Published in final edited form as:

Am J Prev Med. 2017 November ; 53(5): 661–669. doi:10.1016/j.amepre.2017.06.022.

Predicting Sexual Assault Perpetration in the US Army Using Administrative Data

Anthony J. Rosellini, PhD¹, John Monahan, PhD², Amy E. Street, PhD^{3,4}, Maria V. Petukhova, PhD¹, Nancy A. Sampson, BA¹, David M. Benedek, MD⁵, Paul Bliese, PhD⁶, Murray B. Stein, MD, MPH, FRCPC^{7,8}, Robert J. Ursano, MD⁵, and Ronald C. Kessler, PhD¹

¹Department of Health Care Policy, Harvard Medical School, Boston, Massachusetts, USA

²School of Law, University of Virginia, Charlottesville, Virginia, USA

³National Center for PTSD, VA Boston Healthcare System, Boston, Massachusetts, USA

⁴Department of Psychiatry, Boston University School of Medicine, Boston, Massachusetts, USA

⁵Center for the Study of Traumatic Stress, Department of Psychiatry, Uniformed Services University School of Medicine, Bethesda, Maryland, USA

⁶Darla Moore School of Business, University of South Carolina, Columbia, South Carolina, USA

⁷Departments of Psychiatry and Family Medicine & Public Health, University of California San Diego, La Jolla, California, USA

⁸VA San Diego Healthcare System, San Diego, California, USA

Abstract

Introduction—The Department of Defense uses a universal prevention framework for sexual assault prevention, with each branch implementing their own branch-wide programs. Intensive interventions exist, but would be cost-effective only if targeted at high-risk personnel. This study developed actuarial models to identify male U.S. Army soldiers at high risk of administratively-recorded sexual assault perpetration.

Methods—This study investigated administratively-recorded sexual assault perpetration among the 821,807 male Army soldiers serving 2004–2009. Other temporally prior administrative data were used as predictors. Penalized discrete-time (person-month) survival analysis (conducted in 2016) was used to select the smallest possible number of stable predictors to maximize number of sexual assaults among the 5% of soldiers with highest predicted risk of perpetration (*top-ventile concentration of risk [COR]*). Separate models were developed for assaults against non-family and intra-family adults and minors.

Results—4,640 male soldiers were found to be perpetrators against non-family adults, 1,384 against non-family minors, 380 against intra-family adults, and 335 against intra-family minors.

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

Conflict of interest

The remaining authors report nothing to disclose.

Top-ventile COR was 16.2–20.2% predicting perpetration against non-family adults and minors and 34.2–65.1% against intra-family adults and minors. Final predictors consisted largely of measures of prior crime involvement and the presence-treatment of mental disorders.

Conclusions—Administrative data can be used to develop actuarial models that identify a high proportion of sexual assault perpetrators. If a system is developed to routinely consolidate administrative predictors, predictions could be generated periodically to identify those in need of preventive intervention. Whether this would be cost-effective, though, would depend on intervention costs, effectiveness, and competing risks.

Keywords

sexual assault; perpetration; military; prediction model; risk model; violence prediction

Introduction

Sexual assault is a significant concern of the Department of Defense (DoD) and U.S. Army.¹ Similar to civilians, roughly 5% of active duty women report past-year sexual assault victimization.² Although prevalence of sexual assault is believed to be decreasing,³ there have also been increases in the number of victims coming forward to report their experiences.¹ The Department of Defense (DoD) has responded by creating a framework of universal prevention requirements that is used by each branch of the military to implement branch-specific programs to decrease sexual assault.⁴ More intensive preventive interventions exist,^{5,6} but it is likely that they would be cost-effective only if they were targeted at high-risk individuals.^{7,8}

Although there is a rich civilian (including forensic) literature,^{9–11} few studies have examined risk factors for sexual assault perpetration among military personnel.^{12,13} In both civilian and military literatures, personality traits (e.g., psychopathy), attitudes-beliefs (e.g., hyper-masculinity), and prior behavioral problems (e.g., heavy alcohol use, delinquency) have been associated with increased perpetration risk. Knowledge of these risk factors has led to the development of recidivism risk prediction tools in forensic psychiatric settings for prior sexual assault perpetrators.^{14,15} However, adaptation of these tools to a military population would be infeasible due to their reliance on intensive clinical evaluations. No tool has been developed to target male soldiers at high risk of sexual assault perpetration.

An alternative approach used with good success for sexual assault victimization¹⁶ is to develop an actuarial prediction model using the extensive Army/DoD administrative data available for all soldiers. These administrative data include a plethora of information about prior behavioral problems (e.g., prior criminal behavior, presence-treatment of mental and substance use disorders). In addition to identifying soldiers at high risk of sexual assault victimization, this approach has been used recently to predict non-sexual violence perpetration.^{17,18} The goal of the current report is to present results of an effort to develop a comparable model for soldier sexual assault perpetration. Consistent with current research on risk factors for sexual assault perpetration,¹⁹ separate models were developed for perpetration against: (i) non-family adults, (ii) non-family minors, (iii) intra-family adults, and (iv) intra-family minors.

Methods

Sample

The Historical Administrative Data System (HADS) of the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS)²⁰ was used to develop the models. The HADS merged de-identified data across 38 Army/DoD administrative sources (Appendix Table 1) for the 975,057 Regular U.S. Army soldiers (821,807 men) serving between January 1, 2004 and December 31, 2009.²¹ As detailed below, the models were designed to predict the first administratively-recorded occurrence of each outcome using discrete-time survival analysis with person-month as the unit of analysis.²² In order to deal with the enormous sample of person-months (over 32 million) generated by this approach, the logic of case-control analysis was used to select a stratified probability sample of control person-months and then weighted controls by the inverse of their probability of selection to reconstruct population proportions.²³

Measures

Sexual assault perpetration—Six HADS databases were used to obtain information on date, type, and judicial outcome of all reported crimes occurring over the study period. Crime types were coded according to the Bureau of Justice Statistics National Corrections Reporting Program (NCRP) classification system.²⁴ Qualifying sexual assault crimes included (Appendix Table 2): rape (i.e., forcible vaginal intercourse), forcible sodomy (i.e., attempted or forcible oral or anal sex), and “other” sexual assault (i.e., attempted rape, fondling, indecent assault). The outcomes were *founded* cases; that is, cases for which the Army found sufficient evidence to warrant full investigation regardless of whether the investigation resulted in a formal conviction. This focus on founded cases is consistent with other research on administratively-recorded sexual crimes, which typically focus on arrests rather than convictions based on the fact that arrest records reflect actual violent behaviors much more closely than do conviction records.^{1,25,26} Cases of intra-family perpetration were first identified using the Army Central Registry (ACR), a database containing information on reported cases of spouse and child abuse and neglect involving married (currently or previously) Army personnel. Acts of sexual assault perpetration would thus be recorded in the ACR if the victim was a soldier’s dependent or beneficiary. Additional intra-family cases were identified in the other databases using offense code descriptions (e.g., “rape of a family member, wife, by force”). A special flag was used when offense code descriptions involved victims who were minors. For example, the most common offense code for sexual assault perpetration against minors was “child molestation – indecent acts upon a child.” A total of 4,640 men had records indicating an occurrence of sexual assault perpetration against non-family adults, 1,384 against non-family minors, 380 against intra-family adults, and 335 against intra-family minors. All four outcomes included perpetrations against both opposite-sex and same-sex victims, though data were not available to distinguish between the two.

