

Laura Sampson (Orcid ID: 0000-0002-1715-2081)

Word count including abstract, text, and references: 4,939

Number of figures: 4

A machine learning approach to predicting new-onset depression in a military population.

Authors: Laura Sampson, PhD,^{1,2} Tammy Jiang, MPH,¹ Jaimie L. Gradus, DMSc, DSc,¹ Howard J. Cabral, PhD,³ Anthony J. Rosellini, PhD,⁴ Joseph R. Calabrese, MD,⁵ Gregory H. Cohen, PhD,^{1,6} David S. Fink, PhD,⁶ Anthony P. King, PhD,⁷ Israel Liberzon, MD,⁸ Sandro Galea, MD.^{1,9}

¹ Department of Epidemiology, Boston University School of Public Health, Boston, MA. Dr. Sampson, Ms. Jiang, Dr. Gradus, and Dr. Cohen report no financial relationships with commercial interests.

² Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA.

³ Department of Biostatistics, Boston University School of Public Health, Boston, MA. Dr. Cabral reports no financial relationships with commercial interests.

⁴ Center for Anxiety and Related Disorders, Department of Psychological and Brain Science, Boston University, Boston, MA. Dr. Rosellini reports no financial relationships with commercial interests.

⁵ Department of Psychiatry, School of Medicine, Case Western Reserve University, Cleveland, OH. Dr. Calabrese reports no financial relationships with commercial interests.

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1176/rcp2.1023](#).

This article is protected by copyright. All rights reserved.

⁶ Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY. Dr. Fink reports no financial relationships with commercial interests.

⁷ Department of Psychiatry, University of Michigan Medical School, Ann Arbor, MI. Dr. King reports no financial relationships with commercial interests.

⁸ Department of Psychiatry, College of Medicine, Texas A&M, College Station, TX. Dr. Liberzon reports no financial relationships with commercial interests.

⁹ Office of the Dean, Boston University School of Public Health, 715 Albany Street, Boston, MA. Dr. Galea reports no financial relationships with commercial interests.

Previous presentation: A previous version of one part of this analysis was presented as a poster at the Society of Epidemiologic Research Annual Meeting, June 18-21, 2019.

Correspondence and address for reprints:

Laura Sampson, PhD

Department of Epidemiology

Harvard T.H. Chan School of Public Health

677 Huntington Avenue

Boston, MA 02115

Telephone: 617-939-4493

Lasampson@hsph.harvard.edu

Conflict of interest statement: No authors have any conflicts of interest to disclose.

Acknowledgements: This work was supported by the Department of Defense [W81XWH-15-1-0080]. Ms. Jiang and Dr. Gradus were supported by the National Institute of Mental Health [R01MH109507].

Abstract:

Objectives: Depression is one of the most common mental disorders in the United States in both civilian and military populations, but few prospective studies assess a wide range of predictors across multiple domains for new-onset (incident) depression in adulthood. Supervised machine learning methods can identify predictors of incident depression out of many different candidate variables, without some of the assumptions and constraints that underlie traditional regression analyses. The objectives of this study were to identify predictors of incident depression across five years of follow-up using machine learning, and to assess prediction accuracy of the algorithms. **Method:** Data were from a cohort of Army National Guard members free of history of depression at baseline (n = 1951 men and 298 women), interviewed once per year for probable depression. Classification trees and random forests were constructed and cross-validated, using 84 candidate predictors from the baseline interviews. **Results:** Stressors and traumas such as emotional mistreatment and adverse childhood experiences, demographics such as being a parent or student, and military characteristics including paygrade and deployment location were predictive of probable depression. Cross-validated random forest algorithms were moderately accurate (68% for women and 73% for men). **Conclusions:** Events and characteristics throughout the life course, both in and outside of deployment, predict incident depression in

adulthood among military personnel. Although replication studies are needed, these results may help inform potential intervention targets to reduce depression incidence among military personnel. Future research should further refine and explore interactions between identified variables.

