



Published in final edited form as:

Mil Med. 2012 August ; 177(8 0): 47–59.

Prevalence of DSM-IV major depression among U.S. military personnel: Meta-analysis and simulation

Anne M. Gadermann, PhD^{*}, COL Charles C. Engel, MC, USA^{**}, James A. Naifeh, PhD^{**}, Matthew K. Nock, PhD^{***}, Maria Petukhova, PhD^{*}, LCDR Patcho N. Santiago, MC, USN^{**}, Wu Benjamin, MS^{*}, Alan M. Zaslavsky, PhD^{*}, and Ronald C. Kessler, PhD^{*}

^{*}Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston, MA 02115

^{**}Department of Psychiatry, Uniformed Services University of the Health Sciences, 4301 Jones Bridge Road, Bethesda, MD 20814

^{***}Department of Psychology, Harvard University, William James Hall, 1220, 33 Kirkland Street, Cambridge, MA 02138

Abstract

A meta-analysis of 25 epidemiological studies estimated the prevalence of recent DSM-IV major depression among U.S. military personnel. Best estimates of recent prevalence (standard error) were 12.0 percent (1.2) among currently deployed, 13.1 percent (1.8) among previously deployed and 5.7 percent (1.2) among never deployed. Consistent correlates of prevalence were being female, enlisted, young (ages 17 to 25), unmarried and having less than a college education. Simulation of data from a national general population survey was used to estimate expected lifetime prevalence of major depression among respondents with the socio-demographic profile and none of the enlistment exclusions of Army personnel. In this simulated sample, 16.2 percent (3.1) of respondents had lifetime major depression and 69.7 percent (8.5) of first onsets occurred before expected age of enlistment. Numerous methodological problems limit the results of the meta-analysis and simulation. The paper closes with a discussion of recommendations for correcting these problems in future surveillance and operational stress studies.

Keywords

Deployment; Depression; Epidemiology; Prevalence

INTRODUCTION

Major depression (MD) is generally recognized to be among the most burdensome of all disorders in the U.S. population¹ due to its high prevalence² and strong adverse effects on role functioning.³ As exposure to highly stressful life experiences is one of the most consistently documented risk factors for MD,⁴ it is not surprising that exposure to combat has been shown to be a powerful predictor of MD.⁵ Indeed, available research suggests that

MD might be as common as,⁶ or perhaps even more common than,⁷ post-traumatic stress disorder (PTSD) among combat veterans. Yet much more research has been carried out on the prevalence and correlates of PTSD than MD among military personnel.⁸

In an effort to synthesize available data on the prevalence of MD and its relationship to deployment experience, we carried out a quantitative literature review and meta-analysis of the recent literature on the epidemiology of MD among U.S. military personnel. We were mindful in planning this analysis that a recent review found a high range of MD prevalence estimates in studies of military personnel.⁶ The authors of that review cautioned that assessments of MD in the reviewed surveys were typically based on unvalidated screening scales rather than clinical interviews and that many studies used convenience samples rather than probability samples. We consequently limited our review to epidemiological studies that, with a few notable exceptions, used probability sampling methods and validated measures of MD.

Despite considerable information in the reviewed studies on current prevalence of MD, little data exist on lifetime prevalence or age-of-onset of MD among military personnel. Such data could be valuable in determining if military personnel with current MD had first onsets before versus only after enlistment. This information could have important implications in areas such as large-scale public health interventions and Physical Evaluation Boards. Even though direct data on lifetime prevalence are absent, simulation methods can be used to make indirect estimates. Messer et al.⁹ did this to estimate lifetime prevalence of selected mental disorders in the Army from the Epidemiologic Catchment Area (ECA) study,¹⁰ a large community epidemiological survey of mental disorders. We extend the work of Messer et al. here by using similar methods to estimate lifetime prevalence and age-of-onset of MD. The data used to carry out this simulation are from the National Comorbidity Survey Replication,¹¹ a national survey of the prevalence and correlates of DSM-IV mental disorders in the civilian U.S. household population.