Predictors—As mentioned in the introduction, there is a rich civilian literature on risk factors for sexual assault perpetration,^{9–11} but few studies have examined these risk factors among military personnel.^{12,13} A considerably larger literature has examined predictors of any (physical or sexual) violence perpetration among military personnel.^{27–34} As reviewed

by Elbogen and colleagues,³⁵ four broad classes of predictors have been identified in these military studies: socio-demographic and dispositional predictors (e.g., sex, race-ethnicity, personality); historical predictors (e.g., childhood experiences, military career experiences, prior violence); clinical predictors (e.g., mental and physical disorders); and contextual-environmental predictors (e.g., access to weapons). Given that the variables were created from administrative data collected for other purposes, measures were not available for all the significant predictors found in these previous studies. However, 755 HADS variables were identified as being indicators of one or more predictors found to be significant in previous studies. These included 19 socio-demographic, 110 historical (44 defining military career experiences and 66 representing prior crime perpetration-victimization of the soldier), 597 clinical (measures of specific treated mental disorders and broader classes of mental and physical disorders [425 variables]; classes of filled prescriptions [172 variables]), and 29 contextual-environmental (unit characteristics; number of registered weapons owned by the soldier) variables. These variables were defined based on administrative records as of the month prior to the month of occurrence of sexual assault (additional details about variable construction are available by request).

Statistical analysis

Analysis was carried out (in 2016) remotely by Harvard Medical School analysts on a secure server at the University of Michigan Army STARRS Data Coordination Center with approval from the Human Subjects Committees of all Army STARRS collaborating organizations. The governing Institutional Review Boards did not require informed consent because HADS data were de-identified. Analysis began by using cross-tabulations to calculate incidence of the four sexual assault perpetration outcomes expressed for descriptive purposes as number of founded crimes per 100,000 person-years. Model-building was then carried out based on discrete-time person-month survival analysis.²² The latter was used rather than incidence analysis because previous research has shown that the examination of risk factors based on incidence analysis (e.g., number of cases per 100,000 person-years) can yield inaccurate results.³⁶ It is noteworthy in this regard that the person-month survival models allowed for time-varying values of the risk factors, making it possible to examine the possibility of high-risk time periods in the careers of individual soldiers. This model feature, coupled with the fact that the predictors used here are routinely updated for each soldier each month, means that the Army could use results to generate new predicted probabilities of sexual assault perpetration each month for each soldier in the Army as a way of monitoring trends in risk of sexual assault perpetration.

The major challenge in developing these models was the possibility of over-fitting due to the large number of potential predictors. Machine learning methods were used to minimize this problem. Elastic net penalized regression,³⁷ implemented in the R glmnet package³⁸ (R, version 3.2.4.), was used to select the subset of predictors with the most stable cross-validated predictive associations with the outcomes. Cross-validation was used instead of significance testing to select predictors based on the fact that penalized regression searches through vast numbers of possible combinations to select final models. The odds-ratios (ORs) used to characterize the associations of predictors with outcomes in the elastic net models consequently are not accompanied by confidence intervals. A tuning parameter needs to be

specified in elastic net to define the mix of penalties used in estimation. This parameter was set to maximize the cross-validated proportion of observed sexual assault penetration cases among the 5% of soldiers with highest predicted risk (referred to below as *top-ventile concentration of risk [COR]*). An accessible discussion of elastic net penalties and parameter tuning can be found elsewhere.³⁹ Model fit of the final models was characterized by calculating area under the receiver operating characteristic curve (AUC) and top-ventile COR. In addition, all person-months were ranked by their penalized cross-validated predicted probabilities and grouped into 20 categories of equal size (ventiles) to examine the distribution of COR.

Results

Incidence by time in service

4,640 male soldiers were found over the study period to be perpetrators against non-family adults (175.5/100,000 person-years), 1,384 against non-family minors (52.2/100,000 person-years), 380 against intra-family adults (14.3/100,000 person-years), and 335 against intra-family minors (12.6/100,000 person-years) (Appendix Table 3). The total of 6,739 across these four categories (254.6/100,000 person-years) represents 93.5 new founded cases per month over the study period. Incidence varied significantly by time in service for all four outcomes ($\chi^2_6=43.3-492.7$, $ps<.001$). Incidence of perpetration against intra-family minors was highest among men with five or more years in service (18.8–19.6/100,000 person years), while incidence of the other three outcomes was highest among soldiers with less than four years in service (non-family adults: 197.7–257.0/100,000 person-years; non-family minors: 53.3–66.0/100,000 person-years; intra-family adults: 10.9–22.6/100,000 person-years).

Strength of prediction in the final models

The number of final model predictors was between 11 (intra-family minors) and 34 (non-family adults). AUCs were 0.71 for non-family adults, 0.68 for non-family minors, 0.73 for intra-family adults, and 0.83 for intra-family minors (Table 1). Top-ventile COR was 20.2% for non-family adults, 16.2% for non-family minors, 34.2% for intra-family adults, 65.1% for intra-family minors, and 19.2% overall. Given that overall incidence was 254.6/100,000 person-years, 19.2% top-ventile COR across all four types of perpetration translates into an observed incidence of 977.7/100,000 person-years across all four types in the top-risk ventile. And this, in turn, is equivalent to approximately 15 new founded cases of sexual assault perpetration per month among the 18,750 men in the top-risk ventile out of the average of 375,000 male Regular Army soldiers who were in service at a point in time over the study period. COR was dramatically lower in subsequent ventiles (Figure 1), but in most instances remained higher than the expected value of 5% for the next three ventiles and had monotonically decreasing values in later ventiles.

Final model predictors

Common predictors—Soldiers with less than a college education had elevated odds of both non-family types of perpetration. Previously or never deployed soldiers (i.e., soldiers who were not currently deployed) had elevated odds of three outcomes (all other than

perpetration against intra-family minors). Indicators of junior career status (e.g., low rank; non-officer) were associated with increased odds of both non-family outcomes. Various Army command variables were also included in both non-family models.

For prior crime, variables representing perpetration of any crime in the past 12 or 24 months were predictors in all four models. Prior perpetration of major physical violence was included in both intra-family models, while prior perpetration of verbal violence and of non-violent crime were included in both adult models.

In regard to the clinical variables, measures of the presence-treatment of any physical disorder were associated with elevated odds of all four outcomes. Impulse control disorders entered both non-family models. Alcohol and drug problems entered both non-family models as well as the intra-family adult model. Having a problem that was treated with antidepressant pharmacotherapy entered both the non-family and intra-family adult models.

Non-family adult predictors—Unmarried, racial-ethnic minority, and combat support or service support soldiers had elevated odds of perpetration against a non-family adult. Past 12 month demotion and prior perpetration of minor violence and non-violent possession-use of illegal substances were also included in the non-family adult model. Variables for the presence-treatment of adjustment disorder, insomnia, and tobacco use, and non-narcotic analgesic pharmacotherapy were also selected.

Non-family minor predictors—A select number of clinical variables were unique to the non-family minor model, including the presence-treatment of pain, depression, suicidal ideation, injury/poisoning (including suicide attempt), personality disorder, and drug-induced mental illness.

Intra-family adult predictors—The presence-treatment of marital problems was the only unique predictor of intra-family adult perpetration.

Intra-family minor predictors—Prior perpetration of non-family adult sexual violence was only included in the model for intra-family minors, as were variables representing the presence-treatment of stress/adversity and any stressor/stress disorder.

Discussion

The current report used machine learning methods to develop actuarial models for administratively-recorded sexual assault perpetration. The final models had substantially elevated proportions of observed perpetrators in the highest ventiles of predicted risk, especially for the intra-family outcomes.

It is difficult to compare COR to similar operating characteristics of other violence prediction tools developed for use in forensic samples⁴⁰ because of vast differences in populations. However, comparison of overall model strength (AUC), which should be done only with great caution for the same reason,⁴¹ shows that the AUCs of our models are similar to those of leading sexual violence risk tools that require one-on-one clinical assessment.^{40,42} Using conventional interpretive guidelines,⁴³ the AUC of the intra-family

minor model was *excellent* and those of the other three models at the lower-end of *acceptable*. Replication of the analysis in more recent cohorts would be needed to determine if sufficiently strong-stable models could be developed to make intensive prevention efforts cost-effective.