Key words: Depressive disorders, military psychiatry, machine learning, prediction, traumatic events

INTRODUCTION

Depressive disorders are among the most common mental disorders in both civilian and military populations in the United States (U.S.) (1,2). Major Depressive Disorder (MDD) was the second leading cause of disability in the U.S. in 2010 out of all medical conditions (3). Given this high burden, there is a need to identify characteristics of persons at high risk for developing depression, particularly in a military environment where soldiers may frequently deploy to high-stress situations and may be more feasible to monitor and intervene on compared to most civilian populations.

The broad goal of supervised machine learning is optimized prediction, and it comprises algorithmic, data-driven approaches that can handle large numbers of predictor variables (4,5). In particular, classification tree and random forest classifiers construct nonparametric algorithms that promote visual inspection of the data and an understanding of complicated interactions and nonlinear associations that are more difficult to identify and interpret using other methods (4,6,7) or that would not otherwise be detected (8–10), allowing for identification of complex risk profiles without *a priori* hypotheses (7).

Among military populations, most studies investigating predictors of mental health problems have focused on military—and particularly deployment—experiences rather than a full range of characteristics and stressors occurring both in and outside of military service. A broader picture of risk is needed, particularly for part-time soldiers including the National Guard, who frequently transition between military and civilian life. Supervised machine learning methods can identify a wide array of factors associated with incident depression in this group.

Supervised learning has been used to predict psychiatric outcomes including suicide (11–13), posttraumatic stress disorder (PTSD) (14,15), depression in very specific groups (e.g.,

Author Manuscript
elderly populations; 16), comorbid depression among patients with chronic physical conditions (17), and depression treatment response in clinical samples (18–20). To our knowledge, classification trees and random forest have not yet been applied to predicting new-onset (incident) depression in a military population.

Our objectives for this study were to (a) use a range of potentially predictive characteristics and experiences from across the life course to discern which variables and their interactions predict incident depression, using classification tree and random forest algorithms, and (b) assess predictive accuracies of these algorithms using cross-validation, in a cohort of U.S. Army National Guard members.

METHOD

Data source

We used data from the Ohio Army National Guard Mental Health Initiative (OHARNG-MHI), an ongoing cohort study that began in 2008-2009. Details of recruitment are described elsewhere (21). This cohort—and the Ohio Army National Guard in general—is representative of the U.S. Army National Guard population as a whole in terms of many demographic and social factors such as military rank, gender, and age (21,22).

The first and primary cohort of the study (n=2,616 participants at baseline) completed telephone interviews approximately once per year for six years. The baseline interview assessed demographics, mental health disorders, military experiences, and potentially traumatic life events (“traumas,” e.g., major accidents, abuse) that occurred throughout the life course, whereas the follow-up interviews primarily assessed past-year events. In order to mitigate loss of sample size over time and related changes in demographics due to attrition, smaller samples of newer recruits

to the Guard replenished the original group of respondents each year, beginning in the third year of the study, creating a dynamic cohort study design (23). The analytic sample (1,951 men and 298 women) included respondents from OHARNG-MHI who were present for at least one follow-up interview and had no history of depression at baseline (their first interview, regardless of the calendar year of entry into the study).

The Ohio National Guard and the institutional review boards of University Hospitals Case Medical Center, University of Toledo, University of Michigan, Ann Arbor Veterans Administration Medical Center, Columbia University, Boston University Medical Campus, and the Office of Human Research Protections of the US Army Medical Research and Materiel Command approved this study protocol. Respondents provided verbal informed consent after receiving a complete description of the study.

Outcome

Probable depression (henceforth referred to as “depression”) was measured with the Patient Health Questionnaire (PHQ-9) (24) and classified according to the Diagnostic and Statistical Manual for Mental Disorders version IV (DSM-IV) criteria. The construct was validated as part of the parent study, using a Structured Clinical Interview for DSM-IV-TR (SCID) Axis I Disorders (non-patient version) in-person interview among a random subsample of 500 members of the original cohort (25). Any Depressive Disorder, which includes the DSM-IV categories of MDD and Other Depressive Disorder, was used to define depression in this study due to higher sensitivity compared to MDD only (51% vs. 35%), without sacrificing specificity (83%), when validated against the in-person psychiatric interviews.