METHODS

The meta-analysis

The literature search strategy—We searched PubMed (NCBI), Embase (Elsevier) and PsycINFO (EBSCO) for relevant studies published between January 1, 1990 and April 21, 2011 using relevant controlled vocabulary terms for (i) depression, (ii) military personnel and (iii) prevalence. (A detailed description of the full search strategy is available on request.) The search yielded 1,216 non-duplicated articles for review. We also contacted leading researchers in the epidemiology of mental disorders in the military for additional studies and searched for relevant reports. We focused on studies with a sample size of at least 1,000 individuals that provided estimates of recent prevalence of DSM-IV MD based on a validated screening measure (with demonstrated concordance to a diagnostic interview) or a diagnostic interview in a probability sample of individuals currently (at the time of the survey) serving in the U.S. Armed Forces. Studies that focused on clinical populations were excluded.

Two independent raters reviewed the abstracts and, based on the inclusion criteria, identified 32 articles and 13 reports for detailed review. Nineteen of these studies were subsequently excluded because they did not meet the inclusion criteria or used the same data as another publication that was included. (A detailed description is available on request). There were 26 remaining articles and reports (Table 1). One¹² used a subsample of a larger dataset.¹³ Because detailed information on MD by demographic characteristics was only available in the former,¹² that study was included in the analysis examining socio-demographic correlates of MD but not in the regression analyses. We included several studies in which MD was assessed with a version of the Patient Health Questionnaire (PHQ)-9 that included an additional requirement of self-reported functional impairment even though we were unable to find a validation study for this version of the PHQ-9. This was done because, as described below, we were able to develop a calibration for this version of the scale that approximated the more standard version. With regard to random sampling, none of the studies considered here was based on an unrestricted probability sample of the entire Army. The baseline Millennium Cohort Study¹⁴ and the periodic Department of Defense Surveys of Health-Related Behaviors among Military Personnel^{15–19} were based on representative samples of military personnel who were not deployed at the time of sample selection, while the other studies we refer to as probability surveys were based on samples of military personnel in more restrictive sampling frames, but in each case either selected a probability sample of military personnel from the frame or attempted to survey all personnel in the selected units or time periods. Several reports did not use probability sampling methods,^{20–23} but were included because they contained prevalence estimates for individuals currently deployed at the time of data collection that otherwise would have been strongly underrepresented in our analyses. A small number of the studies included respondents in the National Guard or Reserves who might have been recently deactivated at the time of data collection.

Each of the retained studies was entered into a data file for quantitative analysis. The variables included the prevalence estimate, the measure on which the prevalence estimate was based (see below), the sample size, information about whether the assessment was anonymous or not and the deployment status of respondents at the time of data collection (currently deployed, previously deployed or never deployed). A study was coded as anonymous only if this was explicitly stated. Studies coded not anonymous included confidential surveys in which identifying information was available but not disclosed to anyone not connected to the research and surveys that were mandatory for all service members returning from deployment, which were maintained in the permanent medical record. In cases where a single study included respondents with more than one deployment status and a MD prevalence estimate was presented separately by deployment status, the subsamples with different deployment statuses were treated as separate samples and entered as distinct observational records in the data file.^{24–27} In cases where the study included respondents with more than one deployment status but MD prevalence was not reported by deployment status, we treated the study as a single observational record and entered information in the data file for the proportions of respondents that were currently deployed, previously deployed and never deployed. In cases where these proportions were not reported in the study, they were estimated based on the best available information. (A detailed

description of the estimation methods is available on request.) The majority of assessments across studies were from respondents who had previously deployed (83.8%). Smaller proportions had never deployed (14.4%) or were currently deployed (1.8%) at the time of data collection. Studies that assessed MD longitudinally²⁸ or cross-sectionally at two time points (three and twelve months post-deployment)¹³ were treated as separate observational records. The subsamples in one especially large study²⁹ were also treated as separate observational records. This resulted in a total of 37 observational records in the final data file representing a total of 712,698 assessments.