Caution is also needed in interpreting the specific predictors in the final models because our goal was to optimize overall model performance rather than to test hypotheses about specific predictors. The appearance of a given variable in the final models consequently should be taken as a marker of the importance of the set of all associated characteristics in the full predictor set, as in many cases it would be possible to delete a given predictor from the model and achieve comparable model prediction accuracy by instead adding some other correlate of that predictor to the model.¹⁷

In the context of that caution, it is noteworthy that the vast majority of final model predictors involved either prior crime perpetration (all four models) or presence-treatment of mental (all models) or substance (three models) disorders. This is consistent with findings from prior studies of risk factors for violence perpetration among soldiers^{13,27,33} and civilians,^{9–11} and with existing violence risk tools developed in forensic psychiatric samples.^{14,44,45} Indicators of prior severe violence perpetration (i.e., major physical violence; non-family sexual assault) entered only the intra-family models, while mental disorders requiring inpatient treatment entered only the non-family models. This may indicate different risk profiles whereby non-family sexual assault is preceded by severe emotional-substance problems and intra-family perpetration by severe physical violence.

Limitations

The use of administrative data means that it was not possible to include measures of certain dispositional factors (e.g., hostility toward women) found to be significant predictors of sexual assault perpetration in previous studies.^{10,13} Nor did the data include pre-enlistment historical factors (e.g., pre-enlistment perpetration; childhood sexual abuse victimization) found to be significant predictors in previous studies.^{10,12} Model performance may have been improved had such data been available. The reliance on administratively-defined *founded* crimes (i.e., sexual assaults that were reported and thoroughly investigated) also prevented identification of perpetrators who were never apprehended. It is noteworthy in this regard that anonymous surveys suggest that a nontrivial proportion of sexual assaults occurring during military service are never reported to authorities.^{16,46} Perpetrators who are apprehended may be qualitatively different than those who are not (e.g., more violent). The most realistic way to learn of perpetrators who were not apprehended would be to carry out surveys asking male soldiers to self-report perpetration. However, existing evidence suggests that most perpetrators deny such behaviors when asked in surveys.⁴⁷

Conclusions

Machine learning methods can be used to develop actuarial models with high proportions of sexual assault perpetrators classified in the top ventiles of predicted risk. Given that administrative data are routinely collected for all soldiers, it would be possible for the U.S. Army to implement the current models or extensions based on more recent cohorts on an

ongoing basis if resources were devoted to creating a system to regularly consolidate the administrative predictors used here. Although it is beyond the scope of this report to consider all the practical challenges associated with implementing intensive preventive interventions for high-risk soldiers, a few warrant additional comment. First, a recent systematic review of primary prevention strategies for civilian sexual violence perpetration found very few interventions to be effective in reducing sexual violence, with some interventions surprisingly demonstrating iatrogenic effects.⁴⁸ Designing, implementing, and evaluating an intensive intervention for service members at high-risk of sexual assault must thus be done thoughtfully and cautiously, and would likely require legal and ethical review within the DoD. A related secondary issue is the potential negative stigma associated with being required to participate in an intensive sexual assault perpetration prevention program. Sensitivity would be needed when informing soldiers they are high-risk (e.g., high-risk for “negative health-behaviors”?), titling the program (e.g., improving relationship? Managing impulsivity?), and structuring the intervention (e.g., focusing only on sexual behaviors versus combining the program with other intensive prevention programs focused on suicide, substance use, etc.). Special attention may also be needed to maximize soldier engagement. Accordingly, additional research is needed to weigh considerations of likely preventive intervention effectiveness, efficiency, and stigma to determine if the concentrations of risk in the current models are sufficiently high to warrant use in targeting preventive interventions.

Acknowledgments

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

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

Dr. Rosellini, Mrs. Sampson, and Dr. Kessler were responsible for the design of the study reported here. Dr. Stein, Dr. Ursano, and Dr. Kessler were responsible for the design of the Army Study to Assess Risk and Resilience in Servicemembers. Dr. Rosellini, Dr. Monahan, Dr. Street, Dr. Benedek, Dr. Bliese, Dr. Stein, Dr. Ursano, and Dr. Kessler formulated the analysis plan. Dr. Petukhova, Mrs. Sampson, and Dr. Kessler were responsible for data analysis. All authors were involved in analysis interpretation. Dr. Rosellini and Dr. Kessler wrote the first draft of the paper. All authors provided important intellectual content feedback on the first draft of the manuscript. All authors contributed to and have approved the final manuscript.

Dr. Monahan is a co-owner of the Classification of Violence Risk (COVR), Inc. Dr. Stein has been a consultant for Care Management Technologies, received payment for his editorial work from UpToDate and Depression and Anxiety, and had research support for pharmacological imaging studies from Janssen. Dr. Kessler received support for his epidemiological studies from Sanofi Aventis; was a consultant for Johnson & Johnson Wellness and Prevention, Shire, Takeda; and served on an advisory board for the Johnson & Johnson Services Inc. Lake Nona Life Project. Dr. Kessler is a co-owner of DataStat, Inc., a market research firm that carries out healthcare research.

References

1. Department of the U.S. Army. [Accessed October 31, 2016] Army 2020: generating health & discipline in the force ahead of the strategic reset. <https://www.army.mil/e2/c/downloads/232541.pdf>. Published January 19, 2012
2. Black, MC., Merrick, MT. [Accessed May 5, 2017] Prevalence of intimate partner violence, sexual violence, and stalking among active duty women and wives of active duty men – comparisons with women in the U.S. general population. 2010. http://www.sapr.mil/public/docs/research/2010_National_Intimate_Partner_and_Sexual_Violence_Survey-Technical_Report.pdf. Published March 2013
3. United States Department of Defense, Sexual Assault Prevention and Response Office. [Accessed May 5, 2017] Department of Defense annual report on sexual assault in the military – fiscal year 2015. http://sapr.mil/public/docs/reports/FY15_Annual/FY15_Annual_Report_on_Sexual_Assault_in_the_Military_Full_Report.pdf. Published May 5, 2016
4. United States Department of Defense. [Accessed October 31, 2016] 2014–2016 sexual assault prevention strategy. http://sapr.mil/public/docs/prevention/DoD_SAPR_Prevention_Strategy_2014-2016.pdf. Published April 30, 2014
5. Foshee VA, Bauman KE, Ennett ST, et al. Assessing the long-term effects of the Safe Dates program and a booster in preventing and reducing adolescent dating violence victimization and perpetration. *Am J Public Health*. 2004; 94(4):619–624. [PubMed: 15054015]
6. Taylor BG, Stein ND, Mumford EA, Woods D. Shifting Boundaries: an experimental evaluation of a dating violence prevention program in middle schools. *Prev Sci*. 2013; 14(1):64–76. [PubMed: 23076726]
7. Foster EM, Jones D. Can a costly intervention be cost-effective?: an analysis of violence prevention. *Arch Gen Psychiatry*. 2006; 63(11):1284–1291. [PubMed: 17088509]
8. Golubnitschaja O, Costigliola V. General report & recommendations in predictive, preventive and personalised medicine 2012: white paper of the European Association for Predictive, Preventive and Personalised Medicine. *EPMA J*. 2012; 3(1):1–53.
9. Seto MC, Lalumière ML. What is so special about male adolescent sexual offending? A review and test of explanations through meta-analysis. *Psychol Bull*. 2010; 136(4):526–575. [PubMed: 20565168]
10. Tharp AT, DeGue S, Valle LA, et al. A systematic qualitative review of risk and protective factors for sexual violence perpetration. *Trauma Violence Abuse*. 2013; 14(2):133–167. [PubMed: 23275472]
11. Whitaker DJ, Le B, Karl Hanson R, et al. Risk factors for the perpetration of child sexual abuse: a review and meta-analysis. *Child Abuse Negl*. 2008; 32(5):529–548. [PubMed: 18513795]
12. McWhorter SK, Stander VA, Merrill LL, Thomsen CJ, Milner JS. Reports of rape perpetration by newly enlisted male Navy personnel. *Violence Vict*. 2009; 24(2):204–218. [PubMed: 19459400]
13. Stander V, Thomsen C, Merrill L, Milner J. Longitudinal assessment of risk factors for military sexual aggression among enlisted male Navy personnel. *Military Psychology*. In press.
14. Hanson RK, Morton-Bourgon KE. The accuracy of recidivism risk assessments for sexual offenders: a meta-analysis of 118 prediction studies. *Psychol Assess*. 2009; 21(1):1–21. [PubMed: 19290762]
15. Harris, GT., Rice, M., Quinsey, VL., Cormier, CA. Violent offenders: appraising and managing risk. 3. Washington, DC: American Psychological Association; 2015.
16. Street AE, Rosellini AJ, Ursano RJ, et al. Developing a risk model to target high-risk preventive interventions for sexual assault victimization among female U.S. Army Soldiers. *Clin Psychol Sci*. 2016; 4(6):939–956. [PubMed: 28154788]
17. Rosellini AJ, Monahan J, Street AE, et al. Predicting non-familial major physical violent crime perpetration in the US Army from administrative data. *Psychol Med*. 2016; 46(2):303–316. [PubMed: 26436603]