This definition corresponds to reporting a period of at least two weeks in the past year

with two or more co-occurring symptoms, where one of the symptoms is depressed mood or anhedonia (inability to feel pleasure), with a frequency of “more than half the days” or “nearly every day.” Having thoughts of self-harm or suicide is an exception to the frequency criteria, counting as a symptom when reported at any frequency.

Incident depression was established by collapsing up to five years of follow-up data into one binary measure for each participant, to represent whether the individual had new-onset depression at any point during their follow-up.

Predictors

In order to preserve temporality, all potential predictors were collected from the baseline interviews, with the exception of four adverse childhood experiences (ACEs) which were added to the study in the second year for the original cohort (but which were assessed at baseline for the following three cohorts).

Our set of *a priori* predictors included all questions or constructs (i.e., variables created from groups of questions or symptoms) from the baseline surveys, as long as the variables had at least five respondents per cell (category). There were 84 total potential predictors for men and 72 for women; women had fewer potential predictors due to their smaller sample (variables with less than five individuals per cell were either removed, or where possible, categories were combined).

These potential predictors included 12 demographic variables, 7 military characteristic variables (e.g., rank), 8 health-related variables (e.g., other mental health disorders, substance use), 2 social support variables, 8 general life stressors (e.g., financial problems), and 47 traumas, including ACEs and also more recent traumas (e.g., witnessing death), both in and

outside of deployment. All potential predictors and their prevalence in the sample are listed in eTable 1 (appendix), and details of the predictors are provided in the text of the appendix.

Statistical analysis

All analyses were gender-stratified due to known differences in correlates of depression among men and women (26,27). We first ran single classification trees for each gender-specific sample using the *partykit* package in R, specifying no random variable selection at each node, minimums of 20 observations per split and 10 observations for the terminal nodes, and stop criterion based on univariate p -values with a cutoff of $p < 0.01$. We plotted these trees in order to visually evaluate the data structure and identify key predictive variables and their interactions.

Next, we constructed 10-fold cross-validated random forests, which a) consolidate across multiple classification trees to add random variation and avoid overfitting to any particular subsample, and b) test and train the algorithms on different combinations of subsets of the sample. We used the *caret* and *RandomForest* packages with 1,000 trees (28), 5 predictor variables randomly sampled at each node, and minimums of 20 observations per split and 1 observation for the terminal nodes. As described in the appendix, we tuned the algorithms to sample only from a subset of data for each tree (90% of true cases and an equal number of non-cases), in order to adjust for the class imbalance in our sample (29,30). Tuning this parameter provided better sensitivity compared to algorithms calculated from default methods.

Using the cross-validated predicted values, the average area under the receiver operating characteristic curve (AUC) was calculated for each gender-specific algorithm, in addition to the average sensitivity, specificity, and accuracy (the overall proportion of correctly classified individuals). We assessed and plotted variable importance using average decrease in accuracy for

each variable, which represents the reduction in accuracy that would result if a variable were randomly permuted (29). All analyses are explained in detail in the appendix. For the tree classification, missing data was handled using surrogate splits (see appendix). For the cross-validated random forests, a complete-case analysis was performed, as surrogate splits cannot be used across folds.

RESULTS

Men

Incidence of depression over follow-up was 14.9% among men (other descriptive information is provided in the appendix including eTables). Figure 1 shows the single classification tree among men. Past-year PTSD was the most predictive of depression overall. Among men with past-year PTSD, having had casualties in the unit with which they were most recently deployed was the next most predictive variable. Among those without past-year PTSD, parental verbal abuse in childhood (one of the ACEs) was next-most important in predicting depression, and so on, down each branch. The combination of characteristics with the highest probability of incident depression was having both past-year PTSD and reporting a unit casualty during the most recent deployment (n=15, incidence=73.3%). The subgroup with the next highest incidence of depression (53.8%) included men who were parents or guardians of children under the age of 18 and who reported fair or poor general health compared to good or great health, but who reported no traumatic injuries/accidents (other than transportation accidents), no financial problems, no childhood verbal abuse, and no past-year PTSD (n=13).