Measures of major depression in the reviewed studies—The measures of MD in the reviewed studies include the PHQ-2,³⁰ PHQ-8,³¹ PHQ-9 with a severity coding scheme,³² PHQ-9 with a DSM-IV coding scheme (PHQ-9/DSM-IV),³³ PHQ-9/DSM-IV with an impairment requirement,³⁴ the three-item version A Burnam depression screen,³⁵ the eight-item Burnam depression screen³⁶ and the Center for Epidemiologic Studies Depression Scale (CES-D).³⁷ Recall periods were one week, two weeks or one month before interview (Table 1). These recall periods were treated as equivalent in assessing recent prevalence for the analysis. All the measures are screening scales; that is, while they assess some of the key symptoms of DSM-IV MD, they are designed to generate quick estimates of possible diagnosis rather than definitive diagnoses. We were unable to identify any studies using diagnostic interviews that met our inclusion criteria.

Quantitative analysis of the reviewed studies—Quantitative analysis was carried out to examine effects of methodological factors (the measure on which the prevalence estimate was based, sample size, anonymous versus not-anonymous) and deployment status on MD prevalence estimates. Random-effects multi-level regression analysis was used to analyze the data.³⁸ This is the preferred method for quantitative meta-analysis because it allows information about both sample size and study characteristics to be included in the analysis. The random effects model includes terms both for sampling error (sample-size dependent) and model error (representing effects of study-specific variation independent of sample size, such as unobserved variations in measurement methods, population and context). In this way, the model gives more weight to larger than smaller studies but does not allow any single very large study to swamp the effect of smaller studies because weighting takes into consideration the extent to which each observation deviates from the overall pattern in the full sample and down-weights observations that have large deviations. The analysis was carried out initially with the observed study prevalence estimate as the outcome and subsequently repeated with recalibrated measures of prevalence described below. The coefficients in these models were then used to estimate the prevalence of DSM-IV MD separately for deployed, previously deployed and never deployed military personnel by generating a predicted prevalence estimate from the model coefficients separately for respondents in each of the three deployment statuses based on assumptions about calibrations used to equalize estimates across types of measures described below. As we found that anonymity of reports is significantly related to elevated prevalence estimates, the predicted prevalence estimates were made based on the assumption that MD was assessed in an anonymous survey. Standard errors of the prevalence estimates were generated using the jackknife resampling method.³⁹

The simulation

The sample—As noted in the introduction, the simulation study was based on the NCS-R,¹¹ a 2001–03 national face-to-face survey of the prevalence and correlates of DSM-IV mental disorders in the adult (ages 18+) civilian U.S. household population. The response rate was 70.9 percent. The interview was conducted in two parts. Part one, completed by all 9,282 respondents, assessed a core set of DSM-IV mental disorders, while Part two, administered to all Part one respondents who screened positive for at least one Part one disorder (n = 4,235) plus a probability sub-sample of other Part one respondents (n = 1,457), assessed additional disorders and correlates. The Part two sample was weighted to adjust for differential probabilities of selection and the under-sampling of respondents with no Part one disorder. A final post-stratification weight was used to match the Part two sample to the 2000 census on a variety of socio-demographic and geographic variables. These sampling and weighting procedures are discussed in more detail elsewhere.⁴⁰

