
Judicial Guilty	0.0	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
NJP guilty	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AAT	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Any Guilty	0.0	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
Victims	0.2	0.3	0.0	1.0	0.1	1.4	7.6	1.5	0.0	0.7	0.9	10.0	11.3	0.6

Abbreviations: NJP, nonjudicial punishment (article 15); AAT, administrative action taken.

Main Table 2. Distributions of number of workplace (non-familial) violence perpetrations and victimizations in the Historical Administrative Data System (HADS) 2004-2009 Regular Army (n=975,051)

	Number of founded perpetrations/victimizations of each crime type in the HADS 2004-2009						
	1	2	3	4	5	6+	Total
Total							
Perpetrator major physical violence							
Number of people	5174	575	82	36	9	5	5881
% of people	0.5%	0.1%	0.0%	0.0%	0.0%	0.0%	0.6%
Number of acts	5174	1150	246	144	45	41	6800
% of acts	76.1%	16.9%	3.6%	2.1%	0.7%	0.6%	100.0%
Perpetrator minor physical violence							
Number of people	23480	3828	813	221	63	36	28441
% of people	2.4%	0.4%	0.1%	0.0%	0.0%	0.0%	2.9%
Number of acts	23480	7656	2439	884	315	243	35017
% of acts	67.1%	21.9%	7.0%	2.5%	0.9%	0.7%	100.0%
Perpetrator major sexual violence							
Number of people	4917	837	222	74	25	13	6088
% of people	0.5%	0.09	0.023	0.008	0.003	0.0013	0.6253
Number of acts	4917	1674	666	296	125	96	7774
% of acts	63.2%	21.5%	8.6%	3.8%	1.6%	1.2%	100.0%
Victim major physical violence							
Number of people	4010	251	22	0	0	0	4283
% of people	0.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.4%
Number of acts	4010	502	66	0	0	0	4578
% of acts	87.6%	11.0%	1.4%	0.0%	0.0%	0.0%	100.0%
Victim minor physical violence							
Number of people	14644	1595	229	43	8	6	16525
% of people	1.5%	0.2%	0.0%	0.0%	0.0%	0.0%	1.7%
Number of acts	14644	3190	687	172	40	38	18771
% of acts	78.0%	17.0%	3.7%	0.9%	0.2%	0.2%	100.0%
Victim major sexual violence							
Number of people	4380	733	118	26	3	2	5262
% of people	0.4%	0.1%	0.0%	0.0%	0.0%	0.0%	0.5%
Number of acts	4380	1466	354	104	15	14	6333
% of acts	69.2%	23.1%	5.6%	1.6%	0.2%	0.2%	100.0%
Males							
Perpetrator major physical violence							
Number of people	4785	553	76	35	9	5	5463
% of people	0.5%	0.1%	0.0%	0.0%	0.0%	0.0%	0.6%
Number of acts	4785	1106	228	140	45	41	6345
% of acts	75.4%	17.4%	3.6%	2.2%	0.7%	0.6%	100.0%
Perpetrator minor physical violence							
Number of people	20972	3559	772	210	61	35	25609
% of people	2.2%	0.4%	0.1%	0.0%	0.0%	0.0%	2.6%
Number of acts	20972	7118	2316	840	305	236	31787
% of acts	66.0%	22.4%	7.3%	2.6%	1.0%	0.7%	100.0%

Perpetrator major sexual violence							
Number of people	4819	828	222	74	25	13	5981
% of people	0.5%	0.1%	0.0%	0.0%	0.0%	0.0%	0.6%
Number of acts	4819	1656	666	296	125	96	7658
% of acts	62.9%	21.6%	8.7%	3.9%	1.6%	1.3%	100.0%
Victim major physical violence							
Number of people	3431	199	17	0	0	0	3647
% of people	0.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.4%
Number of acts	3431	398	51	0	0	0	3880
% of acts	88.4%	10.3%	1.3%	0.0%	0.0%	0.0%	100.0%
Victim minor physical violence							
Number of people	11161	1075	130	21	5	3	12395
% of people	1.1%	0.1%	0.0%	0.0%	0.0%	0.0%	1.3%
Number of acts	11161	2150	390	84	25	19	13829
% of acts	80.7%	15.5%	2.8%	0.6%	0.2%	0.1%	100.0%
Victim major sexual violence							
Number of people	653	98	15	0	1	0	767
% of people	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%
Number of acts	653	196	45	0	5	0	899
% of acts	72.6%	21.8%	5.0%	0.0%	0.6%	0.0%	100.0%
Females							
Perpetrator major physical violence							
Number of people	389	22	6	1	0	0	418
% of people	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Number of acts	389	44	18	4	0	0	455
% of acts	85.5%	9.7%	4.0%	0.9%	0.0%	0.0%	100.0%
Perpetrator minor physical violence							
Number of people	2508	269	41	11	2	1	2832
% of people	0.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.3%
Number of acts	2508	538	123	44	10	7	3230
% of acts	77.6%	16.7%	3.8%	1.4%	0.3%	0.2%	100.0%
Perpetrator major sexual violence							
Number of people	98	9	0	0	0	0	107
% of people	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Number of acts	98	18	0	0	0	0	116
% of acts	84.5%	15.5%	0.0%	0.0%	0.0%	0.0%	100.0%
Victim major physical violence							
Number of people	579	52	5	0	0	0	636
% of people	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%
Number of acts	579	104	15	0	0	0	698
% of acts	83.0%	14.9%	2.1%	0.0%	0.0%	0.0%	100.0%
Victim minor physical violence							
Number of people	3483	520	99	22	3	3	4130
% of people	0.4%	0.1%	0.0%	0.0%	0.0%	0.0%	0.4%
Number of acts	3483	1040	297	88	15	19	4942

% of acts	70.5%	21.0%	6.0%	1.8%	0.3%	0.4%	100.0%
Victim major sexual violence							
Number of people	3727	635	103	26	2	2	4495
% of people	0.4%	0.1%	0.0%	0.0%	0.0%	0.0%	0.5%
Number of acts	3727	1270	309	104	10	14	5434
% of acts	68.6%	23.4%	5.7%	1.9%	0.2%	0.3%	100.0%

Main Table 3. Comparison of risk per 1,000 person-years of first perpetration and victimization of major physical, minor physical, and major sexual workplace (non-familial) violence in the 2004-2009 HADS[1] (n=975,051) and the 2011-2013 NSS (n=21,832)

	Male				Female				Total			
	HADS		NSS		HADS		NSS		HADS		NSS	
	Risk	SE	Risk	SE	Risk	SE	Risk	SE	Risk	SE	Risk	SE
Perpetration												
Major physical violence	3.2*	0.1	1.9	0.3	1.3	0.1	0.7	0.4	2.9*	0.1	1.8	0.2
Major sexual violence	3.1	0.1	3.1	0.3	0.5	0.1	0.5	0.3	2.7	0.1	2.7	0.3
Minor physical violence	13.8*	0.1	6.7	0.5	9.2	0.3	7.2	1.3	13.1*	0.1	6.8	0.4
Victimization												
Major physical violence	2.2*	0.0	0.8	0.2	2.3*	0.1	0.9	0.5	2.2*	0.0	0.8	0.2
Major sexual violence	0.6	0.0	1.1	0.3	22.5	0.4	25.1	2.6	3.9	0.1	4.1	0.4
Minor physical violence	6.7*	0.1	3.6	0.4	14.2*	0.3	10.2	1.5	7.8*	0.1	4.5	0.4

*Significant difference between HADS and NSS rates at the .05 level, two-sided test.
Abbreviations: HADS, Historic Administrative Data System; NSS, New Soldier Study; SE, standard error.
[1] In order to compare risk in the HADS and NSS, the HADS sample was restricted to soldiers (i.e., person-months) with less than 33 months in service (i.e., the maximum number of NSS follow-up months).

Main Table 4. Prevalence of self-reported verbal, physical, and sexual violence perpetration and victimization in the AAS (n=9,027) and PPDS (n=8,552)

	AAS						PPDS					
	Male		Female		Total		Male		Female		Total	
	%	n	%	n	%	n	%	n	%	n	%	n
I. Perpetrator of verbal violence												
A. Yell, insult, swear, or threaten someone (past 30 days)	16.1	1319	12.3	136	15.6	1455	12.9	1086	8.6	50	12.4	1136
B. Heated argument with someone (past 30 days)	8.3	656	8.5	93	8.3	749	6.0	519	5.7	33	6.0	552
C. Loud argument in public (past 30 days)	3.2	230	2.2	26	3.1	256	2.6	218	2.5	14	2.6	232
D. Any perpetrator of verbal violence (past 30 days)	17.6	1447	13.8	158	17.2	1605	14.2	1197	9.6	56	13.8	1253
II. Perpetrator of physical violence												
A. Physical confrontation during an argument (past 30 days)	3.1	279	1.9	23	3.0	302	3.2	257	2.8	16	3.1	273
B. Sometimes hit people so hard that they got bruises or had to see a doctor	2.6	217	0.9	14	2.4	231	6.6	539	2.8	16	6.2	555
C. Any perpetrator of physical violence	5.1	437	2.7	35	4.8	472	9.1	740	5.2	30	8.7	770
III. Victim of physical or sexual violence, or bullying												
A. Experienced physical assault (during deployment)	2.2	94	1.4	8	2.1	102	1.4	129	0.5	3	1.3	132
B. Experienced physical assault (past 12 months)	0.9	71	2.7	24	1.2	95	1.0	71	2.5	16	1.1	87
C. Victim of sexual assault (during deployment)	0.6	25	3.8	21	1.0	46	0.2	12	1.9	12	0.3	24
C. Sexual assault (during past 12 months)	0.1	12	3.1	21	0.5	33	0.1	6	1.6	9	0.2	15
D. Bullied by unit members (during deployment)	3.2	163	8.5	49	3.8	212	4.1	292	4.3	26	4.1	318
E. Any victim of physical or sexual violence, or bullying	5.0	317	11.4	95	5.7	412	6.2	463	9.6	58	6.5	521

Abbreviations: AAS, All Army Study; PPDS, Pre-post Deployment Study.

Main Table 5. Odds ratios of self-reported perpetration and victimization in the AAS (n=9,027) and PPDS (n=8,552)

	AAS Perpetration verbal violence		PPDS Perpetration verbal violence		AAS Perpetration physical violence		PPDS Perpetration physical violence		AAS Victim of any violence		PPDS Victim of any violence	
	OR	**(95% CI)**	**OR**	**(95% CI)**	**OR**	**(95% CI)**	**OR**	**(95% CI)**	**OR**	**(95% CI)**	**OR**	**(95% CI)**
Gender												
Male	--	--	--	--	--	--	--	--	--	--	--	--
Female	0.7*	(0.6-0.9)	0.6*	(0.5-0.8)	0.5*	(0.4-0.7)	0.6*	(0.4-0.8)	2.5*	(1.9-3.2)	1.6*	(1.1-2.3)
χ^2_1	9.8*		9.7*		13.9*		7.5*		43.8*		7.6*	
Age at interview												
Less than 20	--	--	--	--	--	--	--	--	--	--	--	--
21-24	0.7	(0.3-1.5)	0.8	(0.7-1.1)	0.8	(0.5-1.5)	0.8	(0.6-1.0)	2.5*	(1.2-5.1)	2.3*	(1.5-3.6)
25-29	0.7	(0.3-1.4)	0.7*	(0.5-0.8)	0.4*	(0.2-0.9)	0.7*	(0.5-0.9)	2.6*	(1.3-5.2)	3.5*	(2.1-6.0)
30-44	0.4*	(0.2-0.9)	0.5*	(0.3-0.6)	0.4*	(0.2-0.7)	0.5*	(0.3-0.7)	3.8*	(2.0-7.3)	3.5*	(2.1-5.9)
45+	0.2*	(0.1-0.5)	0.1*	(0.0-0.4)	0.1*	(0.0-0.2)	0.1*	(0.0-0.6)	5.3*	(2.4-11.7)	14.1*	(5.2-38.7)
χ^2_4	42.5*		60.6*		82.3*		18.6*		27.0*		39.7*	
Race/Ethnicity												
Non-Hispanic White	--	--	--	--	--	--	--	--	--	--	--	--
Non-Hispanic Black	0.9	(0.7-1.2)	0.7*	(0.6-1.0)	1.0	(0.8-1.3)	1.3	(1.0-1.7)	1.0	(0.7-1.2)	1.3	(0.6-3.2)
Hispanic	0.9	(0.6-1.5)	0.8	(0.7-1.1)	1.4	(0.6-3.3)	1.0	(0.7-1.4)	0.9	(0.5-1.6)	1.0	(0.7-1.3)
Other	0.7*	(0.5-1.0)	0.9	(0.7-1.2)	1.0	(0.6-2.0)	1.2	(0.9-1.5)	1.1	(0.7-1.6)	1.2	(0.9-1.6)
χ^2_3	8.6*		6.0		1.2		4.5		0.5		1.3	
Rank												
E1-E4	--	--	--	--	--	--	--	--	--	--	--	--
E5-E9	0.9	(0.7-1.1)	0.9	(0.7-1.1)	0.5*	(0.4-0.7)	0.7*	(0.5-0.9)	1.2	(0.8-1.7)	1.5*	(1.2-1.9)
Officer	0.2*	(0.1-0.3)	0.3*	(0.2-0.5)	0.2*	(0.1-0.4)	0.2*	(0.1-0.7)	1.4	(1.0-2.0)	1.5	(0.6-3.5)
χ^2_2	80.5*		28.2*		47.9*		15.5*		3.8		15.9*	
Age of Enlistment												
17-18	--	--	--	--	--	--	--	--	--	--	--	--
19-20	1.0	(0.8-1.2)	0.8*	(0.6-0.9)	1.3	(0.9-2.0)	1.1	(0.9-1.4)	1.0	(0.7-1.4)	1.2	(0.9-1.4)
21-23	0.6*	(0.4-0.8)	0.6*	(0.5-0.7)	0.8	(0.5-1.3)	0.8	(0.6-1.1)	0.8	(0.5-1.2)	0.9	(0.6-1.1)
24+	0.5*	(0.3-0.8)	0.4*	(0.4-0.5)	0.4*	(0.3-0.7)	0.8	(0.5-1.3)	1.1	(0.7-1.6)	1.5	(0.8-3.0)
χ^2_3	16.0*		65.3*		23.3*		5.9		2.3		5.3	
Months in active service												