Figure 2 shows a variable importance plot (of mean decreases in accuracy when each variable is removed) for the top 40 predictive variables among men from the cross-validated random forest (the values for all variables are listed in eTable 2). Reporting general life stressors (having been emotionally mistreated, financial problems, divorce); some demographic characteristics (being a current student, being a parent or guardian, being aged 35 or older); being deployed to a non-conflict area; and various traumas (including two ACEs) had the highest mean decreases in accuracy, meaning they were the top predictors. The cross-validated AUC and accuracy were 0.67 and 73.0%, respectively, with 46.8% sensitivity and 77.0% specificity when using the default threshold of predicted risk of 0.50.

Women

Incidence of depression over follow-up was 24.8% among women. Figure 3 shows the single classification tree. Given the small sample size, only one split of the data was made, for alcohol abuse: women with a history of alcohol abuse at baseline had a 42.5% incidence of depression, whereas those who never had alcohol abuse had a 22.1% incidence of depression.

Figure 4 depicts the variable importance plot for all predictive variables from the cross-validated random forest (these values are also listed in eTable 3). Life stressors (having a family member addicted to drugs or alcohol, having been mistreated); demographics (being Enlisted with a relatively low paygrade, a student, and aged 25 or older); having a close friend or family member seriously injured in an accident other than a car accident; low psychosocial support; and childhood verbal abuse were among the top predictors. The cross-validated AUC was 0.67 and the average accuracy was 68.1%, with 75.3% specificity and 45.9% sensitivity when using the default threshold of predicted risk of 0.50.

DISCUSSION

To our knowledge, this study is the first to use classification trees and random forests to assess predictors of probable incident depression in a military sample. We found that, among both men and women, traumas and ACEs (particularly verbal abuse by a parent or guardian), stressors such as being emotionally mistreated, and demographics such as being a current student were predictive of incident depression during follow-up. Military characteristics (e.g., paygrade), low psychosocial support, and hearing about traumas happening to friends or family (e.g., a friend was in a serious accident) appeared more predictive of depression for women than for men, whereas PTSD, deployment location, personally experienced traumas (including combat-related experiences), and financial problems appeared more predictive among men compared to women.

Among men, recent deployment to a non-conflict area was predictive of depression, compared to being deployed to either a conflict area (Iraq or Afghanistan) or never having been deployed. This may be due to stressful and unexpected domestic deployments to areas affected by natural disasters—which have been increasing in recent years—or to areas of civil unrest after riots or massive protests, which can involve National Guard deployment. These types of domestic deployment may be more distressing for soldiers than combat deployments overseas, because they can involve confronting fellow citizens (at protests that become violent, for example) or witnessing citizens suffer (in natural disaster contexts). This finding should be replicated, but it could indicate that additional resiliency training may be warranted for these unique deployment experiences. We were unable to compare incidence of depression by exact location or type of recent deployments, given small cells and lack of detailed questions on the surveys.

Author Manuscript

Past-year PTSD was the most predictive variable for incident depression among men in the single tree (and moderately predictive in the random forest, suggesting there may have been some overfitting in the single tree). This finding is broadly consistent with both with the only other study to use random forests to predict incident depression in a population-based sample (31) and with many non-machine learning studies that have consistently found comorbidity between PTSD and depression (32–35). The combination of having both past-year PTSD and reporting a unit casualty during the most recent deployment was particularly predictive of depression among men in the classification tree, for which the incidence of depression was 73.3%, or five times larger than the overall incidence of depression among men in this sample.

Among women in our study, PTSD was not predictive of incident depression, but lifetime PTSD status was included in the algorithm instead of past-year status, given the small number of women with PTSD in the past year in an already-small sample of women. This may be the reason why PTSD was not selected among women as being highly predictive, since history of PTSD may have occurred many years before onset of depression, and thus not as clinically or statistically relevant.

Our findings that ACEs and more recent traumas and stressors were predictive of incident depression (for both men and women) is also consistent with a prior machine learning study (31) as well as many non-machine-learning studies that have modeled incident or prevalent depression with similar types of events as exposures or predictors (36–38). Traumas and stressors such as being mistreated have long been known to associate with depression outcomes (39–42), particularly when they occur during childhood, while brain is still developing (42,43).