Sample matching—We sub-sampled Part two NCS-R respondents to create a weighted sub-sample that matched the population of active-duty Army personnel as closely as possible. We focused on Army personnel rather than military personnel more generally because the simulation was carried out as part of planning for the Army Study To Assess Risk and Resilience in Servicemembers (Army STARRS; <http://www.armystarrs.org>). Sub-sampling began by limiting NCS-R respondents to those in the age range 18 to 65 with at least a high school education (or General Education Diploma [GED]) who were employed and had health insurance. We then excluded respondents who would be ineligible for Army service based on: (1) conviction of a felony or serving at least one year in prison; (2) handicaps, including deafness, blindness, paralysis or a missing limb; (3) chronic physical disorders, including cardiovascular disorders (heart attack, stroke, hypertension, heart disease), respiratory disorders (COPD, asthma), diabetes, ulcer, HIV-AIDS, epilepsy or seizure disorder, Crohn's disease, cancer (except skin cancer), severe migraines and extreme obesity; and (4) severe mental disorders, including schizophrenia, other non-affective psychoses, bipolar I disorder and serious suicide attempts that occurred before the imputed age of enlistment. These exclusions are over-inclusive in that they remove people who might have entered the Army with waivers or developed chronic conditions after enlistment. Once the NCS-R sample was restricted in these ways, we selected a series of eight weighting variables available in the NCS-R and the Defense Department Defense Manpower Data Center (DMDC) master personnel dataset for Army personnel who were on active duty in December 2007 (<http://www.virec.research.va.gov/Non-VADataSources/DMDC.htm>). The eight weighting variables were age, sex, race-ethnicity (non-Hispanic black, non-Hispanic white, Hispanic and all others), education (high school graduates including those with a GED, some post-high school education without a bachelor's degree and bachelor's degree or more education), marital status (married, never married and previously married), U.S. citizenship (yes or no), nativity (i.e., born in the US yes or no) and religion (Protestantism, Catholicism, Judaism, Eastern [Buddhism, Hinduism, Islam], other, and atheist or no religion). These variables were selected for weighting because they are known to be significantly related to mental disorders and to have a significantly different distribution among Army personnel than the general U.S. population, although coarseness of some weighting categories makes the matching inexact.

The NCS-R weights were generated by using an exponential weighting function to make the distributions of the eight weighting variables in the adjusted NCS-R sample agree with the distributions in the DMDC dataset. (A detailed description of weighting procedures is available on request.) Regression-based imputation was used to assign an estimated age of enlistment to each NCS-R respondent by estimating a multiple regression equation using cross-tabulations of weighting variables from the DMDC in which the eight variables were used to predict age of enlistment. The regression coefficients from that equation were then applied to the NCS-R dataset to impute individual-level estimates of age of enlistment to match the DMDC distribution conditional on the matching variables.

Measurement of DSM-IV major depression in the NCS-R—Major depression was assessed in the NCS-R with Version 3.0 of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI),⁴¹ a fully-structured lay-administered interview that generates diagnoses for commonly occurring DSM-IV mental disorders. Good concordance was found between CIDI diagnoses and blinded clinical assessments in a NCS-R clinical reappraisal study.⁴² The CIDI yields information on lifetime history, age at first onset and presence of MD in the past year. Diagnoses were assigned based on reports of symptoms, duration and intensity as specified in the DSM-IV.

RESULTS

The meta-analysis

Major depression prevalence estimates in the reviewed studies—Major depression prevalence estimates vary widely across the 37 observational records, from a low of 2.0 percent to a high of 37.4 percent (Figure 1). Figure 1 shows the proportion of studies with prevalence estimates above the levels on the horizontal axis. There are four different lines in the figure based on the cross-classification of two dichotomous distinctions: (i) either using a weight to adjust the 37 different studies for variation in sample size or treating the studies as equal in importance regardless of sample size; and (ii) either using a calibration method described below to adjust prevalence estimates or considering prevalence estimates in the metrics reported in Table I. For current purposes the reader should focus only on the two lines without calibration.

The upper left corner of the figure shows that 100% of studies, by definition, have a prevalence estimate of 0.0% or more. The line for the unweighted (for variation in sample size across studies) and uncalibrated distribution across studies shows that median prevalence (i.e., the prevalence of the study with the 19th highest prevalence out of the 37 studies) is 7.8 percent, the mean 10.3 percent, and the inter-quartile range (IQR: 25th to 75th percentiles) is 4.8 to 15.7 percent. Weighting for variation in sample size across observations substantially reduces estimates both of central tendency (median from 7.8 to 6.1 percent; mean from 10.3 to 8.0 percent) and spread (IQR from 4.8 to 15.7 percent to 3.8 to 10.3 percent).