Less than 20	--	--	--	--	--	--	--	--	--	--	--	--
21-45	1.0	(0.8-1.2)	0.8*	(0.6-1.0)	1.0	(0.6-1.5)	0.8	(0.6-1.1)	2.8*	(1.5-5.1)	2.2*	(1.1-4.2)
46-85	1.1	(0.9-1.3)	0.9	(0.7-1.2)	0.8	(0.5-1.3)	0.9	(0.6-1.2)	3.8*	(2.0-7.5)	4.3*	(2.3-8.2)
86+	0.8	(0.7-1.0)	0.7*	(0.6-1.0)	0.5*	(0.3-0.9)	0.6*	(0.4-0.8)	4.0*	(2.0-7.8)	4.4*	(2.1-9.2)
χ^2_3	10.8*		9.7*		11.0*		10.6*		17.0*		39.7*	
Months in active or non-active service												
Less than 20	--	--	--	--	--	--	--	--	--	--	--	--
21-45	0.9	(0.7-1.1)	0.9	(0.7-1.1)	1.0	(0.7-1.4)	0.8	(0.6-1.2)	2.4*	(1.4-4.3)	2.5*	(1.3-4.8)
46-85	0.9	(0.7-1.2)	1.0	(0.8-1.3)	0.8	(0.5-1.2)	0.8	(0.6-1.0)	3.1*	(1.6-6.0)	5.5*	(2.8-11.0)
86+	0.7*	(0.6-1.0)	0.7*	(0.6-0.9)	0.5*	(0.3-0.8)	0.6*	(0.4-0.9)	3.6*	(1.8-6.9)	5.3*	(2.7-10.6)
χ^2_3	6.2*		11.6*		14.2*		7.8		14.2*		40.6*	
Deployment												
Never Deployed	--	--	--	--	--	--	--	--	--	--	--	--
Currently Deployed	0.3*	(0.2-0.5)	--	--	0.3*	(0.2-0.4)	--	--	0.6	(0.4-1.0)	--	--
Previously Deployed	1.1	(0.9-1.4)	--	--	1.0	(0.6-1.4)	--	--	2.7*	(1.7-4.1)	--	--
χ^2_2	40.9*				60.3*				117.7*			

*Significant at the .05 level, two-sided test.
Abbreviations: AAS, All Army Study; PPDS, Pre-post Deployment Study, OR = odds ratio, CI = confidence interval.

Main Table 6. Classification accuracy, concentration of risk, and standardized risk for predicting the first occurrence of major physical workplace perpetration and victimization in the 2004-2009 HADS (n=975,051)

	Perpetration						Victimization					
	Male		Female		Total		Male		Female		Total	
	Stepwise Logistic	Elastic Net[1]	Stepwise Logistic	Elastic Net[1]	Stepwise Logistic	Elastic Net[1]	Stepwise Logistic	Elastic Net[1]	Stepwise Logistic	Elastic Net[1]	Stepwise Logistic	Elastic Net[1]
AUC	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
Concentration of risk[2]												
Top 1%	14.4	14.3	14.4	11.6	14.8	14.7	8.2	8.1	13.6	13.7	8.3	8.2
2-5%	22.7	22.8	20.7	23.0	22.6	22.5	17.6	17.7	15.2	14.9	16.4	16.3
5-10%	13.7	13.5	16.7	13.4	13.9	14.0	13.4	13.5	13.4	13.7	14.1	14.7
10-45%	35.6	35.8	38.9	40.4	35.9	35.9	42.1	41.9	39.6	39.4	43.0	42.4
45-100%	13.5	13.5	9.3	11.6	12.9	12.8	18.7	18.7	18.2	18.2	18.2	18.5
Standardized rate per 1,000 person-years[2]												
Top 1%	29.4	29.2	13.5	10.5	27.7	27.6	11.1	11.0	19.3	19.2	11.3	11.2
2-5%	11.5	11.6	4.7	5.0	10.6	10.6	6.0	6.0	5.3	5.2	5.6	5.6
5-10%	5.6	5.5	3.0	1.9	5.2	5.3	3.7	3.7	3.8	3.8	3.9	4.0
10-45%	2.1	2.1	1.1	1.2	1.9	1.9	1.6	1.6	1.6	1.6	1.7	1.7
45-100%	0.5	0.5	0.2	0.2	0.4	0.4	0.5	0.5	0.5	0.5	0.5	0.5

Abbreviations: HADS, Historical Administrative Data System; AUC, area under the receiver operating characteristic curve.
[1] All elastic net penalized regression models performed best using a mixing parameter penalty of alpha=0.5.
[2] Final models were used to generate predicted probabilities for each person-month in the sample. These predicted probabilities were then discretized into 5 categories (top 1% of predicted probabilities; 2-5%; 5-10%; 10-45%; and 45-100%) and concentration of risk (i.e., percent of all first occurrences of the outcome occurring among soldiers with a particular predicted probability) and standardized outcome rates (i.e., per 1,000 person-years) were calculated.

Main Table 7. Classification accuracy, concentration of risk, and standardized risk for predicting the first occurrence of major sexual workplace perpetration and victimization in the 2004-2009 HADS (n=975,051)

	Perpetration		Victimization					
	Male		Male		Female		Total	
	Stepwise Logistic	Elastic Net[1]	Stepwise Logistic	Elastic Net[1]	Stepwise Logistic	Elastic Net[1]	Stepwise Logistic	Elastic Net[1]
AUC	0.7	0.7	0.8	0.8	0.8	0.8	0.9	0.9
Concentration of risk[2]								
Top 1%	7.0	7.0	13.3	10.7	12.0	11.9	35.7	34.8
2-5%	14.9	14.9	19.8	20.9	22.0	22.0	37.2	36.9
5-10%	12.8	13.1	15.3	6.9	15.4	15.8	11.3	11.6
10-45%	43.6	42.6	38.7	45.7	40.8	40.5	12.9	13.5
45-100%	21.7	22.5	12.9	15.7	9.8	9.8	3.0	3.2
Standardized rate per 1,000 person-years[2]								
Top 1%	16.2	16.1	3.7	3.1	120.4	119.3	57.8	54.6
2-5%	8.6	8.6	1.4	1.1	55.1	55.2	15.1	15.1
5-10%	5.9	5.9	0.9	1.2	31.0	31.7	3.7	3.8
10-45%	2.8	2.8	0.3	0.4	11.7	11.6	0.6	0.6
45-100%	0.9	0.9	0.1	0.1	1.8	1.8	0.1	0.1

Abbreviations: HADS, Historical Administrative Data System; AUC, area under the receiver operating characteristic curve.
[1] All elastic net penalized regression models performed best using a mixing parameter penalty of alpha=0.5.
[2] Final models were used to generate predicted probabilities for each person-month in the sample. These predicted probabilities were then discretized into 5 categories (top 1% of predicted probabilities; 2-5%; 5-10%; 10-45%; and 45-100%) and concentration of risk (i.e., percent of all first occurrences of the outcome occurring among soldiers with a particular predicted probability) and standardized outcome rates (i.e., per 1,000 person-years) were calculated.

Main Table 8: Risk per 1,000 person-years of major physical, minor physical, and major sexual workplace violence perpetration and victimization by gender and number of occurrences in NSS 2011-2013 sample (n=21,832)

	Male[1]			Female[1]			Total[1]		
	Person-level N	Risk	SE	Person-level N	Risk	SE	Person-level N	Risk	SE
Perpetration major physical violence	53	1.9	0.3	3	0.7	0.4	56	1.8	0.2
Exactly 1	51	1.8	0.3	3	0.7	0.4	54	1.6	0.2
Exactly 2	2	0.1	0.1	0	0.0	0.0	2	0.1	0.1
Perpetration major sexual violence	87	3.1	0.3	3	0.5	0.3	90	2.7	0.3
Exactly 1	86	3.0	0.3	3	0.5	0.3	89	2.7	0.3
Exactly 2	1	0.0	0.0	0	0.0	0.0	1	0.0	0.0
Perpetration minor physical violence	194	6.7	0.5	30	7.2	1.3	224	6.8	0.4
Exactly 1	181	6.2	0.5	30	7.2	1.3	211	6.3	0.4
Exactly 2	13	0.5	0.1	0	0.0	0.0	13	0.4	0.1
Victimization major physical violence	25	0.8	0.2	4	0.9	0.5	29	0.8	0.2
Exactly 1	25	0.8	0.2	4	0.9	0.5	29	0.8	0.2
Victimization major sexual violence	34	1.1	0.3	118	25.1	2.6	152	4.1	0.4
Exactly 1	30	0.8	0.2	104	22.0	2.4	134	3.5	0.3
Exactly 2	4	0.3	0.2	14	3.1	0.9	18	0.6	0.2
Victimization minor physical violence	109	3.6	0.4	49	10.2	1.5	158	4.5	0.4
Exactly 1	102	3.5	0.4	44	8.8	1.4	146	4.1	0.4
Exactly 2	6	0.2	0.1	5	1.3	0.6	11	0.3	0.1
3 or More	1	0.0	0.0	0	0.0	0.0	1	0.0	0.0

Abbreviations: NSS, New Soldiers Study, SE, standard error.
[1] Males: person-level n=18,869, person-month n=366,511; females: person-level n=2,963, person-month n=54,195; total: person-level n=21,832, person-month n=420,706.

Main Table 9. Classification accuracy, concentration of risk, and standardized risk for predicting the first occurrence of major physical, minor physical, and major sexual workplace perpetration and victimization in the 2011-2013 NSS (n=21,832)

	Perpetration						Victimization			
	Major Physical Violence (Males only)		Major Sexual Violence (Males only)		Minor Physical Violence (Total sample)		Major Sexual Violence (Females only)		Minor Physical Violence (Total sample)	
	Stepwise Logistic	Elastic Net[1]	Stepwise Logistic	Elastic Net[1]	Stepwise Logistic	Elastic Net[1]	Stepwise Logistic	Elastic Net[1]	Stepwise Logistic	Elastic Net[1]
AUC	0.8	0.8	0.7	0.7	0.7	0.8	0.7	0.7	0.7	0.7
Concentration of risk[2]										
Top 1%	3.8	5.7	5.7	6.9	8.0	8.9	5.9	3.4	3.8	3.8
2-5%	24.5	24.5	14.9	19.5	12.5	11.2	13.6	16.9	14.6	14.6
5-10%	17.0	15.1	16.1	13.8	8.0	15.6	13.6	9.3	6.3	6.3
10-45%	32.1	34.0	28.7	34.5	46.4	42.0	34.7	40.7	41.8	41.8
45-100%	22.6	20.8	34.5	25.3	25.0	22.3	32.2	29.7	33.5	33.5
Standardized rate per 1,000 person-years[2]										
Top 1%	8.1	9.7	18.7	20.9	57.2	62.1	157.3	112.6	25.0	25.0
2-5%	8.5	8.3	10.4	13.8	20.9	18.0	104.5	135.1	20.9	20.8
5-10%	7.7	7.0	8.8	7.5	10.3	20.3	79.0	54.8	6.6	6.6
10-45%	1.6	1.6	2.4	2.7	8.3	7.6	29.1	31.9	5.4	5.4
45-100%	0.7	0.7	1.7	1.3	2.9	2.5	14.4	13.6	2.6	2.6

Appendix B: Appendix Tables

Appendix Table 1. Stepwise logistic regression and elastic net penalized regression odds ratios for the final model predicting first perpetration of major physical workplace violence, among males in the 2004-2009 HADS (using the best 25 predictors)

	Stepwise Logistic Odds Ratio	Elastic Net (alpha=0.5) Odds Ratio
Core Variables		
Year - 2009	0.8*	0.8
Demographics		
Age - 23+	0.6*	0.6
Race - Non-Hispanic Black	1.9*	1.9
Race - Non-Hispanic White	0.8*	0.8
Education - Graduated high school	0.6*	0.6
Education - Graduated college	0.4*	0.5
Army Career		
MOS Infantry	1.3*	1.3
Currently deployed	0.3*	0.3
Command - TRADOC	0.5*	0.6
Command - N/S America, Europe/Central/Africa, Pacific	1.3*	1.3
Rank - E5+ (including officer)	0.8*	0.8
Months in service - 121+	0.5*	0.5
Last demotion occurred within 12 months	1.5*	1.5
Crime and Drugs		
Perpetrator of any crime in past 12 months (Count of each type of perpetration)	1.4*	1.4
Perpetrator of minor violence in past 24 months (Count of days)	1.6*	1.6
Perpetrator of any crime in past 24 months (Yes/No)	2.0*	2.0
Medical History		
1+ suicide attempts between (not including) the day of the crime and 12 months before (Yes/No)	3.1*	3.1
Conduct/ODD - count of days with outpatient visits in the past 12 months	2.0*	2.0
Count of days with outpatient visits from first of the month to day before crime	0.6*	0.6
Count of days with outpatient visits in the past 3 months	1.3*	1.3
Any stress disorder - occurrence of days with outpatient visits in the past 12 months (Yes/No)	1.3*	1.3
Sedative-Hypnotic prescribed in the past 12 months	1.5*	1.4
Parent Unit		
Median amount of time in the Army of all NCOs (E5-9) in soldier’s unit(s)	1.0	1.0
Median amount of time in the Army of all officers (WO and CO) in soldier’s unit(s)	1.0*	1.0
Median amount of time ever deployed of all NCOs (E5-9) in soldier’s unit(s)	1.0*	1.0

*Significant at the .05 level, two-sided test. Significance could only be calculated for the stepwise logistic regression models (elastic net penalized regression models do not provide significance tests).
Abbreviations: HADS, Historical Administrative Data System; MOS, military occupational specialty; ODD, oppositional defiant disorder; NCO, non-commissioned officer; CO, commissioned officer; WO, warrant officer.