18. Rosellini AJ, Monahan J, Street AE, et al. Using administrative data to identify U.S. Army soldiers at high-risk of perpetrating minor violent crimes. *J Psychiatr Res.* 2017; 84:128–136. [PubMed: 27741501]
19. Proulx, J., Beaugard, E., Lussier, P., Leclerc, B. *Pathways to sexual aggression.* New York: Routledge; 2014.
20. Ursano RJ, Colpe LJ, Heeringa SG, et al. The Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *Psychiatry.* 2014; 77(2):107–119. [PubMed: 24865195]
21. 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. [PubMed: 24318217]
22. Willett JB, Singer JD. Investigating onset, cessation, relapse, and recovery: why you should, and how you can, use discrete-time survival analysis to examine event occurrence. *J Consult Clin Psychol.* 1993; 61(6):952–965. [PubMed: 8113496]
23. Schlesselman, JJ. *Case-control studies: design, conduct, analysis.* New York: Oxford University Press; 1982.
24. United States Department of Justice, Office of Justice Programs, Bureau of Justice Statistics. [Accessed February 24, 2016] National Corrections Reporting Program, 2009 (ICPSR 30799). <http://www.icpsr.umich.edu/icpsrweb/NACJD/studies/30799?archive=NACJD&permit%5B0%5D=AVAILABLE&q=30799&x=0&y=0>. Published May 6, 2011. Updated February 11, 2013
25. Skeem JL, Kenealy P, Monahan J, Peterson J, Appelbaum P. Psychosis uncommonly and inconsistently precedes violence among high-risk individuals. *Clin Psychol Sci.* 2015; 4(1):40–49.
26. Steadman HJ, Monahan J, Pinals DA, Vesselinov R, Robbins PC. Gun violence and victimization of strangers by persons with a mental illness: data from the MacArthur Violence Risk Assessment Study. *Psychiatr Serv.* 2015; 66(11):1238–1241. [PubMed: 26073414]
27. Elbogen EB, Cueva M, Wagner HR, et al. Screening for violence risk in military veterans: predictive validity of a brief clinical tool. *Am J Psychiatry.* 2014; 171(7):749–757. [PubMed: 24832765]
28. Elbogen EB, Johnson SC, Newton VM, et al. Self-report and longitudinal predictors of violence in Iraq and Afghanistan war era veterans. *J Nerv Ment Dis.* 2013; 201(10):872–876. [PubMed: 24080674]
29. Elbogen EB, Johnson SC, Wagner HR, et al. Protective factors and risk modification of violence in Iraq and Afghanistan War veterans. *J Clin Psychiatry.* 2012; 73(6):767–773.
30. Elbogen EB, Johnson SC, Wagner HR, et al. Violent behaviour and post-traumatic stress disorder in US Iraq and Afghanistan veterans. *Br J Psychiatry.* 2014; 204(5):368–375. [PubMed: 24578444]
31. Elbogen EB, Wagner HR, Fuller SR, et al. Correlates of anger and hostility in Iraq and Afghanistan war veterans. *Am J Psychiatry.* 2010; 167(9):1051–1058. [PubMed: 20551162]
32. MacManus D, Dean K, Al Bakir M, et al. Violent behaviour in U.K. military personnel returning home after deployment. *Psychol Med.* 2012; 42(8):1663–1673. [PubMed: 22115074]
33. MacManus D, Dean K, Jones M, et al. Violent offending by UK military personnel deployed to Iraq and Afghanistan: a data linkage cohort study. *Lancet.* 2013; 381(9870):907–917. [PubMed: 23499041]
34. Sullivan CP, Elbogen EB. PTSD symptoms and family versus stranger violence in Iraq and Afghanistan veterans. *Law Hum Behav.* 2014; 38(1):1–9. [PubMed: 23646917]
35. Elbogen EB, Fuller S, Johnson SC, et al. Improving risk assessment of violence among military veterans: an evidence-based approach for clinical decision-making. *Clin Psychol Rev.* 2010; 30(6): 595–607. [PubMed: 20627387]
36. Kraemer HC. Events per person-time (incidence rate): a misleading statistic? *Stat Med.* 2009; 28(6):1028–1039. [PubMed: 19152356]
37. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Series B Stat Methodol.* 2005; 67(2):301–320.
38. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw.* 2010; 33(1):1–22. [PubMed: 20808728]
39. Berk, RA. *Statistical learning from a regression perspective.* New York: Springer; 2008.

40. Tully RJ, Chou S, Browne KD. A systematic review on the effectiveness of sex offender risk assessment tools in predicting sexual recidivism of adult male sex offenders. *Clin Psychol Rev.* 2013; 33(2):287–316. [PubMed: 23334335]
41. Singh JP. Predictive validity performance indicators in violence risk assessment: a methodological primer. *Behav Sci Law.* 2013; 31(1):8–22. [PubMed: 23408459]
42. Whittington R, Hockenhull JC, McGuire J, et al. A systematic review of risk assessment strategies for populations at high risk of engaging in violent behaviour: update 2002–8. *Health Technol Assess.* 2013; 17(50):1–128.
43. Hosmer, DW., Lemeshow, S., Sturdivant, RX. *Applied logistic regression.* 3. New Jersey: John Wiley & Sons; 2013.
44. Hanson RK, Thornton D, Helmus LM, Babchishin KM. What sexual recidivism rates are associated with Static-99R and Static-2002R scores? *Sex Abuse.* 2016; 28(3):218–252. [PubMed: 25810478]
45. Lee SC, Hanson RK. Recidivism risk factors are correlated with a history of psychiatric hospitalization among sex offenders. *Psychol Serv.* 2016; 13(3):261–271. [PubMed: 27504645]
46. Mengeling MA, Booth BM, Torner JC, Sadler AG. Reporting sexual assault in the military: who reports and why most servicewomen don't. *Am J Prev Med.* 2014; 47(1):17–25. [PubMed: 24854779]
47. Kolivas ED, Gross AM. Assessing sexual aggression: addressing the gap between rape victimization and perpetration prevalence rates. *Aggress Violent Behav.* 2007; 12(3):315–328.
48. DeGue S, Valle LA, Holt MK, Massetti GM, Matjasko JL, Tharp AT. A systematic review of primary prevention strategies for sexual violence perpetration. *Aggress Violent Behav.* 2014; 19(4):346–362.