Finally, our findings on financial problems, being a student, being of lower paygrade, and having children may all be related to financial stress, debt, and concern about being able to

provide for one's family, which have been found in non-machine learning contexts to be associated with depression (44–47).

Based on ten-fold cross-validation, our random forest algorithms were moderately accurate overall (73% accuracy for men and 68% for women). These values are in line with other studies predicting depression outcomes; Kautzky and colleagues (18), who used random forests to predict treatment-resistant depression, found accuracies of 68-75%. Similarly, Jin and colleagues (17), who used four different prediction methods including random forests to model depression (also measured using the PHQ-9) among patients with diabetes, found comparable levels of accuracy (approximately 73%).

Limitations of our study include the use of baseline information alone to predict incident depression over follow-up. It follows that we lack (a) time-varying information assessed on the follow-up surveys that could be temporally closer to onset of depression compared to baseline variables, and (b) information on exact timing of prior events and experiences, as the baseline surveys primarily assessed events that occurred at some point in the past, without asking detailed information on timing (with the exception of other mental disorders). However, using only baseline predictors in this study established temporality between our predictors and outcome—a crucial aspect of valid prediction.

Another limitation is our use of the PHQ-9 for measuring depression. Although the PHQ-9 has been validated against a gold standard depression measure within this cohort as well as in many other populations (24,25), it is primarily a screening tool and was not designed as a diagnostic test. Thus, it is possible that there are individuals in this study with incorrectly classified depression status, which could have affected which variables were chosen as being

Author Manuscript

predictive. Future studies should aim to replicate these results using diagnostic measures of depression.

Finally, we used a complete case analysis for the cross-validated random forests. Missing data in this study stems primarily from the fact that ACEs were not asked on the baseline survey for the first (and largest) cohort of participants. For those individuals, the ACEs were assessed in the second wave of the study, at which not all respondents were present. A smaller portion of missing data came from responses of “don’t know” or declining to answer questions such as income. As this is a prediction study and thus we are not aiming to isolate and measure the effect of any particular variable on depression, missing data is not as problematic of an issue as in an explanatory study. Generally, missing data among predictors in prediction modeling is thought to only create bias if missingness is related to the outcome variable (9,48). We have no reason to believe that this is the case in our study, as all predictors are from the baseline interview, at which time the outcome had not yet occurred (with the exception of the four ACEs assessed at wave 2 for the primary cohort, which were missing by design, not by refusal to answer).

Despite these limitations, these results may help inform potential screening interventions for depression in this population. Algorithms represent concrete ways officials might identify characteristics associated with high risk of developing outcomes, regardless of underlying causal relationships; this might be especially useful in a military setting given that military personnel are feasible to monitor. For example, the REACH VET algorithm, built by researchers using machine learning, helped the U.S. Department of Veterans’ Affairs to identify veterans at high risk for suicide (49,50), as part of a crucial undertaking at a time when suicides among military personnel have been increasing.

Future analytic work that aims to predict depression—preferably using larger samples and more specifically timed predictors than we were able to utilize in this study—should aim to replicate our findings and further refine interactions between variables identified here. Machine learning might also be used to predict particular subtypes of depression, given that the overall disorder is heterogeneous and takes on different forms in different individuals; this may improve prediction accuracy. Predictive accuracies of the algorithms could also be compared with individual-level prediction using more traditional types of regressions, or using other types of machine learning algorithms, including ensemble methods such as Super Learner which average across different types of algorithms. Finally, broader environmental and context-level variables—like unit-level characteristics in a military study or residential neighborhood-level characteristics in a general population survey—may be important for prediction of individual incident depression (51,52), and should be included as predictors in future studies, where sampling designs allow.