Appendix Table 2. Stepwise logistic regression and elastic net penalized regression odds ratios for the final model predicting first perpetration of major physical workplace violence, among females in the 2004-2009 HADS (using the best 20 predictors)

	Stepwise Logistic	Elastic Net (alpha=0.5)
	Odds Ratio	Odds Ratio
Demographics		
Race - Non-Hispanic Black	3.7*	1.4
Age of Enlistment - 22+	0.6*	0.8
Army Career		
Currently deployed	0.2*	0.3
Command - FORSCOM	2.3*	1.5
Command - N/S America, Europe/Central/Africa, Pacific	2.0*	1.5
Rank - E5+ (including officer)	0.4*	0.9
Rank - E7+ (including officer)	0.4*	0.6
Last demotion occurred within 12 months	2.1*	1.6
Crime and Drugs		
Perpetrator of verbal violence in past 12 months (Count of days)	59.5*	2.3
Victim of major violence in past 12 months (Yes/No)	2.9*	1.2
Victim of any crime in past 12 months (Yes/No)	1.9*	1.7
Perpetrator of any crime in past 24 months (Yes/No)	2.1*	1.4
Medical History		
Alcohol abuse/dependence - count of days with outpatient visits in the past 12 months	1.4*	1.2
Prior mental disorders - count of days with outpatient visits in the past 12 months	0.3*	0.7
Adjustment disorder - occurrence of days with outpatient visits in the past 12 months (Yes/No)	2.4*	1.5
Drug-induced mental illness - occurrence of days with outpatient visits in the past 12 months (Yes/No)	8.8*	3.4
Non-affective psychosis - occurrence of days with outpatient visits in the past 12 months (Yes/No)	14.0*	3.0
Occurrence of days with outpatient visits in the past 12 months (Yes/No)	1.6*	1.7
Stressors/adversities - count of days in the hospital in the past 12 months	1.3	0.8
Depressive psychosis - occurrence of days in the hospital in the past 12 months (Yes/No)	13.1*	3.0

*Significant at the .05 level, two-sided test. Significance could only be calculated for the stepwise logistic regression models (elastic net penalized regression models do not provide significance tests).
Abbreviations: HADS, Historical Administrative Data System.

Appendix Table 3. Stepwise logistic regression and elastic net penalized regression odds ratios for the final model predicting first perpetration of major physical workplace violence, in the total 2004-2009 HADS (using the best 25 predictors)

	Stepwise Logistic	Elastic Net (alpha=0.5)
	Odds Ratio	Odds Ratio
Demographics		
Age - 23+	0.6*	0.6
Gender - Male	2.5*	2.4
Race - Non-Hispanic Black	2.0*	2.0
Race - Non-Hispanic White	0.7*	0.8
Education - Graduated high school	0.6*	0.6
Education - Graduated college	0.5*	0.5
Army Career		
MOS Infantry	1.3*	1.3
Currently Deployed	0.3*	0.3
Command - TRADOC	0.5*	0.5
Command - AMC/Other/Unknown	0.5*	0.6
Rank - E5+ (including officer)	0.8*	0.8
Months in service - 121+	0.5*	0.6
Last demotion occurred within 12 months	1.5*	1.5
Crime and Drugs		
Perpetrator of any crime in past 12 months (Count of each type of perpetration)	1.4*	1.4
Perpetrator of minor violence in past 24 months (Count of days)	1.6*	1.6
Victim of minor violence in past 24 months (Count of days)	1.7*	1.7
Perpetrator of any crime in past 24 months (Yes/No)	2.0*	1.9
Medical History		
1+ suicide attempts between (not including) the day of the crime and 12 months before (Yes/No)	2.6*	2.6
Count of days with outpatient visits from first of the month to day before crime	0.6*	0.6
Count of days with outpatient visits in the past 3 months	1.3*	1.3
Adjustment disorder - occurrence of days with outpatient visits in the past 12 months (Yes/No)	1.5*	1.5
Occurrence of days with outpatient visits in the past 12 months (Yes/No)	1.3*	1.3
Flag of four or more narcotic or psychotropic medications in 3-month period in the past 3 months	1.4*	1.4
Parent Unit		
Median amount of time in the Army of all NCOs (E5-9) in soldier's unit(s)	1.0	1.0
Median amount of time in the Army of all officers (WO and CO) in soldier's unit(s)	1.0*	1.0

*Significant at the .05 level, two-sided test. Significance could only be calculated for the stepwise logistic regression models (elastic net penalized regression models do not provide significance tests).
Abbreviations: HADS, Historical Administrative Data System; NCO, non-commissioned officer; CO, commissioned officer; WO, warrant officer.

Appendix Table 4. Stepwise logistic regression and elastic net penalized regression odds ratios for the final model predicting first victimization of major physical workplace violence, among males in the 2004-2009 HADS (using the best 20 predictors)

	Stepwise Logistic	Elastic Net (alpha=0.5)
	Odds Ratio	Odds Ratio
Core Variables		
Year - 2004	1.5*	1.5
Demographics		
Age - 23+	0.7*	0.7
Currently Married	0.7*	0.7
Education - Graduated high school	0.7*	0.7
Army Career		
Currently Deployed	0.6*	0.6
Command - FORSCOM	1.6*	1.6
Command - N/S America, Europe/Central/Africa, Pacific	2.9*	2.8
Rank - E5+ (including officer)	0.7*	0.7
Rank - E7+ (including officer)	0.4*	0.4
AFQT score 0-49	1.3*	1.3
Crime and Drugs		
Perpetrator of any crime in past 12 months (Count of each type of perpetration)	1.2*	1.2
Victim of any crime in past 12 months (Yes/No)	1.6*	1.6
Perpetrator of minor violence in past 24 months (Yes/No)	1.6*	1.6
Perpetrator of any crime in past 24 months (Yes/No)	1.6*	1.6
Medical History		
Injury and poisoning - count of days with outpatient visits in the past 12 months	1.2*	1.2
Any substance - occurrence of days with outpatient visits in the past 12 months (Yes/No)	1.6*	1.6
Analgesic - Narcotic short-acting not Schedule II in the past 12 months (times prescribed)	1.2*	1.2
Antianxiety Agent - Antihistamine Type in the past 12 months (Prescribed yes/No)	1.9*	1.9
Parent Unit		
Median amount of time ever deployed of all NCOs (E5-9) in soldier's unit(s)	1.0*	1.0
On average, the percentage of combat deaths over the past year for all duty units	1.6*	1.6

*Significant at the .05 level, two-sided test. Significance could only be calculated for the stepwise logistic regression models (elastic net penalized regression models do not provide significance tests).
Abbreviations: HADS, Historical Administrative Data System; AFQT, Armed forces qualification test; NCO, non-commissioned officer.

Appendix Table 5. Stepwise logistic regression and elastic net penalized regression odds ratios for the final model predicting first victimization of major physical workplace violence, among females in the 2004-2009 HADS (using the best 20 predictors)

	Stepwise Logistic	Elastic Net (alpha=0.5)
	Odds Ratio	Odds Ratio
Army Career		
Currently Deployed	0.6*	0.6
Rank - E5+ (including officer)	0.6*	0.6
Rank - E7+ (including officer)	0.4*	0.4
AFQT score 0-49	1.4*	1.4
Last demotion occurred within 12 months	1.8*	1.8
Crime and Drugs		
Perpetrator of any crime in past 12 months (Count of each type of perpetration)	1.6*	1.6
Victim of major sexual violence in past 12 months (Count of days)	2.2*	2.2
Perpetrator of non-violent drunkenness/vagrancy/disorderly crime in past 24 months (Count of days)	3.3*	3.2
Victim of any crime in past 12 months (Yes/No)	2.5*	2.6
Number of months since you had a positive drug test (Within 3 months)	3.0*	2.9
Medical History		
Marital problems - count of days with outpatient visits in the past 12 months	1.3*	1.3
Count of days with outpatient visits from first of the month to day before crime	0.3*	0.4
Count of days with outpatient visits in the past 3 months	1.4*	1.3
Drug dependence - occurrence of days with outpatient visits in the past 12 months (Yes/No)	2.8*	2.8
Anxiety - occurrence of days with outpatient visits in the past 12 months (Yes/No)	1.7*	1.7
Occurrence of days in the hospital from first of the month to day before crime (Yes/No)	6.8*	5.6
Analgesic - Non-Narcotic - Analgesics prescribed in the past 12 months	1.2*	1.2
Parent Unit		
Number of different duty units over past 12 months	1.3*	1.3
Median amount of time ever deployed of all NCOs (E5-9) in soldier's unit(s)	1.0*	1.0
On average, the percentage of combat deaths over the past year for all duty units	4.0*	3.7

*Significant at the .05 level, two-sided test. Significance could only be calculated for the stepwise logistic regression models (elastic net penalized regression models do not provide significance tests).
Abbreviations: HADS, Historical Administrative Data System; AFQT, Armed forces qualification test; NCO, non-commissioned officer.

Appendix Table 6. Stepwise logistic regression and elastic net penalized regression odds ratios for the final model predicting first victimization of major physical violence, in the total 2004-2009 HADS (using the best 20 predictors)

	Stepwise Logistic	Elastic Net (alpha=0.5)
	Odds Ratio	Odds Ratio
Demographics		
Age - 23+ yrs	0.8*	0.8
Currently Married	0.8*	0.8
Education - Graduated high school or higher	0.7*	0.7
Army Career		
Currently Deployed	0.6*	0.6
Command - FORSCOM	1.5*	1.4
Command - N/S America, Europe/Central/Africa, Pacific	2.4*	2.3
Rank - E5+ (including officer)	0.7*	0.7
Rank - E7+ (including officer)	0.4*	0.4
AFQT score 0-49	1.3*	1.3
Crime and Drugs		
Perpetrator of any crime in past 12 months (Count of each type of perpetration)	1.2*	1.2
Victim of non-violent crime in past 12 months (Yes/No)	0.5*	0.5
Victim of any crime in past 12 months (Yes/No)	3.0*	2.9
Perpetrator of any crime in past 24 months (Yes/No)	1.6*	1.6
Medical History		
Injury and poisoning - count of days with outpatient visits in the past 12 months	1.2*	1.2
Any substance disorder - occurrence of days with outpatient visits in the past 12 months (Yes/No)	1.6*	1.6
Analgesic - Narcotic short-acting not Schedule II prescribed in the past 12 months	1.2*	1.2
Parent Unit		
Number of different duty units over past 12 months	1.1*	1.1
Median amount of time in the Army of all officers (WO and CO) in soldier’s unit(s)	1.0*	1.0
Median number of months all lower enlisted soldiers (E1-E4) have been on duty in unit	1.0*	1.0
On average, the percentage of combat deaths over the past year for all duty units	1.6*	1.6

*Significant at the .05 level, two-sided test. Significance could only be calculated for the stepwise logistic regression models (elastic net penalized regression models do not provide significance tests).
Abbreviations: HADS, Historical Administrative Data System; AFQT, Armed forces qualification test; WO, warrant officer; CO, commissioned officer.