Appendix

Appendix Table 1

Descriptions of administrative datasets included in the Army STARRS Historical Administrative Data Systems (HADS)

Database acronym	Description
AFMETS	ARMED FORCES MEDICAL EXAMINER TRACKING SYSTEM (AFMETS): Variables include manner of death and cause of death, including self-inflicted.
MDR	MEDICAL DATA REPOSITORY (MDR): Contains information about medical, dental, pharmaceutical, and ancillary claims data for both in network and purchased care as well as both inpatient and outpatient treatment. Data are collected on both Army personnel and their beneficiaries.
DCIPS	DEFENSE CASUALTY INFORMATION PROCESSING SYSTEM (DCIPS): Casualty data system for Army, Navy, Air Force, and Marines. It contains information on all casualties including deceased active duty members (and cause of death).
DMDC/Casualty	DEFENSE MANPOWER DATA CENTER (DMDC)/CASUALTY: Provides information on war deaths and active duty deaths, wounds, injuries, and illnesses.
DMDC/CTS	DEFENSE MANPOWER DATA CENTER (DMDC)/CONTINGENCY TRACKING SYSTEM (CTS): Collection of activation, mobilization, and deployment data. Provides information to DoD decision makers and includes a CTS Deployment File used for tracking the location of deployed personnel.
DMDC/Master Personnel & DMDC/Transaction files (2 distinct datasets)	DEFENSE MANPOWER DATA CENTER (DMDC)/MASTER PERSONNEL & TRANSACTION FILES: The Active Duty Master File provides an inventory of all individuals on active duty (excluding reservists on active duty for training) at a point in time. It is a standardized and centralized database of present and past members of the active duty force. Personal data elements include social security number, education level, home of record, date of birth, marital status, number of dependents, race, ethnic group, and name. Military data elements include Service, pay grade, Armed Forces Qualification Test percentile (enlisted only), source of commission (officers only), military primary duty and secondary occupation, Unit Identification Code, months of service, duty location, Estimated Termination of Service date, basic active service date, date of current rank, pay entry base date, foreign language ability, and major command code.

Database acronym	Description
DODSER	DEPARTMENT OF DEFENSE SUICIDE EVENT REPORT (DODSER): Provides risk and protective factor information for suicide events. This file will contain unsuccessful attempts and completed suicide cases.
TMDS	THEATER MEDICAL DATA STORE (TMDS): Used to track, analyze, view and manage Soldier medical treatment information recorded on the battlefield. Features include accessibility and visibility of service members' deployed medical records, outpatient and inpatient treatment records created in theater facilities, treatment records from other applications, reports on movement of patients, and patient status and injury/illnesses.
DMDC/DEERS	DEFENSE MANPOWER DATA CENTER (DMDC)/DEFENSE ENROLLMENT ELIGIBILITY REPORTING SYSTEM (DEERS): A DoD file containing personnel data and data related to DoD affiliation, benefits, employment, and pay.
MEDPROS/Readiness	MEDICAL PROTECTION SYSTEM (MEDPROS)/READINESS: The comprehensive Medical Readiness data includes all medical and dental readiness requirements IAW AR 600-8-101. They include immunizations, permanent physical profiles/duty limitations, eyeglasses/inserts, blood type, medical warning tags, personal deployment meds, pregnancy screening, DNA, HIV and dental status among other data elements.
DMDC/Perstempo	DEFENSE MANPOWER DATA CENTER (DMDC)/PERSONNEL TEMPO: Provides information on Soldier movement. Perstempo is a Congressionally mandated definition of time in which servicemembers engaged in official duties spend off-duty time away from housing associated with their garrison duty (time away from home station). The definition also applies to contingency operations such as Operation Iraqi Freedom (OIF) and Operation Enduring Freedom (OEF).
ACR ^a	ARMY CENTRAL REGISTRY (ACR): A victim-based registry documenting spouse and child abuse events involving Army service members.
AWD	ARMY WAIVER DATA (AWD): Information on waivers granted to new recruits to join the Army. The waivers are provided for four primary categories: 1) dependence (married with over two children), 2) conduct, 3) drug and alcohol, and 4) medical.
DAMIS	DRUG AND ALCOHOL MANAGEMENT INFORMATION SYSTEM (DAMIS): Includes data from the Army Drug And Alcohol Prevention Training program and Army Substance Abuse program.
DMDC/MEPCOM	DEFENSE MANPOWER DATA CENTER (DMDC)/MILITARY ENTRANCE PROCESSING COMMAND (MEPCOM): Contains transaction records on all individuals processed and examined at Military Entrance Processing Stations (MEPS) since July 1970. Examination data prior to FY 1976 is incomplete. Data elements in the USMEPCOM files are divided into three groups: Personal, Medical, and Administrative. The primary user is the Office of the Undersecretary of Defense (Personnel and Readiness) where Armed Services Vocational Aptitude Battery (ASVAB) data provides the basis for aptitude-related studies.
DMDC/Payroll	DEFENSE MANPOWER DATA CENTER (DMDC)/Payroll: Contains elements with information such as demographics, special and incentive pays, basic pay and allowances. This file is also used to answer pay-related questions for the General Accounting Office, Military Services research centers, and other government agencies.
DTMS	DIGITAL TRAINING MANAGEMENT SYSTEM (DTMS): Contains information on training requirements and standards, training planned and scheduled by Army units, and training accomplished by individuals and units in the Army.
MEDPROS/DHA	MEDICAL PROTECTION SYSTEM (MEDPROS)/DEPLOYMENT HEALTH ASSESSMENT (DHA): The DHA cycle begins when a Soldier completes the DD Form 2795 (Pre-Deployment Health Assessment) prior to deployment. Soldiers complete the DD Form 2796 Post Deployment Health Assessment (PDHA) prior to redeployment/demobilization. This triggers their Post Deployment Health Readiness Assessment (PDHRA) eligibility. At 90 days after completion of the DD Form 2796, Soldiers become eligible to participate in a PDHRA screening. The targeted window for PDHRA is 90–180 days after deployment. Once the Soldier is screened, follow-up appointments and treatment will be scheduled as needed.
MEDPROS/PHA	MEDICAL PROTECTION SYSTEM (MEDPROS)/PERIODIC HEALTH ASSESSMENT (PHA): The PHA is an annual physical and consists of two parts. Part one is a self-assessment completed by the Soldier; Part two is the provider assessment. PHA results consist of the health assessment, height, weight, physical and psychiatric assessment (PULHES), and potential for deployability within six months. These items will be entered into MEDPROS at point of service. The PHA is intended to improve Individual Medical Readiness and to evaluate deployability.
RRPS	RISK AND REDUCTION PROGRAM SYSTEM (RRPS): The RRPS collects data at the unit level that may show that members of the unit are at a higher risk. The Risk Reduction Program collects the number of incidents of high risk behavior by battalion level or separate company.