REFERENCES

1. Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, et al. The Epidemiology of Major Depressive Disorder: Results from the National Comorbidity Survey Replication (NCS-R). *JAMA J Am Med Assoc.* 2003 Jun 18;289(23):3095.
2. Gadermann AM, Engel CC, Naifeh JA, Nock MK, Petukhova M, Santiago PN, et al. Prevalence of DSM-IV major depression among U.S. military personnel: meta-analysis and simulation. *Mil Med.* 2012 Aug;177(8 Suppl):47–59.
3. Murray CJL, Abraham J, Ali MK, Alvarado M, Atkinson C, Baddour LM, et al. The state of US health, 1990-2010. *JAMA J Am Med Assoc.* 2013 Aug 14;310(6):591.
4. Sutton CD. Classification and regression trees, bagging, and boosting. *Handb Stat.* 2005;24:303–29.
5. James G, Witten D, Hastie T, Tibshirani R. Tree-based methods. In: Casella G, Fienberg S, Olkin I, editors. *An Introduction to Statistical Learning with Applications in R.* New York, NY: Springer Texts in Statistics; 2013. p. 303–32.
6. Morgan J. Technical Report No. 1: Classification and regression tree analysis. 2014. Available from: <https://www.bu.edu/sph/files/2014/05/MorganCART.pdf>
7. Hastie, T., Tibshirani, R., & Friedman J. *The elements of statistical learning.* 1st ed. New York, NY: Springer; 2001.
8. Fayyad UM, Grinstein GG, Wierse A. *Information Visualization in Data Mining and Knowledge Discovery (The Morgan Kaufmann Series in Data Management Systems).* 1st ed. Burlington, Massachusetts: MK/Morgan Kaufmann Publishers; 2001.
9. Shmueli G. To Explain or to Predict? *Stat Sci.* 2010;25(3):289–310.

- Author Manuscript
10. Faghmous JH. Machine Learning. In: El-Sayed AM, Galea S, editors. System Science and Population Health. New York, NY: Oxford University Press; 2017. p. 129–37.
 11. Gradus JL, King MW, Galatzer-Levy I, Street AE. Gender differences in machine learning models of trauma and suicidal ideation in veterans of the Iraq and Afghanistan wars. *J Trauma Stress*. 2017 Aug;30(4):362–71.
 12. Kessler RC, Stein MB, Petukhova M V, Bliese P, Bossarte RM, Bromet EJ, et al. Predicting suicides after outpatient mental health visits in the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *Mol Psychiatry*. 2017 Apr 19;22(4):544–51.
 13. Gradus JL, Rosellini AJ, Horváth-Puhó E, Street AE, Galatzer-Levy I, Jiang T, et al. Prediction of sex-specific suicide risk using machine learning and single-payer health care registry data from Denmark. *JAMA Psychiatry*. 2020 Jan 1;77(1):25.
 14. Marinić I, Supek F, Kovacić Z, Rukavina L, Jendricko T, Kozarić-Kovacić D. Posttraumatic stress disorder: diagnostic data analysis by data mining methodology. *Croat Med J*. 2007 Apr;48(2):185–97.
 15. Karstoft K-I, Galatzer-Levy IR, Statnikov A, Li Z, Shalev AY, members of Jerusalem Trauma Outreach and Prevention Study (J-TOPS) group F members of the JTO and PS (J-T). Bridging a translational gap: using machine learning to improve the prediction of PTSD. *BMC Psychiatry*. 2015 Mar 16;15:30.
 16. Sau A, Bhakta I. Predicting anxiety and depression in elderly patients using machine learning technology. *Healthc Technol Lett*. 2017 Dec 1;4(6):238–43.
 17. Jin H, Wu S, Capua P Di. Development of a clinical forecasting model to predict comorbid depression among diabetes patients and an application in depression screening