Appendix Table 7. Stepwise logistic regression and elastic net penalized regression odds ratios for the final model predicting first perpetration of sexual workplace violence, among males in the 2004-2009 HADS (using the best 20 predictors)

	Stepwise Logistic	Elastic Net (alpha=0.5)
	Odds Ratio	Odds Ratio
Core Variables		
Year - 2004	0.7*	0.8
Demographics		
Age - 23+	0.8*	0.8
Race - Non-Hispanic Black	1.8*	1.8
Race - Hispanic	1.5*	1.4
Education - Graduated high school or higher	0.8*	0.8
Army Career		
Currently Deployed	0.4*	0.4
Command - N/S America, Europe/Central/Africa, Pacific	1.3*	1.3
Command - Special Ops	0.5*	0.6
Rank - E7+ (including officer)	0.5*	0.5
Rank - Officer	0.5*	0.6
AFQT score 0-49	1.2*	1.2
Last demotion occurred within 12 months	1.5*	1.5
Crime and Drugs		
Perpetrator of any crime in past 12 months (Count of each type of perpetration)	1.2*	1.2
Perpetrator of non-workplace (familial) sexual violence in past 24 months (Count of days)	11.8*	11.5
Perpetrator of any crime in past 24 months (Yes/No)	2.0*	2.0
Number of months since you had a positive test (Within 12 months)	0.6*	0.6
Medical History		
Count of days with outpatient visits in the past 3 months	1.3*	1.3
Occurrence of days with outpatient visits from first of the month to day before crime (Yes/No)	0.4*	0.4
Occurrence of days with outpatient visits in the past 12 months (Yes/No)	1.3*	1.3
Parent Unit		
Number of different duty units over past 12 months	1.2*	1.2

*Significant at the .05 level, two-sided test. Significance could only be calculated for the stepwise logistic regression models (elastic net penalized regression models do not provide significance tests).
Abbreviations: HADS, Historical Administrative Data System; AFQT, Armed forces qualification test.

Appendix Table 8. Stepwise logistic regression and elastic net penalized regression odds ratios for the final model predicting first victimization of sexual workplace violence, among males in the 2004-2009 HADS (using the best 20 predictors)

	Stepwise Logistic Odds Ratio	Elastic Net (alpha=0.5) Odds Ratio
Core Variables		
Year - 2007	1.5*	1.0
Year - 2004	0.4*	1.0
Demographics		
Age - 23+ yrs	0.6*	0.7
Army Career		
Duty MOS - Basic Training	1.6*	1.0
Currently deployed	0.7*	1.0
Command - N/S America, Europe/Central/Africa, Pacific	1.8*	1.0
Rank - E5+ (including officer)	0.3*	0.6
Months in service - 25+	0.6*	0.6
Crime and Drugs		
Perpetrator of sexual violence in past 12 months (Count of days)	7.8*	5.4
Victim of any crime in past 12 months (Count of each type of victimization)	1.8*	1.0
Number of months since you had a positive test (Within 12 months)	3.4*	1.9
Medical History		
Adjustment disorder - count of days with outpatient visits in the past 12 months	1.4*	1.1
Injury and poisoning - count of days with outpatient visits in the past 12 months	1.3*	1.0
Count of days with outpatient visits from first of the month to day before crime	0.4*	1.0
Count of days with outpatient visits in the past 12 months	1.2*	1.1
Occurrence of days with outpatient visits in the past 3 months	2.1*	1.2
Occurrence of days with hospitalizations from first of the month to day before crime	5.6*	1.0
Antidepressants prescribed (SSRI) in the past 12 months	2.0*	1.2
Parent Unit		
Number of different duty units over past 12 months	1.3*	1.1
Median number of months all officers (CO and WO) have been on duty in this unit	1.0*	1.0

*Significant at the .05 level, two-sided test. Significance could only be calculated for the stepwise logistic regression models (elastic net penalized regression models do not provide significance tests).
Abbreviations: HADS, Historical Administrative Data System; MOS, military occupational specialty; CO, commissioned officer; WO, warrant officer.

Appendix Table 9. Stepwise logistic regression and elastic net penalized regression odds ratios for the final model predicting first victimization of sexual workplace violence, among females in the 2004-2009 HADS (using the best 20 predictors)

	Stepwise Logistic	Elastic Net (alpha=0.5)
	Odds Ratio	Odds Ratio
Core Variables		
Year - 2004	0.5*	0.6
Demographics		
Age - 20+	0.8*	0.8
Age - 23+	0.6*	0.6
Currently Married	0.7*	0.8
Race - Non-Hispanic White	1.9*	1.9
Education - Graduated high school or higher	0.7*	0.7
Army Career		
Command - FORSCOM	1.4*	1.4
Command - N/S America, Europe/Central/Africa, Pacific	2.0*	2.0
Rank - E5+ (including officer)	0.5*	0.5
Rank - E7+ (including officer)	0.3*	0.4
Months in service - 13+	0.6*	0.6
Months in service - 25+	0.6*	0.6
Crime and Drugs		
Perpetrator of sexual violence in past 12 months (Yes/No)	7.9*	7.6
Victim of non-workplace (familial) sexual violence in past 12 months (Yes/No)	8.8*	8.7
Victim of any crime in past 24 months (Yes/No)	1.6*	1.6
Medical History		
Injury and poisoning - count of days with outpatient visits in the past 12 months	1.2*	1.2
Count of days with outpatient visits from first of the month to day before crime	0.6*	0.7
Count of days with outpatient visits in the past 12 months	1.3*	1.3
PTSD - occurrence of days in the hospital in the past 12 months (Yes/No)	4.4*	4.2
Parent Unit		
Number of different duty units over past 12 months (Yes/No)	1.2*	1.2

*Significant at the .05 level, two-sided test. Significance could only be calculated for the stepwise logistic regression models (elastic net penalized regression models do not provide significance tests).
Abbreviations: HADS, Historical Administrative Data System.

Appendix Table 10. Stepwise logistic regression and elastic net penalized regression odds ratios for the final model predicting first victimization of sexual workplace violence, in the total 2004-2009 HADS (using the best 20 predictors)

	Stepwise Logistic	Elastic Net (alpha=0.5)
	Odds Ratio	Odds Ratio
Core Variables		
Year - 2004	0.5*	0.9
Demographics		
Age - 23+	0.6*	0.6
Currently Married	0.8*	1.0
Gender - Male	0.0*	0.1
Race - Non-Hispanic White	1.8*	1.2
Education - Graduated high school or higher	0.7*	0.8
Army Career		
Command - N/S America, Europe/Central/Africa, Pacific	1.7*	1.1
Rank - E5+ (including officer)	0.5*	0.6
Rank - E7+ (including officer)	0.3*	0.9
Months in service - 13+	0.7*	0.7
Months in service - 25+	0.6*	0.6
Crime and Drugs		
Perpetrator of sexual violence in past 12 months (Count of days)	5.2*	1.9
Victim of non-workplace (familial) sexual violence in past 12 months (Yes/No)	8.5*	13.6
Victim of any crime in past 24 months (Yes/No)	1.6*	1.1
Number of months since you had a positive test (Within 3 months)	2.6*	1.1
Medical History		
Injury and poisoning - count of days with outpatient visits in the past 12 months	1.2*	1.2
Count of days with outpatient visits from first of the month to day before crime	0.5*	1.0
Count of days with outpatient visits in the past 3 months	1.3*	1.0
Count of days with outpatient visits in the past 12 months	1.2*	1.2
Parent Unit		
Number of different duty units over past 12 months	1.3*	1.2

*Significant at the .05 level, two-sided test. Significance could only be calculated for the stepwise logistic regression models (elastic net penalized regression models do not provide significance tests).
Abbreviations: HADS, Historical Administrative Data System.
.

Appendix Table 11. Stepwise logistic regression and elastic net penalized regression odds ratios for the final model predicting first perpetration of major physical workplace violence, among males in the 2011-2013 NSS (using the best 5 predictors)

	Stepwise Logistic Odds Ratio	Elastic Net (alpha=0.5) Odds Ratio
Demographics		
Race/Ethnicity - Non-Hispanic Black	5.0*	4.3
Army career		
Months in the Army - Less than 12	0.3*	0.3
Medical history		
Any medical failure at accession	3.3*	2.7
Personality traits		
Introversion	0.8*	0.8
Stress		
Experienced severe or very severe stress in past 12 months	2.7*	2.5

*Significant at the .05 level, two-sided test. Significance could only be calculated for the stepwise logistic regression models (elastic net penalized regression models do not provide significance tests).
Abbreviations: NSS, New Soldiers Study.

Appendix Table 12. Stepwise logistic regression and elastic net penalized regression odds ratios for the final model predicting first perpetration of sexual workplace violence, among males in the 2011-2013 NSS (using the best 8 predictors)

	Stepwise Logistic	Elastic Net (alpha=0.5)
	Odds Ratio	Odds Ratio
Demographics		
Race/Ethnicity - Non-Hispanic Black	3.8*	3.2
Race/Ethnicity - Hispanic	2.8*	2.4
Army Career		
Months in the Army - 3 or less	0.3*	0.4
Months in the Army - 8 or less	1.7*	1.3
Personality		
Introversion	0.8*	0.9
Mental disorders		
Lifetime ODD (Yes/No)	2.0*	1.5
Childhood adversity		
Physically abused at home	1.5*	1.6
Social networks		
Count of sexual partners in the past 12 months	1.1*	1.1

*Significant at the .05 level, two-sided test. Significance could only be calculated for the stepwise logistic regression models (elastic net penalized regression models do not provide significance tests).
Abbreviations: NSS, New Soldiers Study.

Appendix Table 13. Stepwise logistic regression and elastic net penalized regression odds ratios for the final model predicting first perpetration of minor physical workplace violence, in the total 2011-2013 NSS (using the best 13 predictors)

	Stepwise Logistic	Elastic Net (alpha=0.5)
	Odds Ratio	Odds Ratio
Demographics		
Race/Ethnicity - Non-Hispanic Black	2.0*	2.0
Education - Some College/College Graduate	0.3*	0.4
Currently/Previously Married	1.7*	1.6
Army career		
Months in the Army - 4 or less	0.1*	0.1
AFQT Score (percentile) - 43+	0.7*	0.7
Personality		
Introversion	0.9*	1.0
Mental disorders		
Number of years with lifetime insomnia	1.1*	1.1
GAD symptoms - 1+	0.5*	0.5
Number of years with problem behaviors - 1+	1.6*	1.8
Number of years with anger attacks - 4+	1.7*	1.7
Lifetime Substance Disorders (Yes/No)	1.7*	1.9
Stress		
Experienced severe or very severe stress in past 12 months regarding health of loved one	1.5*	1.4
Childhood adversity		
Sent to juvenile detention center	4.1*	4.0

*Significant at the .05 level, two-sided test. Significance could only be calculated for the stepwise logistic regression models (elastic net penalized regression models do not provide significance tests).
Abbreviations: NSS, New Soldiers Study.

Appendix Table 14. Stepwise logistic regression and elastic net penalized regression odds ratios for the final model predicting first victimization of sexual workplace violence, among females in the 2011-2013 NSS (using the best 5 predictors)

	Stepwise Logistic	Elastic Net (alpha=0.5)
	Odds Ratio	Odds Ratio
Demographics		
Race/Ethnicity - Non-Hispanic Black	0.4*	0.3
Army career		
Months in the Army - 4 or less	0.4*	0.5
Personality		
Impulsivity-sensation seeking	1.2*	1.1
Pre-enlistment trauma		
Sexual Assault - 1 or more	2.0*	2.7
Childhood adversity		
Count of types of childhood adversities - 10+	2.3*	2.8

*Significant at the .05 level, two-sided test. Significance could only be calculated for the stepwise logistic regression models (elastic net penalized regression models do not provide significance tests).
Abbreviations: NSS, New Soldiers Study.

Appendix Table 15. Stepwise logistic regression and elastic net penalized regression odds ratios for the final model predicting first victimization of minor physical workplace violence in the total 2011-2013 NSS (using the best 10 predictors)

	Stepwise Logistic	Elastic Net (alpha=0.5)
	Odds Ratio	Odds Ratio
Core variables		
Quarter of NSS interview - 8	2.2*	2.3
Demographics		
Gender - Male	0.4*	0.3
Born in the US	3.8*	2.7
Army career		
Months in the Army - 6 or less	0.2*	0.3
Basic Combat Training Site - Fort Leonard Wood	1.5*	1.6
Personality		
Second order personality scale - Thoughtfulness	0.7*	0.8
Mental Disorders		
Number of years with lifetime insomnia - 2+	1.8*	1.3
Stress		
Experienced severe or very severe stress in past 12 months	1.9*	1.9
Trauma		
Count of types of traumas - 5+	1.5*	1.3
Childhood adversity		
Child Adversity - In foster home	2.1*	1.6

*Significant at the .05 level, two-sided test. Significance could only be calculated for the stepwise logistic regression models (elastic net penalized regression models do not provide significance tests).
Abbreviations: NSS, New Soldiers Study.

Appendix C: Predicting Suicides After Psychiatric Hospitalization in US Army Soldiers

Original Investigation

Predicting Suicides After Psychiatric Hospitalization in US Army Soldiers

The Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS)

Ronald C. Kessler, PhD; Christopher H. Warner, MD; Christopher Ivany, MD; Maria V. Petukhova, PhD; Sherri Rose, PhD; Evelyn J. Bromet, PhD; Millard Brown III, MD, MB; Tianxi Cai, ScD; Lisa J. Colpe, PhD, MPH; Kenneth L. Cox, MD, MPH; Carol S. Fullerton, PhD; Stephen E. Gilman, ScD; Michael J. Gruber, MS; Steven G. Heeringa, PhD; Lisa Lewandowski-Romps, PhD; Junlong Li, PhD; Amy M. Millikan-Bell, MD, MPH; James A. Naifeh, PhD; Matthew K. Nock, PhD; Anthony J. Rosellini, PhD; Nancy A. Sampson, BA; Michael Schoenbaum, PhD; Murray B. Stein, MD, MPH; Simon Wessely, PhD; Alan M. Zaslavsky, PhD; Robert J. Ursano, MD; for the Army STARRS Collaborators

IMPORTANCE The US Army experienced a sharp increase in soldier suicides beginning in 2004. Administrative data reveal that among those at highest risk are soldiers in the 12 months after inpatient treatment of a psychiatric disorder.