Database acronym	Description
SADMS ^a	SEXUAL ASSAULT DATA MANAGEMENT SYSTEM (SADMS): Centralized repository of relevant data regarding the entire lifecycle of sexual assault cases, involving victims and/or alleged offenders.
TRAC2ES	TRANSCOM REGULATING AND COMMAND AND CONTROL EVACUATION SYSTEM (TRAC2ES): A tracking system for all medical transfers across the world for all DOD services.
ASMIS-R	ARMY SAFETY MANAGEMENT INFORMATION SYSTEM-REVISED (ASMIS-R): Contains incident information data. Components consist of the reporting, analysis, and tracking of data, as well as efforts to provide preventative measures via risk assessments. ReportIT provides for the collection, verification, and timely dissemination of incident investigation and safety inspection data. AnalyzeIT serves as the business intelligence suite designed to identify trends and problem areas via data mining techniques and ad hoc analysis. TrackIT is designed to monitoring recommended controls and logging hazards. PreventIT encompasses risk assessment and guidance tools for units, battalions, commanders, civilians, and contractors.
CIMS/AC12 ^a	CRIMINAL INVESTIGATION DIVISION INFORMATION MANAGEMENT SYSTEM (CIMS)/AUTOMATED CRIMINAL INVESTIGATION/CRIMINAL INTELLIGENCE (AC12): Event level crime record database. Separate system from ASCRC (see below); Entity IDs are associated with offenses.
CIMS/ASCRC ^a	CRIMINAL INVESTIGATION DIVISION INFORMATION MANAGEMENT SYSTEM (CIMS)/AUTOMATED SYSTEM CRIME RECORD CENTER (ASCRC): Receives, maintains, accounts for, disseminates information from, and disposes of Army crime records; retrieves and correlates data and statistics from the records and provides to authorized recipients; coordinates automation of crime records data and information; and serves as functional proponent for the electronic imaging of crime record.
COPS/VRS	CENTRALIZED OPERATIONS POLICE SUITE (COPS)/VEHICLE REGISTRATION SYSTEM (VRS): Provides information on privately owned vehicles of civilians and military personnel utilized at military installations.
PDCAPS	PHYSICAL DISABILITY CASE PROCESSING SYSTEM (PDCAPS): Contains information on fitness and applicable disability benefits of Soldiers with duty related impairments, including clinical information of all medical conditions the Soldier manifests to determine whether each meets medical retention standards.
ACMIS ^a	ARMY COURT MARTIAL INFORMATION SYSTEM (ACMIS): Contains data on Soldiers (including Officers) arraigned on court-martial charges at either a Special or General Court-Martial. It contains some basic information such as their rank, General Technical (GT) score (if available), social security numbers, and birthdays. The database includes such information as where the Soldier was tried, which General Court-Martial Jurisdiction convened the case, the offenses charged, how the accused plead, the findings in the case as well as the sentence (if any). It also contains information concerning cases required to proceed through the appellate court process and outcome in the appellate court system. It includes the location of the offenses and trial, and tracks the dates of certain events in the court-martial process. It does not contain information concerning whether or not an accused has committed suicide.
ANAM	AUTOMATED NEUROPSYCHOLOGICAL ASSESSMENT METRICS (ANAM): This is an Army-developed and Army-owned, objective, computer-based test of human performance and neuropsychological functioning used to assess the cognitive performance of all active duty, reserve, and National Guard Soldiers.
ATRRS	ARMY TRAINING AND REQUIREMENTS RESOURCE SYSTEM (ATRRS): This is a system of record for management of personnel input to training and is the repository for training requirements, programs, personnel data and training costs.
COPS/MPRS ^a	CENTRALIZED OPERATIONS POLICE SUITE (COPS)/MILITARY POLICE REPORTING SYSTEM (MPRS): Centralized database that contains subsystems supporting the Military Police Corps Army wide. Contains information included in law enforcement reports.
COPS/Violations	CENTRALIZED OPERATIONS POLICE SUITE (COPS)/VIOLATIONS: Violation information contained in MPRS system (e.g., incident, time of incident).
COPS/ACIS	CENTRALIZED OPERATIONS POLICE SUITE (COPS)/ARMY CORRECTIONAL INFORMATION SYSTEM (ACIS): Has information for the confinement facility on new prisoners, currently confined prisoners, departed prisoners, and a monthly summary report of confinement facility activity data.
SFT	SOLDIER FITNESS TRACKER (SFT): Formerly known as the Comprehensive Soldier Fitness Global Assessment Tool (CSF-GAT), this source provides baseline information on strength in four areas – emotional, social, spiritual and family. Also provides an opportunity to track self-development and growth in these areas over time. The Chief of Staff of the

Database acronym	Description
	Army is requiring all Soldiers to take the GAT by May 2010. The objective of this program is to enhance performance and build resilience.
EORS	EQUAL OPPORTUNITY REPORTING SYSTEM (EORS): Database is used to collect, record, and maintain racial, ethnic group, and gender data and statistics needed to support the Army EO Program, to include Affirmative Action Program reporting requirements. No health related data. Contains information such as rank, race, complaint type, category, gender, timeline case start to end, resolution.
ITAPDB	INTEGRATED TOTAL ARMY PERSONNEL DATABASE (ITAPDB): Consolidated human resource database providing the Army and DOD with a single repository for personnel information for all components of the Army.
WWAS	WOUNDED WARRIOR ACCOUNTABILITY SYSTEM (WWAS): Integrated applications that support the accurate, timely, and effective tracking and management of Warfighters in the Wounded Warrior Lifecycle.

^aDatabase was used to identify cases of sexual assault perpetration. Cases were de-duplicated by creating a hierarchy of the criminal justice data systems, identifying instances when a Soldier had a record of a single type of crime recorded across multiple data systems within a two day offense window, and removing the duplicated record from whichever data system was lower in the hierarchy. For example, if a Soldier had one record of sexual assault perpetration in the Army Court Martial Information System (ACMIS) with an offense date of May 1, 2006 and another record of minor violence perpetration in the Military Police Reporting System (MPRS) with an offense date of May 2, 2016, the MPRS record was considered a duplicate case and removed from the dataset.

Appendix Table 2

The National Corrections Reporting Program (NCRP) offenses used to identify criminal perpetration and victimization

I. Sexual Assault Violent Crime

A. Rape/Sodomy/Sexual Assault

Rape – Force

Rape – Statutory – No Force

Forcible Sodomy

Sexual Assault – Other

Lewd Act With Children

B. Family-Related Sexual Violence

Family-Related Sexual Offenses (*Flagged as intra-familial perpetration*)

II. Major Physical Violent Crime

A. Murder/Homicide/Manslaughter

Murder

Unspecified Homicide

Voluntary/Nonnegligent Manslaughter

Manslaughter – Non-Vehicular

B. Kidnapping

Kidnapping

C. Aggravated Arson

Aggravated Arson

D. Aggravated Assault

Aggravated Assault

Aggravated Assault On A Public Officer

Aggravated Assault Of Military Official

Violent Offenses, Other – Aggravated Assault

E. Family-Related Aggravated Assault

Family-Related Aggravated Assault

Child Abuse

F. Robbery

Armed Robbery

Unarmed Robbery

III. Minor Physical Violent Crime

A. Simple Assault

Simple Assault

Simple Assault On A Public Officer

Simple Assault Of Military Official

Violent Offenses, Other – Simple Assault

B. Family-Related Simple Assault (*Excluded from outcome in the current study*)

Family-Related Simple Assault

C. Other Physical Violence

Violent Offenses, Other – Violent Offenses

IV. Verbal Violent Crime

A. Blackmail/Extortion/Intimidation

Blackmail/Extortion/Intimidation

B. Rioting

Rioting

C. Harassment (Non-Sexual)

Invasion Of Privacy – Other Harassment

Other Public Order Offenses – Harassment

D. Family-Related Emotional Abuse

Family-Related Emotional Abuse

E. Sexual Verbal Violence

Morals/Decency – Sexual Harassment

Invasion Of Privacy – Sexual Harassment

V. Non-Violent Offenses**V-1. Reckless Endangerment**

Manslaughter – Vehicular

Hit/Run Driving

Hit/Run Driving – Property Damage

Driving While Intoxicated

Driving Under The Influence (DUI)

Driving Under The Influence (DUI) – Drugs

Violent Offenses, Other – Reckless Endangerment

V-2. Possession/Use/Other Violation Of Illegal Substance

Possession/Use – Heroin

Possession/Use – Cocaine Or Crack

Possession/Use – Other Controlled Substances

Possession/Use – Marijuana/Hashish

Possession/Use – Drug Unspecified
Heroin Violation – Offense Unspecified
Cocaine Or Crack Violation – Offense Unspecified
Controlled Substance – Offense Unspecified
Marijuana/Hashish Violation – Offense Unspecified
Drug Offenses – Violation/Drug Unspecified

V-3. Weapon Offense

Weapon Offense

V-4. Drunkenness/Vagrancy/Disorderly Conduct

Drunkenness/Vagrancy/Disorderly Conduct
Liquor Law Violations, Other

V-5. Other Non-Violent Offenses

- A. Burglary
 - Burglary
- B. Simple Arson
 - Simple Arson
- C. Forgery/Fraud
 - Forgery/Fraud
- D. Theft
 - Auto Theft
 - Grand Larceny – Theft Over \$200
 - Petty Larceny – Theft Under \$200
 - Larceny/Theft – Value Unknown
 - Embezzlement
 - Stolen Property – Receiving
 - Stolen Property – Trafficking
 - Unauthorized Use Of Vehicle
 - Public Order Offenses, Other – Theft
- E. Property Offenses
 - Destruction Of Property
 - Trespassing
 - Property Offenses – Other
- F. Commercialized Vice, Illegal Substances
 - Trafficking – Heroin
 - Trafficking – Cocaine Or Crack
 - Trafficking – Other Controlled Substances
 - Trafficking – Marijuana/Hashish
 - Trafficking – Drug Unspecified
- G. Commercialized Vice, Other (Other non-violent crime)
 - Commercialized Vice
 - Liquor Law Violations – Commercialized Vice
- H. Court-Ordered Violations
 - Escape From Custody