One possible reason for the wide variation in these prevalence estimates is that the different screening scales might differ in sensitivity and specificity. Published validity studies are consistent with this possibility, suggesting that the PHQ-9 with severity scoring has the

highest sensitivity and that the eight-item Burnam scale has the highest specificity (Table 2). These validation studies typically administered a gold standard clinical reappraisal interview to a probability sub-sample of people shortly after they were administered the screening scale. The clinical interviewers typically were blinded to the screening scale scores. These validity studies were conducted in civilian populations, though, mainly in primary care samples, making it unclear whether the estimates of sensitivity and specificity in these studies generalize to non-patient samples of military personnel.

It is possible to adjust prevalence estimates in a screening scale to approximate estimates of “true” prevalence if information is available on the sensitivity and specificity of the screening scale. For example, if we know that sensitivity is 50% (i.e., half of the true cases are detected by the screening scale) and specificity is 100% (i.e., all the true non-cases are classified as non-cases by the screening scale), then the estimated prevalence is only 50% as high as the true prevalence, meaning that the best estimate of true prevalence is two times the prevalence estimate in the screening scale. A standard formula exists for converting prevalence estimates in screening scales to estimates of true prevalence based on information about sensitivity and specificity.⁴³ We used that formula to adjust the prevalence estimates based on the screening scales in each of the studies reported in Table II. The estimates of sensitivity and specificity used in doing this were the published estimates of sensitivity and specificity for these screening scales. However, these transformations yielded implausible estimates of true MD prevalence. This was most clear in quite a few studies where estimated prevalence was strongly negative (i.e., not merely within sampling error of 0.0%, but substantially less than 0.0%). A negative prevalence, of course, is impossible. So what does it mean to find that adjusted prevalence estimates are negative? (Detailed results of this estimation exercise are available on request but are not presented here because so many of the estimates are implausible.) It means, quite simply, that the sensitivity and specificity estimates in the published validity studies of the screening scales do not apply to the samples considered here. That is, the true sensitivity and specificity of the screening scales in the studies where they were used must have been different than the sensitivity and specificity estimated in the methodological studies of the screening scales. There are a variety of reasons why this might be the case, but the most plausible one is that the samples used in the original validity studies of the screening scales might have been different than those in the substantive studies reported in Table II (e.g., more severe cases of MD, which would lead to differences in rates of detection). We have no way to produce more accurate estimates of sensitivity and specificity for the studies in Table II, as these studies did not include the blinded clinical reappraisal interviews with probability sub-samples of respondents that would be required to calculate independent estimates of sensitivity and specificity for these specific studies. Based on these facts, we abandoned the attempt to correct prevalence estimates in these studies for differential sensitivity and specificity.

Yet the substantial variation in prevalence estimates across these studies raises the possibility that between-measure variation in concordance with clinical diagnoses could be important. In the absence of being able to correct for this variation by using sensitivity-specificity adjustments, we turned to a different method: calibration of the prevalence estimates across studies to a common metric by making use of the fact that a number of

epidemiological studies – some of them in military samples and others in civilian samples – have reported MD prevalence estimates based on two or more of the measures used in the 37 observational records considered here. Access to these within-study pairs of prevalence estimates allowed us to create prevalence ratios to transform prevalence estimates based on one measure to estimates based on another measure. The common metric we transformed to was the DSM-IV coding scheme for the PHQ-9. The latter scheme, used in five of the 25 studies considered here, requires at least five of the nine PHQ-9 questions about criterion A symptoms of MD to be reported as having occurred more than half the days over the recall period and for at least one of these questions to involve either depressed mood or anhedonia. One study reported prevalence estimates based on this PHQ-9/DSM-IV coding scheme as well as on the PHQ-9 severity coding scheme and the PHQ-2.⁴⁴ Three separate articles from a second study reported prevalence estimates based on this same set of three coding schemes.^{30, 32, 33} Two other studies reported prevalence estimates based on both the PHQ-9/DSM-IV and the version of the PHQ-9 coding scheme that requires impairment.^{13, 25} We used the weighted average ratios of prevalence estimates based on these alternative coding schemes to transform prevalence estimates based on other measures to PHQ-9/DSM-IV prevalence estimates.