OBJECTIVE To develop an actuarial risk algorithm predicting suicide in the 12 months after US Army soldier inpatient treatment of a psychiatric disorder to target expanded posthospitalization care.

DESIGN, SETTING, AND PARTICIPANTS There were 53 769 hospitalizations of active duty soldiers from January 1, 2004, through December 31, 2009, with *International Classification of Diseases, Ninth Revision, Clinical Modification* psychiatric admission diagnoses. Administrative data available before hospital discharge abstracted from a wide range of data systems (sociodemographic, US Army career, criminal justice, and medical or pharmacy) were used to predict suicides in the subsequent 12 months using machine learning methods (regression trees and penalized regressions) designed to evaluate cross-validated linear, nonlinear, and interactive predictive associations.

MAIN OUTCOMES AND MEASURES Suicides of soldiers hospitalized with psychiatric disorders in the 12 months after hospital discharge.

RESULTS Sixty-eight soldiers died by suicide within 12 months of hospital discharge (12.0% of all US Army suicides), equivalent to 263.9 suicides per 100 000 person-years compared with 18.5 suicides per 100 000 person-years in the total US Army. The strongest predictors included sociodemographics (male sex [odds ratio (OR), 7.9; 95% CI, 1.9-32.6] and late age of enlistment [OR, 1.9; 95% CI, 1.0-3.5]), criminal offenses (verbal violence [OR, 2.2; 95% CI, 1.2-4.0] and weapons possession [OR, 5.6; 95% CI, 1.7-18.3]), prior suicidality [OR, 2.9; 95% CI, 1.7-4.9], aspects of prior psychiatric inpatient and outpatient treatment (eg, number of antidepressant prescriptions filled in the past 12 months [OR, 1.3; 95% CI, 1.1-1.7]), and disorders diagnosed during the focal hospitalizations (eg, nonaffective psychosis [OR, 2.9; 95% CI, 1.2-7.0]). A total of 52.9% of posthospitalization suicides occurred after the 5% of hospitalizations with highest predicted suicide risk (3824.1 suicides per 100 000 person-years). These highest-risk hospitalizations also accounted for significantly elevated proportions of several other adverse posthospitalization outcomes (unintentional injury deaths, suicide attempts, and subsequent hospitalizations).

CONCLUSIONS AND RELEVANCE The high concentration of risk of suicide and other adverse outcomes might justify targeting expanded posthospitalization interventions to soldiers classified as having highest posthospitalization suicide risk, although final determination requires careful consideration of intervention costs, comparative effectiveness, and possible adverse effects.

JAMA Psychiatry. doi:10.1001/jamapsychiatry.2014.1754
Published online November 12, 2014.

Author Affiliations: Author affiliations are listed at the end of this article.

Group Information: The Army STARRS collaborators are listed at the end of the article.

Corresponding Author: Ronald C. Kessler, PhD, Department of Health Care Policy, Harvard Medical School, 180 Longwood Ave, Ste 215, Boston, MA 02115-5899 (kessler@hcp.med.harvard.edu).

The US Army suicide rate, although historically below the civilian rate, has increased since 2004[1] to exceed the civilian rate.[2] Despite numerous efforts to address this problem, including universal interventions (eg, Ask/Care/Escort prevention education and depression, posttraumatic stress disorder, and suicide screening in all primary care encounters) and high-risk interventions (eg, postdeployment screening),[3] the US Army suicide rate has continued to increase. One potentially important group for targeted interventions is soldiers recently discharged from inpatient psychiatric treatment. Such patients have long been known to have a high risk of suicide.[4] US military administrative data document an 8-fold elevated suicide risk in the 3 months after psychiatric hospitalization and a 5-fold elevated risk for the remainder of the 12 months after hospitalization.[5] A report[6] on the similar patterns among civilians called for expansion of posthospitalization suicide preventive interventions, noting that such interventions in the United Kingdom (eg, required outpatient visits within 1 week of hospital discharge, assertive outreach for missed outpatient appointments, 24-hour community crisis teams, and intensive community support for patients difficult to engage in traditional services) were associated with significant before-after reductions in posthospitalization suicides.[7]

Suicide is a rare outcome even among recently discharged psychiatric inpatients[8]; therefore, the benefits of providing intensive posthospitalization suicide prevention interventions to all recently discharged inpatients are low. A more rational allocation of treatment resources would be to combine relatively inexpensive universal interventions[9] with more intensively targeted high-risk interventions.[4] However, this tiered approach would require developing a reliable risk stratification scheme. The US Department of Veterans Affairs (VA) and the US Department of Defense (DoD) called for this kind of differentiation in their Clinical Practice Guideline (CPG) entitled *Assessment and Management of Patients at Risk for Suicide*.[10] However, the CPG provided little concrete guidance on how these assessments should be implemented. Research has consistently revealed that health care professionals are not accurate in making such assessments.[11-14]

One potentially promising approach to assessing posthospitalization suicide risk would be to use administrative data available during hospitalization to generate an actuarial posthospitalization suicide risk algorithm. Previous research has revealed that actuarial suicide prediction is much more accurate than prediction based on clinical judgment.[11-14] An increasing number of computerized risk algorithms are being used as clinical decision support tools in other areas of medicine and have been found to improve clinical processes.[15,16] Skepticism exists about developing such an algorithm for posthospitalization suicide interventions based on the relatively weak associations found in previous research[17] on in-hospital predictors and subsequent suicides. However, a stronger risk algorithm might be developed in the US Army because of the availability of integrated administrative data for all US Army personnel. Absence of such data in the general population is widely recognized as an impediment to big data health care solutions.[18] A number of empirical studies[19-23] have documented strong predictive associations between integrated US Army and DoD administrative data and subsequent US Army suicides, although none attempted to develop a risk algorithm for posthospitalization suicides. The objective of this study was to develop such an algorithm using administrative data from the Historical Administrative Data System (HADS) of the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS).[24]

Methods

Sample

Creation and analysis of the consolidated and deidentified data system were approved by the Human Subjects Committees of the Uniformed Services University of the Health Sciences for the Henry M. Jackson Foundation (the primary grantee), the University of Michigan Institute for Social Research (site of the Army STARRS Data Enclave), and Harvard Medical School (site of data analysis). Obtaining informed consent from individual soldiers, most of whom were no longer in service at the time the HADS was constructed, was not required because the data were deidentified.

There were 53 769 regular US Army hospitalizations from January 1, 2004, through December 31, 2009, with any *International Classification of Diseases, Ninth Revision, Clinical Modification* (*ICD-9-CM*) psychiatric admission diagnosis exclusive of tobacco use disorders (eTable 1 at http://www.armystarrs.org/publications). These hospitalizations involved 40 820 soldiers (30 763 with 1 hospitalization, 6929 with 2, and 3128 with >2), representing 0.9% of all regular US Army soldiers in any 12-month period. We excluded the 13 936 additional hospitalizations in which nicotine dependence was the only psychiatric diagnosis because these were invariably for physical disorders and nicotine dependence was noted based on withdrawal during hospitalization. There was no elevated posthospitalization suicide risk among these soldiers. We also excluded the 406 additional hospitalizations that occurred through emergency departments because of a suicide attempt without an accompanying *ICD-9-CM* psychiatric diagnosis. Four of these 406 soldiers died in the hospital, whereas none of the others died by suicide in the next 12 months. On the basis of evidence from another study[25] indicating that predictors of posthospitalization suicide vary with time since discharge and elevated risk persists 12 months after discharge, a discrete-time person-month survival file was created to examine suicides in the 12 months after hospital discharge, censoring all person-months at the beginning of new hospitalizations or terminations of active duty and allowing interactions between substantive predictors and time since hospital discharge. All person-months with suicide were coded 1 on the outcome, and all others were coded 0. This file contained 334 936 person-months for a mean of 6.2 months (334 936 per 53 760 months) after hospital discharge. This low mean reflects high rates of termination of service and subsequent hospitalization within 12 months of each hospitalization.

jamanetwork/2014/psy/11_12_2014/yoi140086pap PAGE: left 2 SESS: 40 OUTPUT: Oct 21 11:51 2014

Measures

The HADS includes data from 38 US Army and DoD administrative data systems[26] (eTable 2 at http://www.armystarrs.org/publications). In a comprehensive review of published studies of predictors of civilian posthospitalization suicides, Troister et al[27] found 5 replicated classes of predictors: (1) sociodemographics (the most consistent being male sex and recent job loss), (2) history of prior suicidal behaviors, (3) quality of care (eg, low continuity of care), (4) time since hospital discharge (inversely related to suicide risk), and (5) other psychopathological risk factors (the most consistent being nonaffective psychosis, mood disorders, and multiple comorbid psychiatric disorders). Other studies[17,28,29] found similar predictors. We extracted HADS variables operationalizing these predictors and added US Army career variables found to predict military suicides,[19-22] unit variables, criminal justice variables (violent crime victimization or perpetration), and measures of registered weapons. All predictors other than those that involved the hospitalization were defined as of the month *before* hospitalization, whereas predicted suicides were in the 12 months *after* hospital discharge.

We cast a wide net in extracting HADS measures of the predictor constructs. For example, we distinguished 23 categories of psychiatric diagnoses defined largely by aggregated *ICD-9-CM* codes (eg, attention-deficit/hyperactivity learning disorders [*ICD-9-CM* codes 314.0-315.9]), 8 additional categories of behavioral stressors (eg, marital problems, other stressors or adversities, suicidal ideation, and self-damaging behavior), and summary measures of any prior admission diagnoses, admission count variables, and parallel outpatient variables (eTable 1 at http://www.armystarrs.org/publications). We also included National Drug Code psychotropic medication codes collapsed into 15 categories (eg, antianxiety, antidepressant, and antipsychotic) and 25 subcategories (eg, selective serotonin reuptake inhibitor, s*erotonin-norepinephrine reuptake inhibitor*, and tricyclic antidepressant) based on the First Databank Enhanced Therapeutic Classification System (http://www.fdbhealth.com) (eTable 3 at http://www.armystarrs.org/publications). A total of 421 individual variables were constructed (eTable 4 at http://www.armystarrs.org/publications).

Because the HADS data systems were not developed for research, more data were missing and inconsistent in some (eg, sociodemographic) component data sets than in research data sets. However, because the HADS data sets are updated monthly, missing values typically appeared in earlier and/or later months, allowing nearest neighbor imputations. Remaining missing values were resolved using randomly selected multiple imputations.[30] Inconsistencies were reconciled using rational imputations (eg, a soldier classified female one month but male other months was recoded male).

Statistical Analysis

Discrete-time (person-month) survival analysis[31] was used to predict suicides in the 12 months after hospitalization in 3 steps. First, functional forms of bivariate associations were examined and predictors transformed (usually sets of nested dichotomies but some collapsed-truncated continuous variables) to explore nonlinear multivariate associations. Second, all predictors were discretized and analyzed with 100 regression trees in distinct bootstrap pseudo-samples using the R package rpart program[32] to prevent overfitting[33] and allow detecting interactions among predictors.[25,28] Third, predictors having significant bivariate associations and interactions emerging in 10% or more of regression trees were included as predictors in multivariate survival models.

A central challenge in the third step was multicollinearity among the 421 predictors. The classic way to address this problem is with stepwise analysis,[34] but this approach overfits.[35] Machine learning methods reduce overfitting.[36,37] The machine learning method we used was the elastic net,[38] a penalized regression method that provides stable and sparse estimates of model parameters by explicitly penalizing overfitting with a composite penalty $\lambda\{\text{MPP} \times P_{lasso} + (1 - \text{MPP}) \times P_{ridge}\}$, where MPP is a mixing parameter penalty with values between 0 and 1 that controls relative weighting between 2 types of penalties: the lasso penalty and the ridge penalty. The parameter λ controls the total amount of penalization.[39] The ridge penalty handles multicollinearity by shrinking all coefficients smoothly toward 0 but retains all variables in the model.[40] The lasso penalty allows simultaneous coefficient shrinkage and variable selection, tending to select at most one predictor in each strongly correlated set but at the expense of giving unstable estimates in the presence of high multicollinearity.[41] The elastic net approach of combining the ridge and lasso penalties has the advantage of yielding more stable and accurate estimates than either the ridge or lasso alone while maintaining model parsimony.[38]

The 3-step approach of combining regression trees with penalized regression for variable selection enabled us to incorporate possible interactions and nonlinearities in a clinically meaningful way while controlling for possible overfitting. The R package glmnet program[42] was used to estimate penalized models with MPPs of 0.1, 0.4, 0.7, and 1.0 (an MPP of 0.0 was not used because of multicollinearity in the full predictor set). Internal 10-fold cross-validation selected the coefficient in front of the penalty. Comparative fit across the 20 specifications (ie, 4 MPP values for each of 5 constraints on the number of predictors) was evaluated by inspecting the area under the receiver operating characteristic curve (AUC) and concentration of risk (CR). The CR is the proportion of observed suicides after hospitalizations in each ventile (ie, 20 groups of hospitalizations of equal frequency) ordered by predicted suicide risk. Suicide risk of each hospitalization was calculated using coefficients to project risk as of 12 months after hospital discharge regardless of observed hospitalization data and censoring and standardized by time of hospitalization to adjust for temporal variation in suicide risk. Given that the number of hospitalizations per ventile was much larger than the number of suicides, we focused on the CR in the highest-risk ventile in selecting the best penalized model.