Flight To Avoid Prosecution
 Parole Violation
 Probation Violations
 Contempt Of Court
 Offenses Against Courts, Legislatures And Commissions
 Obstruction – Law Enforcement
 I. Traffic Offense – Minor
 Traffic Offenses – Minor
 J. Military Specific Non-Violent Offenses
 Disobey Army Orders
 AWOL/Desertion/Treason
 K. Family-Related Non-Violent Offenses
 Family-Related Offenses – Contempt Of Court
 Family-Related Offenses – Reckless Endangerment
 Family-Related Offenses – Domestic Disturbance
 Family-Related Offenses – Custody Offenses
 Family-Related Offenses – Morals/Decency – Other Moral/Decency Offense
 L. Other Non-Violent Offenses
 Violent Offenses, Other – Non-Violent Offenses
 Habitual Offender
 Morals/Decency – Other Moral/Decency Offense
 Immigration Violations
 Invasion Of Privacy – Other
 Contributing To The Delinquency Of A Minor
 All Other Public Order Offenses
 Bribery And Conflict Of Interest
 Juvenile Offenses
 Felony – Unspecified
 Misdemeanor Unspecified
 Other/Unknown

Appendix Table 3

Incidence/100,000 person-years of first founded sexual assault perpetration by victim characteristics and time-in-service^a

Months-in-service	Non-family sexual assault													
	Adult victims							Minor victims						
	Incidence/ 100,000 person-years			Distribution of crimes		Population distribution		Incidence/ 100,000 person-years			Distribution of crimes		Population distribution	
	Est.	(se)	(n) ^b	%	(se)	% ^c	(se)	Est.	(se)	(n) ^b	%	(se)	% ^c	(se)
0–12	257.0	9.4	(830)	17.9	0.6	12.2	0.1	53.3	4.3	(171)	12.4	0.9	12.1	0.3
13–24	246.2	9.6	(728)	15.7	0.5	11.2	0.1	63.5	4.9	(189)	13.7	0.9	11.2	0.2

Months-in-service	Non-family sexual assault													
	Adult victims							Minor victims						
	Incidence/ 100,000 person-years			Distribution of crimes		Population distribution		Incidence/ 100,000 person-years			Distribution of crimes		Population distribution	
	Est.	(se)	(n) ^b	%	(se)	% ^c	(se)	Est.	(se)	(n) ^b	%	(se)	% ^c	(se)
25–36	205.4	9.0	(573)	12.3	0.5	10.6	0.1	66.0	5.2	(180)	13.0	0.9	10.3	0.2
37–48	197.7	9.8	(447)	9.6	0.4	8.6	0.1	56.9	5.2	(130)	9.4	0.8	8.6	0.2
49–60	170.7	10.3	(297)	6.4	0.4	6.6	0.1	47.3	5.3	(84)	6.1	0.6	6.7	0.2
61–120	177.3	6.0	(958)	20.6	0.6	20.4	0.2	57.8	3.4	(314)	22.7	1.1	20.5	0.3
121+	100.2	3.6	(807)	17.4	0.6	30.5	0.2	39.1	2.3	(316)	22.8	1.1	30.5	0.3
Total	175.5	2.7	(4,640)	100.0	--	100.0	--	52.2	1.5	(1,384)	100.0	--	100.0	--
χ^2_6	492.7			--		--		48.2			--		--	
Intra-family sexual assault														
0–12	10.9	1.9	(37)	9.7	1.5	12.9	0.7	1.9	0.8	(6)	1.8	0.7	11.7	1.1
13–24	22.6	3.2	(61)	16.1	1.9	10.2	0.6	5.7	1.5	(18)	5.4	1.2	11.8	1.1
25–36	20.8	3.0	(58)	15.3	1.8	10.5	0.6	6.1	1.5	(19)	5.7	1.3	11.8	1.1
37–48	14.3	2.6	(35)	9.2	1.5	9.3	0.6	7.6	2.0	(17)	5.1	1.2	8.5	0.8
49–60	22.5	3.9	(41)	10.8	1.6	6.9	0.5	12.4	2.9	(23)	6.9	1.4	7.0	0.7
61–120	14.7	1.7	(84)	22.1	2.1	21.5	0.7	18.8	2.1	(98)	29.3	2.5	19.7	1.0
121+	8.4	1.1	(64)	16.8	1.9	28.8	0.7	19.6	1.7	(154)	46.0	2.7	29.6	1.1
Total	14.3	0.8	(380)	100.0	--	100.0	--	12.6	0.8	(335)	100.0	--	100.0	--
χ^2_6	43.3			--		--		91.7			--		--	

Boldface indicates statistical significance ($p < 0.05$).

^a A total of 6,723 founded first offenses of the four sexual assault perpetration outcomes occurred between January 1, 2004 and December 31, 2009 among the 821,807 men in active duty service over that time period.

^b n = number of first founded sexual assault perpetrations in the time interval represented by the row.

^c Percent of the total population person-months of 31,249,056 in the time interval represented by the row.

Expanded description of elastic net

Elastic net regression minimizes the problem of collinearity among predictors by explicitly penalizing over-fit with a mixing parameter penalty (MPP) $\lambda \{ \text{MPP} \times P_{\text{lasso}} + (1 - \text{MPP}) \times P_{\text{ridge}} \}$ that controls relative weighting between two types of penalties, the lasso penalty (P_{lasso}) and the ridge penalty (P_{ridge}). 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. The ridge penalty handles multicollinearity by shrinking all coefficients smoothly towards 0 but retains all variables in the model.¹ The elastic net approach of combining the ridge and lasso penalties has the advantage of yielding more stable and accurate estimates than either ridge or lasso alone while maintaining model parsimony.² The parameter λ controls the total amount of penalization.³

Elastic net was implemented with a range of fixed MPP values in the “training sample” (person-months from 2004–2007): 0.01, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 0.99. The coefficients estimated in those models were then applied to calculate predicted probabilities of the outcome for each person-month in the validation sample (person-months from 2008–2009) for each value of MPP. The final MPP value was then selected to maximize concentration of risk (COR) among the 5% of Soldiers with highest predicted risk in the validation sample.

The optimal MPP for each outcome was: non-family perpetration against adults (MPP=0.50), non-family perpetration against minors (MPP=0.60), intra-family perpetration against adults (MPP=0.30), intra-family perpetration against minors (MPP=0.20).

References

1. Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*. 1970; 12(1):55–67.
2. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Series B Stat Methodol*. 2005; 67(2):301–320.
3. Berk, RA. *Statistical Learning from a Regression Perspective*. New York, NY: Springer; 2008.

Figure 1. Concentration of risk within each ventile of predicted risk derived from the final models^a

^a Ventiles are 20 groups created by dividing the total sample into equally sized groups defined by rank order of predicted risk from the final models. Only the nine final models that had a concentration of risk in the top-ventile of risk at least three times the expected value are presented here.

Table 1

Performance of the final elastic net models

Assault type	AUC	Top-ventile concentration of risk, % ^a	Top-ventile incidence/100,000 person-years	Number of variables in final model
I. Non-family sexual assault perpetration against...				
Adult	0.71	20.2	707.1	(34)
Minor	0.68	16.2	166.9	(19)
II. Intra-family sexual assault perpetration against....				
Adult	0.73	34.2	98.0	(13)
Minor	0.83	65.1	154.9	(11)

Abbreviations: AUC, area under the receiver operating characteristic curve.

^aConcentration of risk refers to the percent of all observed occurrences of the outcome in a ventile or ventiles of the predicted risk distribution.