In cases where no study existed that presented prevalence estimates based on both the PHQ-9/DSM-IV coding scheme and one of the other measures, we made indirect calibrations using a third measure. For example, while there were no studies that included prevalence estimates based on both the CES-D and the PHQ-9/DSM-IV coding scheme, one study presented both CES-D and PHQ-2 prevalence estimates,⁴⁵ while two others presented both PHQ-2 and PHQ-9/DSM-IV prevalence estimates,^{30, 33, 44} allowing us to multiply these two separate ratios together to generate a synthetic CES-D versus PHQ-9/DSM-IV calibration ratio to adjust prevalence estimates in the one study that used the CES-D to estimate depression prevalence.

As shown in Figure 1, this calibration exercise substantially reduces the average prevalence estimates from median and mean of 6.1 and 8.0 percent in the weighted raw data to 3.0 and 4.5 percent in the weighted calibrated data as well as in the IQR (from 6.5 percent [between 3.8 and 10.3 percent] in the raw data to 2.7 percent [between 2.3 and 5.0 percent] in the calibrated data).

Multiple regression analysis—The test for the variance of random intercepts is significant in the random effects model of the raw outcomes ($\chi^2_1 = 13.0$, $p < 0.001$), indicating significant heterogeneity among observations. This supports the decision to use the random effects model. Methodological and substantive variables are both significant predictors of variation in MD prevalence estimates across the 37 raw observational records (Table 3). With regard to methodological factors, prevalence estimates differ significantly by type of screening scale ($\chi^2_7 = 37.1$, $p < 0.001$) and are significantly higher in anonymous than identified surveys (odds ratio [OR] = 3.1, $t = 4.2$, $p = 0.002$). With regard to substantive factors, prevalence estimates are significantly higher among the currently deployed (OR = 2.7, $t = 3.4$, $p = 0.006$) and previously deployed (OR = 3.2, $t = 12.7$, $p < 0.001$) than the never deployed. It is important to recognize that results are based on a multivariate, which means that ORs for each predictor are net of those for other predictors.

The results are different when the same model is used to predict variation in recalibrated MD prevalence estimates. The most dramatic difference is that estimated prevalence is no longer predicted significantly by type of screening scale ($\chi^2_7 = 2.2$, $p = 0.95$), indicating that the recalibration exercise was successful. However, survey anonymity remains significantly associated with elevated prevalence (OR = 2.9, $t = 4.0$, $p = 0.002$). Furthermore, the currently deployed and previously deployed continue to have significantly higher prevalence estimates than the never deployed, although these ORs are somewhat lower than when the model is estimated on the raw data (OR = 2.2, $t = 2.7$, $p = 0.020$ for currently deployed; OR = 2.5, $t = 9.8$, $p < 0.001$ for previously deployed). These significant ORs change only modestly in a model that deletes predictors for type of screening scale.

Major depression prevalence estimates based on the best-fitting regression model—Based on the assumption that the higher MD prevalence estimates in anonymous surveys are more accurate than the lower estimates in non-anonymous surveys, the parameters of the best-fitting random-effects model for the calibrated data were used to generate best estimates of MD prevalence for anonymous surveys. As noted above in the section on analysis methods, the jackknife resampling method was used to generate estimates of standard error (SE). Best estimates of current MD prevalence (SE) are 12.0 percent (1.2) for the currently deployed, 13.1 percent (1.8) for the previously deployed and 5.7 percent (1.2) for the never deployed.