Once a best penalized model was selected, a conventional discrete-time survival model with a logistic link function was estimated using the same predictors as the best penalized model to examine how much the penalty reduced model fit. Because the variance inflation factor of coefficients in this model revealed estimates to be unstable, we also used forward stepwise analysis with a .05-level entry criterion to select a stable subset of predictors for a reduced version of the logistic model. Coefficients in this reduced logistic model were then exponentiated to create odds ratios (ORs) for ease of interpretation. Ventiles from the best penalized model were then collapsed into risk strata using the logic of stratum-specific likelihood ratios.[43] The CR, AUC, and the standardized (for amount of uncensored time observed after each hospitalization) suicide rates per 100 000 person-years were calculated for these risk strata. Finally, parallel rates of risk were calculated for unintentional injury deaths, attempted suicides, and subsequent hospitalizations in the same ventiles to evaluate other adverse outcomes associated with posthospitalization suicide risk.

Results

Patterns of Posthospitalization Suicide

Sixty-eight hospitalized soldiers died by suicide within 12 months of hospital discharge (263.9 suicides per 100 000 person-years vs 18.5 suicides per 100 000 in the total US Army),[23] representing 12.0% of all US Army suicides. An additional 157 hospitalized soldiers died in other ways, and 22 010 others terminated active duty for other reasons (eg, administrative separation and retirement) within 12 months of hospital discharge.

Bivariate Associations of Predictors With Suicide

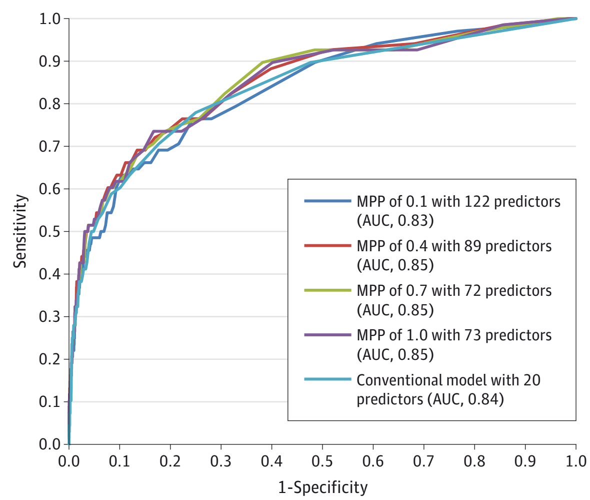
No interactions emerged in more than 10% of regression trees. However, 131 of the 421 bivariate associations (31.1%) between individual predictors and suicides were significant at the .05 level (eTables 5-9 and eTables 11-15 at http://www.armystarrs.org/publications). All these variables were used in the penalized multivariate models.

Selecting a Best Penalized Survival Model

A 10-fold cross-validation revealed that AUC was maximized across the 20 penalized survival models for an MPP of 1.0 (lasso) with 73 predictors and an MPP of 0.1 to 0.7 with 72 to 122 predictors (**Figure 1**). Because the lasso model yielded the best cross-validated CR in the highest-risk ventile (52.9%) (**Table 1**), we estimated a conventional discrete-time survival model with a logistic link function using the same 73 predictors. This model had a much higher AUC (AUC, 0.89) and CR (CR, 61.8%) in the highest-risk ventile than the lasso model with the same predictors, but this was because of overfitting (variance inflation factor >5 for 6 coefficients). Forward stepwise analysis selected a more stable set of predictors in a reduced logistic model, and this model, which contained 20 predictors, had a slightly lower AUC (AUC, 0.84) and CR (CR, 50.0%) in the highest-risk ventile than the lasso model.

Caution is needed in interpreting predictors in the reduced logistic model because the variable selection algorithm maximized overall prediction accuracy rather than individual coefficient accuracy. It is nonetheless noteworthy that the model included variables in all predictor classes (**Table 2**):

Figure 1. Receiver Operating Characteristic (ROC) Curves for Discrete-Time (Person-Month) Elastic Net Penalized Survival Models With Different Mixing Parameter Penalties (MPPs) and for a Conventional Discrete-Time Survival Model Predicting Posthospitalization Suicide



Elastic net penalized survival models were estimated with different MPPs, allowing up to 421 predictors. The best cross-validated model was an MPP of 1.0 with 73 predictors. A conventional discrete-time survival model that contained the same 73 predictors was unstable (variance inflation factor >5.0 for 6 predictors). As a result, we used forward stepwise analysis with a .05-level entry criterion to select a more stable subset of the 73 predictors. Twenty predictors entered that model. The ROC curve shown here for the conventional model is based on those 20 predictors. AUC indicates area under the receiver operating characteristic curve.

3 sociodemographic characteristics (male sex, enlistment at ≥27 years of age, and US Armed Forces Qualification Test score >50th percentile; ORs, 1.9 [95% CI, 1.0-3.5] to 7.9 [95% CI, 1.9-32.6]), access to firearms (number of registered pistols; OR, 1.3; 95% CI, 1.0-1.6), crime perpetration (weapons possession or verbal assault; ORs, 2.2 [95% CI, 1.2-4.0] to 5.6 [95% CI, 1.7-18.3]), prior suicidality (ORs, 1.6 [95% CI, 1.1-2.5] to 2.9 [95% CI, 1.7-4.9]), prior psychiatric treatment (ORs, 0.3 [95% CI, 0.2-0.6] to 5.6 [95% CI, 1.8-17.7]), and characteristics of the focal hospitalization (ORs, 0.4 [95% CI, 0.2-0.7] to 6.0 [95% CI, 2.1-17.4]). The 2 ORs less than 1.0 were for (1) being above the 50th percentile on the ratio of number of psychiatric hospitalizations to time in service and (2) posttraumatic stress disorder during current hospitalization.

CR and Conditional Risk Distributions

Inspection of the CR across predicted risk ventiles led to creation of 4 risk strata. Most suicides occurred in the highest-risk stratum (which was made up of the 5% of hospitalizations in the highest-risk ventile; CR, 52.9%) (**Figure 2**). The CR was lower (CR, 8.8%) in the second stratum (made up of the 5% of hospitalizations in the second-highest ventile), lower still (CR, 4.2%) in a third stratum (made up of the 35% of hospitalizations in the next 7 ventiles), and lowest (CR, 0.8%) in the fourth stratum (made up of the 55% of suicides in the lowest 11 ventiles).

Table 1. CR, AUC, and N_p Values by Mixing Parameter Penalty[a]

Allowed Predictor	Mixing Parameter Penalty 0.1	0.4	0.7	1.0
25				
CR	26.5	29.4	35.3	36.8
AUC	0.71	0.75	0.77	0.79
n_p	30	27	26	30
50				
CR	29.4	41.2	42.6	50.0
AUC	0.74	0.80	0.82	0.84
n_p	53	51	53	56
100				
CR	45.6	51.5	51.5	52.9
AUC	0.82	0.85	0.85	0.85
n_p	109	89	72	73
200				
CR	48.5	51.5	51.5	52.9
AUC	0.84	0.85	0.85	0.85
n_p	122	89	72	73
421				
CR	48.5	51.5	51.5	52.9
AUC	0.84	0.85	0.85	0.85
n_p	122	89	72	73

Abbreviations: AUC, area under the receiver operating characteristic curve; CR, concentration of risk; n_p, number of selected predictors.

[a] The CR is the proportion of all observed posthospitalization suicides that occurred in the 12 months after hospital discharge (or <12 months if the soldier terminated services before 12 months after hospital discharge) that occurred after the 5% of hospitalizations classified by the model as having highest risk of suicide. See the Statistical Analysis section for a discussion of elastic net models and mixing parameter penalties.

Suicide risk ranged from 1338.8 per 100 000 hospitalizations in the highest-risk stratum to 20.3 per 100 000 hospitalizations in the lowest-risk stratum (**Table 3**). However, because mean time in service after hospital discharge was considerably less than 12 months, suicide risk per 100 000 person-years was considerably higher than per 100 000 hospitalizations: 3824.1 per 100 000 person-years in the highest-risk stratum to 40.9 per 100 000 in the lowest-risk stratum.

Stability of Estimates

The CR in the highest-risk stratum did not differ significantly, depending on whether (1) hospitalization was in a facility with a mental health inpatient unit vs a general medical facility without such a unit (48.2% vs 66.7; χ^2_1 = 1.7; *P* = .19); (2) the suicide occurred before vs after September 1, 2008 (median date of suicides during the study period; 38.7% vs 70.3%; χ^2_1 = 2.4; *P* = .12); or (3) the suicide did vs did not occur within 3 months of hospital discharge (median time to postdischarge suicide; 52.6% vs 56.7%; χ^2_1 = 0.0; *P* = .99).

Associations of Suicide Risk With Other Adverse Outcomes

Soldiers in the highest-risk stratum also had elevated risks of other adverse outcomes in the year after hospital discharge, including unintentional injury deaths (CR, 10.1%; χ^2_1 = 7.1; *P* = .008), suicide attempts (CR, 9.1%; χ^2_1 = 332.7; *P* < .001), and subsequent hospitalizations (7.5%; χ^2_1 = 893.4; *P* < .001). Soldiers in the highest predicted suicide risk stratum had 7 unintentional injury deaths, 830 suicide attempts, and 3765 subsequent hospitalizations within 12 months of hospital discharge

Table 2. ORs (95% CIs) and VIFs for the Discrete-Time Logistic Survival Model[a]

Variable	OR (95% CI)	VIF[b]
Sociodemographics		
Male sex (yes/no)	7.9 (1.9-32.6)[c]	1.0
Age of enlistment ≥27 y (yes/no)	1.9 (1.0-3.5)[c]	1.0
AFQT score >50th percentile (yes/no)	3.3 (1.7-10.0)[c]	1.0
Access to firearms		
No. of registered pistols	1.3 (1.0-1.6)[c]	1.0
Crime perpetration		
No. of verbal assault offenses in past 12 mo	2.2 (1.2-4.0)[c]	1.0
Any nonviolent weapons offense in past 24 mo (yes/no)	5.6 (1.7-18.3)[c]	1.0
Suicidal behavior		
Any prior suicide attempt since enlistment (yes/no)	2.9 (1.7-4.9)[c]	1.0
No. of outpatient visits with suicidal ideation in past 12 mo	1.6 (1.1-2.5)[c]	1.1
Other prior treatment		
≥6 Outpatient visits with a mental health professional in past 12 mo (yes/no)	1.9 (1.0-3.6)[c]	1.4
No. of antidepressant prescriptions filled in past 12 mo	1.3 (1.1-1.7)[c]	1.1
No. of psychiatric hospitalizations/time in service >50% percentile (yes/no)	0.3 (0.2-0.6)[c]	1.2
Any prior inpatient psychiatric treatment in past 12 mo (yes/no)	1.8 (0.8-3.7)	1.8
No. of inpatient days in past 12 mo by diagnosis		
Major depression	2.2 (1.1-4.4)[c]	1.4
Somatoform or dissociative disorder	5.6 (1.8-17.7)[c]	1.0
Characteristics of focal hospitalization		
Hospitalized in a civilian psychiatric hospital or civilian facility with a psychiatric unit (yes/no)	1.6 (1.0-2.7)[c]	1.0
Disorders diagnosed during current hospitalization (yes/no)		
PTSD	0.4 (0.2-0.7)[c]	1.1
Suicidal ideation	2.4 (1.3-4.7)[c]	1.0
Nonaffective psychosis	2.9 (1.2-7.0)[c]	1.0
Somatoform or dissociative disorder	3.6 (1.2-10.8)[c]	1.0
Hearing loss	6.0 (2.1-17.4)[c]	1.0

Abbreviations: AFQT, US Armed Forces Qualification Test; OR, odds ratio; PTSD, posttraumatic stress disorder; VIF, variance inflation factor.

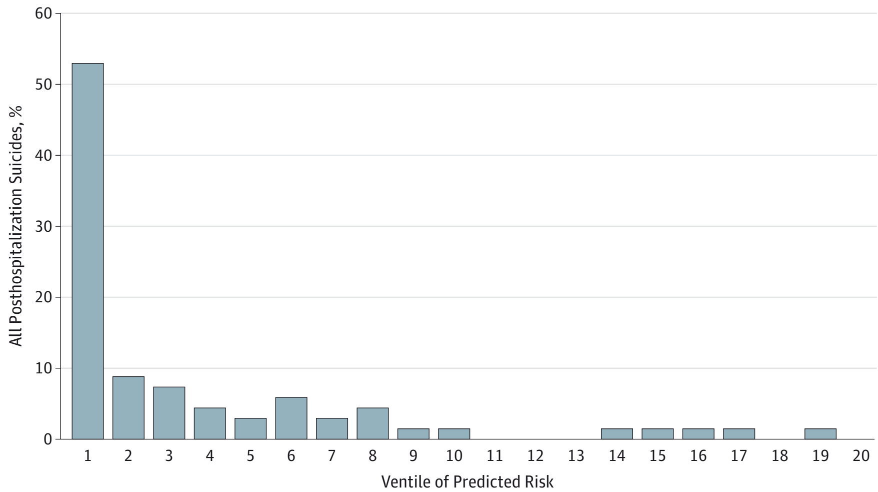
[a] The best penalized survival model was a lasso model with 73 predictors from the total of 421 predictors considered. A conventional discrete-time survival model that contained those same 73 predictors was unstable (VIF >5.0 for 6 predictors). As a result, we used forward stepwise analysis with a .05-level entry criterion to select a more stable subset of the 73 predictors. The coefficients for the 20 predictors that entered are presented here.