- policy making. *Prev Chronic Dis*. 2015;12.
18. Kautzky A, Dold M, Bartova L, Spies M, Vanicek T, Souery D, et al. Refining prediction in treatment-resistant depression. *J Clin Psychiatry*. 2018 Feb 21;79(1):16m11385.
 19. Zhdanov A, Atluri S, Wong W, Vaghei Y, Daskalakis ZJ, Blumberger DM, et al. Use of machine learning for predicting escitalopram treatment outcome from electroencephalography recordings in adult patients with depression. *JAMA Netw Open*. 2020 Jan 3;3(1):e1918377.
 20. Pearson R, Pisner D, Meyer B, Shumake J, Beevers CG. A machine learning ensemble to predict treatment outcomes following an Internet intervention for depression. *Psychol Med*. 2018 Nov 5;1–12.
 21. Calabrese JR, Prescott M, Tamburrino M, Liberzon I, Slembariski R, Goldmann E, et al. PTSD comorbidity and suicidal ideation associated with PTSD within the Ohio Army National Guard. *J Clin Psychiatry*. 2011 Aug 15;72(08):1072–8.
 22. Department of Defense (United States). 2011 Demographics: Profile of the Military Community. Washington, D.C; 2012.
 23. Kalton G. Some issues in the design and analysis of longitudinal surveys. *Proc 59th ISI World Stat Congr* 25-30 August 2013, Hong Kong. 2013;2611–6.
 24. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001 Sep;16(9):606–13.
 25. Prescott MR, Tamburrino M, Calabrese JR, Liberzon I, Slembariski R, Shirley E, et al. Validation of lay-administered mental health assessments in a large Army National Guard cohort. *Int J Methods Psychiatr Res*. 2014 Mar 1;23(1):109–19.
 26. Kessler RC. The epidemiology of depression among women. In: Keyes CL., Goodman

- SH, editors. *Women and Depression: A Handbook for the Social, Behavioral, and Biomedical Sciences*. Cambridge, United Kingdom: Cambridge University Press; 2006. p. 22–40.
27. Oquendo MA, Turret J, Grunebaum MF, Burke AK, Poh E, Stevenson E, et al. Sex differences in clinical predictors of depression: a prospective study. *J Affect Disord*. 2013 Sep 25;150(3):1179–83.
28. Hastie T, Tibshirani R, Friedman J. Random Forests. In: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. New York, NY: Springer; 2013. p. 587–603.
29. Strobl C, Malley J, Tutz G. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol Methods*. 2009 Dec;14(4):323–48.
30. Chen C, Liaw A, Brieman L. Using random forest to learn imbalanced data: Technical Report No. 666. University of California, Berkley. Berkeley, CA; 2004 [cited 2019 Jul 14]. Available from: <https://statistics.berkeley.edu/tech-reports/666>
31. Rosellini AJ, Liu S, Anderson GN, Sbi S, Tung ES, Knyazhanskaya E. Developing algorithms to predict adult onset internalizing disorders: An ensemble learning approach. *J Psychiatr Res*. 2020 Feb 1;121:189–96.
32. Caramanica K, Brackbill RM, Liao T, Stellman SD. Comorbidity of 9/11-related PTSD and depression in the World Trade Center Health Registry 10-11 years postdisaster. *J Trauma Stress*. 2014 Dec;27(6):680–8.
33. Rytwinski NK, Scur MD, Feeny NC, Youngstrom EA. The co-occurrence of major depressive disorder among individuals with posttraumatic stress disorder: A meta-

- analysis. *J Trauma Stress*. 2013 Jun;26(3):299–309.
34. Plana-Ripoll O, Pedersen CB, Holtz Y, Benros ME, Dalsgaard S, de Jonge P, et al. Exploring comorbidity within mental disorders among a Danish national population. *JAMA Psychiatry*. 2019 Mar 1;76(3):259.
35. Kessler RC, Sonnega A, Bromet E, Hughes M, Nelson CB. Posttraumatic stress disorder in the National Comorbidity Survey. *Arch Gen Psychiatry*. 1995 Dec 1;52(12):1048.
36. Rudenstine S, Cohen G, Prescott M, Sampson L, Liberzon I, Tamburrino M, et al. Adverse childhood events and the risk for new-onset depression and post-traumatic stress disorder among U.S. National Guard Soldiers. *Mil Med*. 2015 Sep;180(9):972–8.
37. Kessler RC. The effects of stressful life events on depression. *Annu Rev Psychol*. 1997 Feb;48(1):191–214.
38. Hovens JGFM, Giltay EJ, Spinhoven P, van Hemert AM, Penninx BWJH. Impact of childhood life events and childhood trauma on the onset and recurrence of depressive and anxiety disorders. *J Clin Psychiatry*. 2015 Jul 22;76(07):931–8.
39. Muscatell KA, Slavich GM, Monroe SM, Gotlib IH. Stressful life events, chronic difficulties, and the symptoms of clinical depression. *J Nerv Ment Dis*. 2009 Mar;197(3):154–60.
40. Keller MC, Neale MC, Kendler KS. Association of different adverse life events with distinct patterns of depressive symptoms. *Am J Psychiatry*. 2007 Oct;164(10):1521–9.
41. Hammen C. Stress and depression. *Annu Rev Clin Psychol*. 2005;1:293–319.
42. Elliott GR, Eisdorfer C, editors. *Stress and human health: Analysis and implications of research. A study by the Institute of Medicine/National Academy of Sciences OM/NAS*. New York, NY: Springer; 1982.