Final model coefficients (odds-ratios) for founded sexual assault perpetration against non-family adults and minors^a

Table 2

Predictor	Adult victims		Minor victims	
	%	(se)	%	(se)
I. Socio-demographics ^b				
Marital status – Previously or never married	39.8	0.2	1.1	--
Education – Less than college	75.3	0.2	1.2	75.1 0.3
Race/ethnicity – Non-Hispanic Black	18.2	0.2	1.2	--
Race/ethnicity – Non-White	34.7	0.2	1.1	--
II. Historical ^b				
a. Army career				
Command in past year – FORSCOM	--	--	--	47.4 0.4
Command in past year – Non-TRADOC	--	--	--	85.7 0.3
Command in past year – Area-based component	18.0	0.2	1.2	--
Previously or never deployed	76.2	0.2	1.5	76.2 0.3
Returned from deployment 7+ months ago	64.5	0.2	1.1	--
MOS – Combat support or combat service support	59.1	0.2	1.1	--
Rank – Junior enlisted (E1–E4)	45.0	0.2	1.0	45.0 0.4
Rank – Non-officer	83.8	0.2	1.3	83.8 0.3
Less than 5 years in service	42.5	0.2	1.1	--
Demotion past 12 months	2.8	0.1	1.2	--
b. Prior crime perpetration				
Minor violence past 12 months (Count) ^c	0.7	0.0	1.4	--
Verbal violence past 12 months	0.2	0.0	1.5	--
Possession or use of illegal substance past 12 months	0.8	0.0	1.3	--
Non-violent crime past 12 months (Count) ^c	3.4	0.1	1.1	--
Any crime past 12 months	5.2	0.1	1.1	5.6 0.2
Any crime past 24 months	8.2	0.1	1.4	8.7 0.2
III. Clinical ^b				
a. Outpatient treatment of...				

Predictor	Adult victims		Minor victims	
	%	(se)	%	OR (se)
Any pain past 12 months	--	--	46.9	0.4 1.1
Any physical disorder past 12 months	92.8	0.1 1.7	--	--
Any physical disorder past 12 months (Cat.) ^d	--	--	93.7	0.2 1.0
Adjustment disorder past 3 months	0.6	0.0 1.4	--	--
Depression past 3 months	--	--	1.6	0.1 1.3
Tobacco use past 12 months	8.2	0.1 1.0	--	--
Any substance disorder past 3 months	--	--	1.1	0.1 1.3
Other impulse control disorder past 3 months (Cat.) ^d	0.0	0.0 1.9	--	--
Any impulse control disorder past 3 months	--	--	0.7	0.1 1.7
Any mental disorder past 3 months	7.6	0.1 1.3	--	--
Any mental disorder past 12 months	20.0	0.2 1.0	20.7	0.3 1.1
b. Inpatient treatment of...				
Suicide ideation past 12 months	--	--	0.0	0.0 3.0
Injury or poisoning past 12 months	--	--	0.6	0.1 2.0
Tobacco use past 3 months	0.1	0.0 1.5	--	--
Alcohol disorder past 3 months	0.1	0.0 1.8	--	--
Personality disorders past 3 months	--	--	0.0	0.0 7.6
Drug-induced mental illness past 3 months	--	--	0.0	0.0 8.2
Insomnia past 3 months	0.0	0.0 2.6	--	--
Any impulse control disorder past 3 months	--	--	0.0	0.0 3.2
Any mental disorder past 3 months	0.3	0.0 1.2	--	--
c. Medications				
Antidepressant (SSRI) use past 3 months (Cat.) ^d	1.4	0.1 1.3	--	--
Antidepressant (Sedative) use past 3 months (Cat.) ^d	0.4	0.0 1.2	--	--
Narcotic or alcohol treatment agents past 12 months	0.1	0.0 1.6	--	--
Analgesic NSAID use past 3 months (Cat.) ^d	13.7	0.2 1.2	--	--
Analgesic or antipyretic use past 12 months (Cat.) ^d	14.6	0.2 1.1	--	--
Any analgesic use past 12 months (Cat.) ^d	48.2	0.2 1.0	--	--

Abbreviations: OR, odds ratio; FORSCOM, forces command; TRADOC, training and doctrine command; MOS, military occupational specialty; Count, count variable; Cat, categorical variable; SSRI, selective serotonin reuptake inhibitor; NSAID, nonsteroidal anti-inflammatory drug.

^aDistinct case-control samples were used to develop the models for the two non-family sexual assault perpetration outcomes. The analysis samples included all person-months with the outcome plus a probability sample of all other person-months in the population stratified by sex and marital status. All records in the control sample were weighted by the inverse of probability of selection.

^bVariables were coded dichotomously (yes/no) unless otherwise noted.

^cCount variables for prior crime perpetration reflect the total number of crimes of the specified type and over the specified time period. The percentages reported for these variables reflect the proportion of the sample with non-zero values.

^dCategorical variables for treatment visits and prescriptions reflect the total number of visits (coded 0-4; 0=0 visits; 1=1-2 visits; 2=3-5 visits; 3=6-10 visits; 4=11+ visits) and prescriptions (coded 0-3; 0=never prescribed; 1=prescribed once; 2=prescribed twice; 3=prescribed 3+ times) of the specified type and over the specified time period. The percentages reported for these variables reflect the proportion of the sample with non-zero values.

Table 3

Final model coefficients (odds-ratios) for founded sexual assault perpetration against intra-family adults and minors^a

Predictor	Adult victims		Minor victims	
	%	(se)	%	(se)
I. Historical ^b				
a. Army career				
Previously or never deployed	76.3	0.8	1.5	--
b. Prior crime (perpetration)				
Major physical violence past 12 months	0.8	0.1	1.3	0.8
Major physical violence past 24 months	1.4	0.2	1.4	1.2
Non-family sexual assault past 12 months	--	--	--	0.1
Non-family sexual assault past 24 months	--	--	--	0.2
Verbal violence past 12 months	0.2	0.1	1.6	--
Non-violent crime past 24 months (Count) ^c	6.7	0.5	1.2	--
Any crime past 12 months	5.7	0.4	1.6	0.8
Any crime past 24 months	9.3	0.5	1.3	10.0
II. Clinical ^b				
a. Outpatient treatment of...				
Any physical disorder past 3 months (Cat.) ^d	55.0	0.9	1.2	52.3
A prior mental disorder past 3 months (Cat.) ^{d,e}	0.2	0.1	1.8	0.7
Non-dependent drug abuse past 12 months (Cat.) ^d	0.6	0.2	1.4	--
Marital problems past 3 months	0.9	0.1	1.3	--
Stressors/adversities past 3 months	--	--	--	1.0
Any stressor/stress disorder past 12 months	--	--	--	8.8
Any mental disorder past 3 months ^e	8.1	0.5	1.3	--
Any mental disorder past 3 months (Cat.) ^{d,e}	--	--	--	8.5
b. Medications				
Antidepressant (SNRI) use past 12 months	0.4	0.1	3.1	--

Abbreviations: OR, odds ratio; Count, count variable; Cat, categorical variable; SNRI, serotonin-norepinephrine reuptake inhibitor.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

^fDistinct case-control samples were used to develop the models for the two intra-family sexual assault perpetration outcomes. The analysis samples included all person-months with the outcome plus a probability sample of all other person-months in the population stratified by sex and marital status. All records in the control sample were weighted by the inverse of probability of selection.

^gVariables were coded dichotomously (yes/no) unless otherwise noted.

^cCount variables for prior crime perpetration reflect the total number of crimes of the specified type and over the specified time period. The percentages reported for these variables reflect the proportion of the sample with non-zero values.

^dCategorical variables for treatment visits and prescriptions reflect the total number of visits (coded 0–4; 0=0 visits; 1=1–2 visits; 2=3–5 visits; 3=6–10 visits; 4=11+ visits) and prescriptions (coded 0–3; 0=never prescribed; 1=prescribed once; 2=prescribed twice; 3=prescribed 3+ times) of the specified type and over the specified time period. The percentages reported for these variables reflect the proportion of the sample with non-zero values.

^ePrior mental disorder refers to International Classification of Diseases, Ninth Revision Clinical Modification (ICD-9-CM) V-codes capturing a personal history of mental disorder(s) (e.g. V11.1.1 A personal history of affective disorders).