[b] The VIF for the coefficient associated with predictor X_i 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 X_i is the dependent variable, and all the other 19 predictors of suicide are included as predictors of X_i. A VIF greater than 5.0 is typically considered an indicator of high multicollinearity.[44]

[c] Significant at the .05 level (2-sided test). However, note that the predictors were selected using stepwise analysis and the current *P* values are consequently inexact.

jamanetwork/2014/psy/11_12_2014/yoi140086pap PAGE: right 5 SESS: 40 OUTPUT: Oct 21 11:51 2014

Figure 2. Concentration of Risk of Posthospitalization Suicides by Ventile of Predicted Risk Based on the Discrete-Time Penalized Survival Model With a Mixing Parameter Penalty of 1.0



Ventiles are 20 groups of hospitalizations of equal frequency (2688 or 2689 hospitalizations), dividing the total sample of 53 769 hospitalizations into groups defined by level of predicted suicide risk.

(492,666.2 per 100 000 person-years). At least one of these outcomes occurred after 46.3% of the highest-risk hospitalizations.

Discussion

Although risk factors for suicide are widely known, synthesizing this information to optimize suicide prediction has been an elusive goal up to now. This study addressed this problem by using machine learning to generate an actuarial suicide risk algorithm from US Army and DoD administrative data, finding that 52.9% of suicides occurred after the 5% of hospitalizations with highest predicted risk. Although interventions in this high-risk stratum would not solve the entire US Army suicide problem given that posthospitalization suicides account for only 12% of all US Army suicides, the algorithm would presumably help target preventive interventions. Before clinical implementation, though, several key issues must be addressed.

The first question is whether the risk algorithm is sufficiently stable to predict future suicides given that it is based on only 68 prior suicides. It is noteworthy that the machine learning methods used to create the algorithm were designed explicitly to maximize stability of predictions. Within-sample stability analyses found that the CR did not vary significantly by type of inpatient facility, year of hospitalization, or number of months since hospital discharge; however, this does not guarantee future stability. Algorithm stability will consequently be tested again in the 2010-2013 US Army suicide data in a future study to address this question.

The second question is whether the risk algorithm improves on clinical judgment. The study was unable to examine this issue empirically because the US Army electronic medi-

Table 3. CR and Conditional Risk of Posthospitalization Suicides by Risk Strata Across All Hospitalizations

	Strata of Predicted Suicide Risk Based on the Lasso Model[a]				
Variable	Highest-Risk Stratum (First Ventile)	Second Ventile	Third to Ninth Ventiles	Lowest-Risk Stratum (10th-20th Ventiles)	Total
Observed No. of suicides	36	6	20	6	68
CR, %[b]	52.9	8.8	4.2	0.9	NA
No. per 100 000 person-years					
Hospitalizations	1338.8	223.3	106.3	20.3	126.5
Person-years	3824.1	538.7	221.1	40.9	263.9
No. of hospitalizations	2689	2687	18 820	29 573	53 769

Abbreviations: CR, concentration of risk; NA, not applicable.

[a] Ventiles of suicide risk are 20 groups of hospitalizations of equal frequency (n = 2688-2689 hospitalizations) dividing the total of 53 769 hospitalizations into groups defined by level of predicted suicide risk. The third through ninth ventiles were collapsed into a single risk stratum based on the fact that observed suicide risk was comparable in these 7 ventiles. The 10th through 20th ventiles were collapsed into a final risk stratum based on similar evidence.

[b] The CR, which is defined as the proportion of all the observed outcomes of the type that occurred in the 12 months after hospital discharge (or <12 months if the soldier terminated services before 12 months after hospital discharge) that occurred in the risk ventile represented by the column heading. The CR is defined separately for each of the 2 highest risk ventiles and then as a per-ventile mean for the next 7 ventiles treated as a single risk stratum and then final 11 ventiles treated as a separate risk stratum.

jamanetwork/2014/psy/11_12_2014/yoi140086pap PAGE: left 6 SESS: 40 OUTPUT: Oct 21 11:51 2014

cal record does not include a structured field where health care professionals must record suicide risk assessments. In addition, documentation of suicide risk assessment in clinical notes was not consistent during the study period. However, with improved documentation after the VA and DoD CPG, comparison of actuarial to clinical prediction may be possible in the future. As noted in the Introduction, though, previous research has indicated that actuarial suicide prediction is much more accurate than prediction based on clinical judgment.[11-14] This evidence is consistent with a large body of literature reporting that actuarial methods are superior to expert judgments in many areas of prediction.[45,46] At the same time, the comprehensive suicide risk assessments required by the new VA and DoD CPG[10] will generate information not included in administrative records. As a result, our algorithm should be seen as a component of this comprehensive clinical assessment rather than a substitute for this assessment.

The third question is whether suicide is sufficiently common in the highest-risk stratum and available interventions sufficiently powerful to make targeted posthospitalization interventions efficient compared with alternative ways of deploying the same clinical resources. Our results shed no light on this question. The potential for harm also has to be taken into consideration because intensive posthospitalization interventions might lead to undue scrutiny by nonmedical leaders that adversely affect soldier careers. This concern is all the more important given that most soldiers identified as being high risk do not commit suicide. Although a formal analysis of comparative risks and benefits is beyond the scope of this report, it is noteworthy that the highest-risk stratum had significantly elevated risks of other adverse outcomes and that prevalence of at least one such outcome was present after 46.3% of highest-risk hospitalizations. Ameliorative effects of expanded high-risk interventions on these outcomes (ie, unintentional injury deaths, suicide attempts, and subsequent hospitalizations) are plausible because numerous risk factors for suicide (eg, depression and substance abuse) are also risk factors for these other outcomes[2,47,48] and most suicide prevention interventions recommended for high-risk patients are likely to affect these outcomes as well.[7,10] These presumed benefits would have to be considered in a broad-based evaluation of risks and benefits of any future targeted high-risk posthospitalization preventive interventions.

The major limitations of our analysis involve errors in the administrative data used as predictors (missing and inconsistent values and errors in *ICD-9-CM* diagnoses). In addition, the algorithm could almost certainly be improved if more nuanced risk factor data were available. Because the new VA and DoD CPG contains a checklist of risk factors health care professionals are urged to assess in evaluating suicide risk, creation of a system to record these assessments in the electronic medical record along with the health care professional's clinical global impression of patient suicide risk might increase the completeness of these assessments and provide a rich source of information for future risk algorithm refinement.

Conclusions

The high concentration of risk of suicides and other adverse outcomes might justify targeting expanded posthospitalization interventions to soldiers classified as having highest posthospitalization suicide risk, although final determination requires careful consideration of intervention costs, comparative effectiveness, and possible adverse effects.

ARTICLE INFORMATION

Submitted for Publication: January 29, 2014; final revision received June 12, 2014; accepted July 21, 2014.

Published Online: November 12, 2014. doi:10.1001/jamapsychiatry.2014.1754.

Author Affiliations: Department of Health Care Policy, Harvard Medical School, Boston, Massachusetts (Kessler, Petukhova, Rose, Gruber, Rosellini, Sampson, Zaslavsky); Department of Behavioral Medicine, Blanchfield Army Community Hospital, Fort Campbell, Kentucky (Warner); US Army Office of the Surgeon General, Falls Church, Virginia (Ivany, Brown); Department of Psychiatry, Stony Brook University School of Medicine, Stony Brook, New York (Bromet); Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts (Cai, Li); National Institute of Mental Health, Bethesda, Maryland (Colpe, Schoenbaum); US Army Public Health Command, Aberdeen Proving Ground, Maryland (Cox, Millikan-Bell); Center for the Study of Traumatic Stress, Department of Psychiatry, Uniformed Services University of the Health Sciences, Bethesda, Maryland (Fullerton, Naifeh, Ursano); Department of Social and Behavioral Sciences, Harvard School of Public Health, Boston, Massachusetts (Gilman); Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts (Gilman); Institute for Social Research, University of Michigan, Ann Arbor (Heeringa, Lewandowski-Romps); Department of Psychology, Harvard University, Cambridge, Massachusetts (Nock); Department of Psychiatry, University of California, San Diego, La Jolla (Stein); Deapartment of Family and Preventive Medicine, University of California, San Diego, La Jolla (Stein); Veterans Affairs San Diego Healthcare System, San Diego, California (Stein); King's Centre for Military Health Research, King's College London, London, United Kingdom (Wessely).

Author Contributions: Dr Kessler had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Study concept and design: Kessler, Cai, Colpe, Fullerton, Gilman, Heeringa, Naifeh, Rosellini, Schoenbaum, Wessely, Ursano.
Acquisition, analysis, or interpretation of data: Kessler, Warner, Ivany, Petukhova, Rose, Bromet, Brown, Cox, Gruber, Heeringa, Lewandowski-Romps, Li, Millikan-Bell, Nock, Rosellini, Sampson, Schoenbaum, Stein, Zaslavsky, Ursano.
Drafting of the manuscript: Kessler, Rosellini, Schoenbaum, Ursano.
Critical revision of the manuscript for important intellectual content: Kessler, Warner, Ivany, Petukhova, Rose, Bromet, Brown, Cai, Colpe, Cox, Fullerton, Gilman, Gruber, Heeringa, Lewandowski-Romps, Li, Millikan-Bell, Naifeh, Nock, Rosellini, Sampson, Stein, Wessely, Zaslavsky, Ursano.
Statistical analysis: Kessler, Ivany, Petukhova, Rose, Cai, Gruber, Heeringa, Li, Rosellini, Sampson, Schoenbaum, Zaslavsky.
Obtained funding: Kessler, Ursano.
Administrative, technical, or material support: Kessler, Brown, Colpe, Fullerton, Heeringa, Lewandowski-Romps, Millikan-Bell, Naifeh, Nock, Rosellini, Ursano.
Study supervision: Kessler, Bromet, Sampson, Schoenbaum, Stein, Wessely, Ursano.

Conflict of Interest Disclosures: Dr Kessler reported being a consultant for AstraZeneca, Analysis Group, Bristol-Myers Squibb, Cerner-Galt Associates, Eli Lilly & Company, GlaxoSmithKline Inc, HealthCore Inc, Health Dialog, Hoffman-LaRoche Inc, Integrated Benefits Institute, J & J Wellness & Prevention Inc, John Snow Inc, Kaiser Permanente, Lake Nona Institute, Matria Inc, Mensante, Merck & Co Inc, Ortho-McNeil Janssen Scientific Affairs, Pfizer Inc, Primary Care Network, Research Triangle Institute, Sanofi-Aventis Groupe SA, Shire US Inc, SRA International Inc, Takeda Global Research & Development, Transcept Pharmaceuticals Inc, and Wyeth-Ayerst. Dr Kessler reported serving on advisory boards for Appliance

Computing II, Eli Lilly & Company, Mindsite, Ortho-McNeil Janssen Scientific Affairs, Johnson & Johnson, Plus One Health Management, and Wyeth-Ayerst and receiving research support for his epidemiologic studies from Analysis Group Inc, Bristol-Myers Squibb, Eli Lilly & Company, EPI-Q, GlaxoSmithKline, Johnson & Johnson Pharmaceuticals, Ortho-McNeil Janssen Scientific Affairs, Pfizer Inc, Sanofi-Aventis Groupe SA, Shire US Inc, and Walgreens Co. Dr Kessler reported owing a 25% share in DataStat Inc. Dr Stein reported being a consultant for Healthcare Management Technologies and receiving research support for pharmacologic imaging studies from Janssen. No other disclosures were reported.

Funding/Support: The Army STARRS was sponsored by the US Department of the Army and funded under cooperative agreement U01MH087981 with the National Institute of Mental Health, National Institutes of Health, US Department of Health and Human Services.

Role of the Funder/Sponsor: As a cooperative agreement, scientists employed by the National Institute of Mental Health (Drs Colpe and Schoenbaum) and US Army liaisons and consultants (Dr Cox and Steven Cersovsky, MD, MPH) collaborated to develop the study protocol and data collection instruments, supervise data collection, interpret results, and prepare reports. Although a draft of the manuscript was submitted to the US Army and the National Institute of Mental Health for review and comment before submission, this was with the understanding that comments would be only advisory.

Group Information: The Army STARRS coprincipal investigators were Robert J. Ursano, MD (Uniformed Services University of the Health Sciences), and Murray B. Stein, MD, MPH (University of California, San Diego, and Veterans Affairs San Diego Healthcare System). The site principal investigators were Steven G. Heeringa, PhD (University of Michigan), and Ronald C. Kessler, PhD (Harvard Medical School). The National Institute of Mental Health collaborating scientists were Lisa J. Colpe, PhD, MPH, and Michael Schoenbaum, PhD. The US Army liaisons and consultants were Steven Cersovsky, MD, MPH, and Kenneth L. Cox, MD, MPH.