43. Kaufman J, Plotsky PM, Nemeroff CB, Charney DS. Effects of early adverse experiences on brain structure and function: clinical implications. *Biol Psychiatry*. 2000 Oct 15;48(8):778–90.
44. Stansfeld SA, Head J, Marmot MG. Explaining social class differences in depression and well-being. *Soc Psychiatry Psychiatr Epidemiol*. 1998;33(1):1–9.
45. Ettman CK, Cohen GH, Galea S. Is wealth associated with depressive symptoms in the United States? *Ann Epidemiol*. 2020 Mar 1;43:25-31.e1.
46. Gathergood J. Debt and depression: Causal links and social norm effects. *Econ J*. 2012 Sep 1;122(563):1094–114.
47. Musliner KL, Munk-Olsen T, Eaton WW, Zandi PP. Heterogeneity in long-term trajectories of depressive symptoms: Patterns, predictors and outcomes. *J Affect Disord*. 2016 Mar 1;192:199–211.
48. Ding, Y. and Simonoff J. An investigation of missing data methods for classification trees applied to binary response data. *J Mach Learn Res*. 2010;11:131–170.
49. REACH VET, Recovery Engagement and Coordination for Health - Veterans Enhanced Treatment: Predictive Analytics for Suicide Prevention. Veterans Health Administration. 2017. Available from: [https://www.dspo.mil/Portals/113/Documents/2017_Conference/Presentations/REACH VET Predictive Modeling.pdf?ver=2017-08-10-132615-843](https://www.dspo.mil/Portals/113/Documents/2017_Conference/Presentations/REACH_VET_Predictive_Modeling.pdf?ver=2017-08-10-132615-843)
50. Ravindranath M. How the VA uses algorithms to predict suicide. *Politico*. 2019 [cited 2019 Aug 4]. Available from: <https://www.politico.com/story/2019/06/25/va-veterans-suicide-1382379>
51. Galea S, Keyes KM. Population health science and the challenges of prediction. *Ann*

Intern Med. 2017 Oct 3;167(7):511.

52. Helbich M, Hagenauer J, Roberts H. Relative importance of perceived physical and social neighborhood characteristics for depression: a machine learning approach. *Soc Psychiatry Psychiatr Epidemiol.* 2020 May 14;55(5):599–610.

FIGURE LEGENDS AND FOOTNOTES:

Figure 1. Classification tree for incident depression during follow-up among men (n = 1,951).

Figure 1 FOOTNOTES:

^aDuring most recent deployment.

PTSD = posttraumatic stress disorder.

In grey boxes: n = number of individuals with selected combination of predictors; “no” = proportion without incident depression in this group; “yes” = proportion with incident depression in this group.

Figure 2. Variable importance plot from 10-fold cross-validated random forest for incident depression during follow-up, among men with no missing data (n = 1,409).

Figure 2 FOOTNOTES:

^aDuring most recent deployment.

PTSD = posttraumatic stress disorder.

Figure 3. Classification tree for incident depression during follow-up among women (n = 298).

Figure 3 FOOTNOTES:

In grey boxes: n = number of individuals with selected combination of predictors; “no” = proportion without incident depression in this group; “yes” = proportion with incident depression in this group.

Figure 4. Variable importance plot from 10-fold cross-validated random forest for incident depression during follow-up, among women with no missing data (n = 251).

Figure 4 FOOTNOTES:

^aDuring most recent deployment.