Other team members were Pablo A. Aliaga, MA (Uniformed Services University of the Health Sciences), David M. Benedek, MD (Uniformed Services University of the Health Sciences), Susan Borja, PhD (National Institute of Mental Health), Gregory G. Brown, PhD (University of California, San Diego), Laura Campbell-Sills, PhD (University of California, San Diego), Catherine L. Dempsey, PhD, MPH (Uniformed Services University of the Health Sciences), Richard Frank, PhD (Harvard Medical School), Carol S. Fullerton, PhD (Uniformed Services University of the Health Sciences), Nancy Gebler, MA (University of Michigan), Robert K. Gifford, PhD (Uniformed Services University of the Health Sciences), Stephen E. Gilman, ScD (Harvard School of Public Health), Marjan G. Holloway, PhD (Uniformed Services University of the Health Sciences), Paul E. Hurwitz, MPH (Uniformed Services University of the Health Sciences), Sonia Jain, PhD (University of California, San Diego), Tzu-Cheg Kao, PhD (Uniformed Services University of the Health Sciences), Karestan C. Koenen, PhD (Columbia University), Lisa Lewandowski-Romps, PhD (University of Michigan), Holly Herberman Mash, PhD (Uniformed Services University of the Health Sciences), James E. McCarroll, PhD, MPH (Uniformed Services University of the Health Sciences), Katie A. McLaughlin, PhD (Harvard Medical School), James A. Naifeh, PhD (Uniformed Services University of the Health Sciences), Matthew K. Nock, PhD (Harvard University), Rema Raman, PhD (University of California, San Diego), Sherri Rose, PhD (Harvard Medical School), Anthony Joseph Rosellini, PhD (Harvard Medical School), Nancy A. Sampson, BA (Harvard Medical School), LCDR Patcho Santiago, MD, MPH (Uniformed Services University of the Health Sciences), Michaelle Scanlon, MBA (National Institute of Mental Health), Jordan Smoller, MD, ScD (Harvard Medical School), Michael L. Thomas, PhD (University of California, San Diego), Patti L. Vegella, MS, MA (Uniformed Services University of the Health Sciences), Christina Wassel, PhD (University of Pittsburgh), and Alan M. Zaslavsky, PhD (Harvard Medical School).

Disclaimer: The contents are solely the responsibility of the authors and do not necessarily represent the views of the US Department of Health and Human Services, the National Institute of Mental Health, the US Department of the Army, or the US Department of Defense.

Additional Contributions: John Mann, MD, Maria Oquendo, MD, Barbara Stanley, PhD, Kelly Posner, PhD, and John Keilp, PhD, Department of Psychiatry, Columbia University, College of Physicians and Surgeons, and New York State Psychiatric Institute, New York, contributed to the early stages of the US Army STARRS development.

REFERENCES

1. Armed Forces Health Surveillance Center. Deaths by suicide while on active duty, active and reserve components, US Armed Forces, 1998-2011. *Med Surveill Monthly Rep.* 2012;19(6):7-10.

2. Nock MK, Deming CA, Fullerton CS, et al. Suicide among soldiers: a review of psychosocial risk and protective factors. *Psychiatry*. 2013;76(2):97-125.

3. Zamorski MA. Suicide prevention in military organizations. *Int Rev Psychiatry*. 2011;23(2):173-180.

4. Valenstein M, Kim HM, Ganoczy D, et al. Higher-risk periods for suicide among VA patients receiving depression treatment: prioritizing suicide prevention efforts. *J Affect Disord*. 2009;112(1-3):50-58.

5. Luxton DD, Trofimovich L, Clark LL. Suicide risk among US Service members after psychiatric hospitalization, 2001-2011. *Psychiatr Serv*. 2013;64(7):626-629.

6. Olfson M, Marcus SC, Bridge JA. Focusing suicide prevention on periods of high risk. *JAMA*. 2014;311(11):1107-1108.

7. While D, Bickley H, Roscoe A, et al. Implementation of mental health service recommendations in England and Wales and suicide rates, 1997-2006: a cross-sectional and before-and-after observational study. *Lancet*. 2012;379(9820):1005-1012.

8. Paton MB, Large MM, Ryan CJ. Debate: clinical risk categorisation is valuable in the prevention of suicide and severe violence–no. *Australas Psychiatry*. 2014;22(1):10-12.

9. Berrouiguet S, Gravey M, Le Galudec M, Alavi Z, Walter M. Post-acute crisis text messaging outreach for suicide prevention: a pilot study. *Psychiatry Res*. 2014;217(3):154-157.

10. US Department of Veterans Affairs and US Department of Defense. *Assessment and Management of Patients at Risk for Suicide.* Washington, DC: US Dept of Veterans Affairs and US Dept of Defense; 2013.

11. Erdman HP, Greist JH, Gustafson DH, Taves JE, Klein MH. Suicide risk prediction by computer interview: a prospective study. *J Clin Psychiatry*. 1987;48(12):464-467.

12. Gustafson DH, Greist JH, Stauss FF, Erdman H, Laughren T. A probabilistic system for identifying suicide attemptors. *Comput Biomed Res*. 1977;10(2):83-89.

13. Gustafson DH, Tianen B, Greist JH. A computer-based system for identifying suicide attemptors. *Comput Biomed Res*. 1981;14(2):144-157.

14. Nock MK, Park JM, Finn CT, Deliberto TL, Dour HJ, Banaji MR. Measuring the suicidal mind: implicit cognition predicts suicidal behavior. *Psychol Sci*. 2010;21(4):511-517.

15. Bright TJ, Wong A, Dhurjati R, et al. Effect of clinical decision-support systems: a systematic review. *Ann Intern Med*. 2012;157(1):29-43.

16. Garg AX, Adhikari NK, McDonald H, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA*. 2005;293(10):1223-1238.

17. Large M, Sharma S, Cannon E, Ryan C, Nielssen O. Risk factors for suicide within a year of discharge from psychiatric hospital: a systematic meta-analysis. *Aust N Z J Psychiatry*. 2011;45(8):619-628.

18. Weber GM, Mandl KD, Kohane IS. Finding the missing link for big biomedical data. *JAMA*. 2014;311(24):2479-2480.

19. Bachynski KE, Canham-Chervak M, Black SA, Dada EO, Millikan AM, Jones BH. Mental health risk factors for suicides in the US Army, 2007-8. *Inj Prev*. 2012;18(6):405-412.

20. Bell NS, Harford TC, Amoroso PJ, Hollander IE, Kay AB. Prior health care utilization patterns and suicide among U.S. Army soldiers. *Suicide Life Threat Behav*. 2010;40(4):407-415.

21. Black SA, Gallaway MS, Bell MR, Ritchie EC. Prevalence and risk factors associated with suicides of Army soldiers 2001–2009. *Mil Psychol*. 2011;23(4):433-451.

22. Hyman J, Ireland R, Frost L, Cottrell L. Suicide incidence and risk factors in an active duty US military population. *Am J Public Health*. 2012;102(suppl 1):S138-S146.

23. Schoenbaum M, Kessler RC, Gilman SE, et al; Army STARRS Collaborators. Predictors of suicide and accident death in the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS): results from the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *JAMA Psychiatry*. 2014;71(5):493-503.

24. Ursano RJ, Colpe LJ, Heeringa SG, Kessler RC, Schoenbaum M, Stein MB; Army STARRS collaborators. The Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *Psychiatry*. 2014;77(2):107-119.

25. Pirkola S, Sohlman B, Wahlbeck K. The characteristics of suicides within a week of

discharge after psychiatric hospitalisation: a nationwide register study. *BMC Psychiatry*. 2005;5:32.

26. Kessler RC, Colpe LJ, Fullerton CS, et al. Design of the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *Int J Methods Psychiatr Res*. 2013;22(4):267-275.

27. Troister T, Links PS, Cutcliffe J. Review of predictors of suicide within 1 year of discharge from a psychiatric hospital. *Curr Psychiatry Rep*. 2008;10(1):60-65.

28. Bickley H, Hunt IM, Windfuhr K, Shaw J, Appleby L, Kapur N. Suicide within two weeks of discharge from psychiatric inpatient care: a case-control study. *Psychiatr Serv*. 2013;64(7):653-659.

29. Park S, Choi JW, Kyoung Yi K, Hong JP. Suicide mortality and risk factors in the 12 months after discharge from psychiatric inpatient care in Korea: 1989-2006. *Psychiatry Res*. 2013;208(2):145-150.

30. Rubin DB. *Introduction. Multiple Imputation for Nonresponse in Surveys.* Hoboken, NJ: John Wiley & Sons; 2008:1-26.

31. Efron B. Logistic regression, survival analysis, and the Kaplan-Meier curve. *J Am Stat Assoc*. 1988;83(402):414-425.

32. Thernau T, Atkinson B, Ripley B. Rpart: Recursive Partitioning. R Package 4.1-0. http://CRAN.R-project.org/package=rpart. Accessed December 15, 2013.

33. Zhang H, Singer BH. *Recursive Partitioning and Applications.* 2nd ed. New York, NY: Springer; 2010.

34. Draper NR, Smith H. *Applied Regression Analysis.* 2nd ed. Hoboken, NJ: John Wiley & Sons; 1981.

35. Berk RA. *Regression Analysis: A Constructive Critique.* Vol 11. New York, NY: Sage; 2004.

36. Bellazzi R, Zupan B. Towards knowledge-based gene expression data mining. *J Biomed Inform*. 2007;40(6):787-802.

37. van der Laan MJ, Rose S. *Targeted Learning: Causal Inference for Observational and Experimental Data.* New York, NY: Springer; 2011.

38. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Series B Stat Methodol*. 2005;67(suppl):301-320.

39. Berk RA. *Statistical Learning From a Regression Perspective.* New York, NY: Springer; 2008.

40. Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*. 1970;12(1):55-67.

41. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Series B Stat Methodol*. 1996;58:267-288.

42. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33(1):1-22.

43. Schmitz N, Kruse J, Tress W. Application of stratum-specific likelihood ratios in mental health screening. *Soc Psychiatry Psychiatr Epidemiol*. 2000;35(8):375-379.

44. Stine RA. Graphical interpretation of variance inflation factors. *Am Stat*. 1995;49(1):53-56.

45. Dawes RM, Faust D, Meehl PE. Clinical versus actuarial judgment. *Science*. 1989;243(4899):1668-1674.

46. Grove WM, Zald DH, Lebow BS, Snitz BE, Nelson C. Clinical versus mechanical prediction: a meta-analysis. *Psychol Assess*. 2000;12(1):19-30.

47. Victor SE, Klonsky ED. Correlates of suicide attempts among self-injurers: a meta-analysis. *Clin Psychol Rev*. 2014;34(4):282-297.

48. Scott-Parker B, Watson B, King MJ, Hyde MK. The influence of sensitivity to reward and punishment, propensity for sensation seeking, depression, and anxiety on the risky behaviour of novice drivers: a path model. *Br J Psychol*. 2012;103(2):248-267.

jamanetwork/2014/psy/11_12_2014/yoi140086pap PAGE: right 9 SESS: 40 OUTPUT: Oct 21 11:51 2014

Appendix D: Quad Chart

Behavioral-based predictors of workplace violence in the Army STARRS

PI: Ronald C. Kessler, PhD **Org:** Harvard Medical School

Study/Product Aim(s)

To develop practical risk prediction indices for workplace violence perpetration and victimization based on analyses of data in the Army Study to Assess Risk and Resilience in Servicemembers (A-STARRS). A-STARRS data include administrative records from many different Army/DoD data-bases, self-report survey data, neurocognitive test results, and blood samples. This large and rich databases creates a unique opportunity to develop practical risk prediction indices for workplace violence perpetration and victimization.

Approach and Military Relevance

The preliminary A-STARRS data report very high proportions of Soldiers who either perpetrate and/or are victimized by workplace violence. The results of the proposed research could generate new information about the most powerful risk and resilience factors for preventing workplace violence in the Army and targeting outreach efforts for high risk Soldiers.

ARMY STARRS

Army Study to Assess Risk and Resilience in Servicemembers

Behavioral-based predictors of workplace violence in the Army STARRS

Estimated Timeline

Activities Year	1	2	3	4
Build HADS survival files. Construct Basic models of HADS outcomes				
Add conceptually-guided interactions & cross-validated data-mining models				
Expand models of HADS outcomes with survey and neurocognitive predictors				
Develop & test final workplace violence Perpetration-victimization indices				

Q8 Report

Projected Goals/Milestones

Year 1: Due to IRB and DUA delays, work was delayed. Coded survey data. Developed a coding scheme to code violent crimes into NCRP categories.

Year 2: Disaggregated, selected, and examined rates of administratively recorded workplace violence perpetration and victimization in the HADS and NSS. Examined rates of self-reported workplace violence in the AAS and PPDS. Constructed HADS and NSS predictors. Built person-month data files for HADS and NSS data mining models for predicting first occurrence of workplace violence outcomes. Examined bivariate associations. Used cross-validated data mining to maximize prediction accuracy and concentration of risk and identify final predictor sets for 15 HADS and NSS outcomes. Examined the relative strength of the HADS and NSS predictors

Year 3: Finalize HADS and NSS data mining models. Develop parallel data mining models predicting recurrence of workplace violence in the HADS. Develop parallel data mining models predicting first occurrence of self-reported workplace violence in the AAS and PPDS. Continue analyses that distinguish competing interpretations of significant predictors. Complete all data mining analyses.

Year 4: Carry out conceptually-guided interaction analyses based on the stressor-emotion and rational choice models of workplace violence. Finalize computer programs to predict perpetration and victimization. Finalize analyses and recommendations for optimal targets for interventions.