Other correlates of major depression—A number of surveys report MD prevalence by socio-demographic variables and/or by branch of service. We calculated ORs for these estimates within studies and then summarized these results by computing weighted (by sample size) averages of ORs across studies. Women are found consistently to have higher rates of MD than men with a mean (range) OR of 1.6 (1.1 to 1.9) across studies (Table 4). Prevalence also is higher among respondents with no more than high school education (3.0 [2.0 to 3.6]) or some college education (1.8 [1.6 to 2.1]) relative to college graduates. Prevalence generally is unrelated to race-ethnicity. Prevalence is consistently higher among enlisted (2.8 [1.9 to 3.6]) personnel than warrant officers (1.1 [0.9 to 1.2]) or commissioned officers (the contrast category, with an implicit OR of 1.0). In addition, MD generally is estimated to be more common among younger (up to ages 24 or 25) than older (older than 24 or 25) respondents (2.0 [1.0 to 2.2]) and among the unmarried (either never married or previously married) than the married (1.8 [1.0 to 2.0]). The studies that compared MD across services report consistently higher prevalence in the Army (2.0 [1.6 to 2.1]), Navy (1.7 [1.3 to 1.8]) and Marines (2.0 [1.4 to 2.3]) than the Air Force.

The simulation

Current depression prevalence estimates in the simulation data—A question can be raised how the prevalence estimates reported above for current MD compare to the general U.S. population. The comparable prevalence estimate (SE) in the simulated NCS-R data is 1.3 percent (0.6). To be clear, this is the rate we would expect in a representative sample of people in the U.S. population who have the same socio-demographic profile (e.g., age, sex, race-ethnicity and education) and history of pre-enlistment health problems as the members of the U.S. Army. The 1.3 percent MD prevalence estimate is substantially lower

than the estimates reported above in the meta-analysis. Even though the NCS-R simulation uses a different measure of MD than any of the meta-analysis surveys, the NCS-R measure has been validated in the general population and shown to yield a prevalence estimate very similar to the estimate based on blinded clinical reappraisal interviews using DSM-IV criteria.² We would consequently expect that the meta-analysis estimates, based on calibration to the PHQ-9/DSM-IV, would be comparable to the simulated NCS-R/DSM-IV estimates.

Lifetime depression prevalence estimates in the simulation data—As noted in the introduction, much less is known about lifetime prevalence than current prevalence of MD among military personnel. The simulated NCS-R data show that 16.2 percent of the sample has a lifetime history of MD, that 69.7 percent of these lifetime cases (i.e., 11.3 percent of the total sample) had first onsets before enlistment (i.e., before the age when we would have expected them to enlist based on their socio-demographic profile) and that the remaining 30.3 percent of lifetime cases (i.e., 5.5 percent of the total sample) had first onsets only after enlistment (Table 5). The majority of those with current MD had first onsets prior to enlistment (77.9 percent). The latter is higher than the 69.7 percent of lifetime cases with pre-enlistment MD, suggesting that MD persistence is higher among early-onset than later-onset cases. This higher persistence is indirectly indicated by the fact that the ratio of current to lifetime prevalence is higher among respondents with pre-enlistment (8.8 percent) than post-enlistment (5.4 percent) MD.

Given the much higher prevalence of current MD among Army personnel than expected from the simulations, it is unclear from these data what percent of actual Army personnel with current MD had first onsets prior to enlistment. The high current prevalence estimates from the meta-analysis could reflect either a dramatic increase in current prevalence among lifetime cases once they enter the Army, a dramatic increase in post-enlistment first onset or a combination. If it is true that 11.3 percent of actual Army personnel had a history of MD before enlistment, and if post-enlistment onsets were double the estimate in the simulation data (i.e., 11.0 percent rather than 5.5 percent), then the ratio of current to lifetime prevalence would have to be at least 25 percent among the never deployed (5.7/22.3) and 50 percent among the currently (12.0/22.3) and previously (13.1/22.3) deployed to generate the estimates of current prevalence found in the meta-analysis data.

DISCUSSION

The meta-analysis reported here was limited by the fact that a wide range of MD screening scales were used in the different studies and by the fact that prevalence estimates vary significantly by type of screening scale. We attempted to address this problem by transforming the screening scale prevalence estimates to a common metric based on the results of published validity studies. This attempt failed, though, as some “corrected” prevalence estimates were less than zero. This means that the sensitivity and specificity estimates reported in the published validity studies, all of which were carried out in civilian samples and mostly among primary care patients, do not apply to the military samples considered here. The calibration approach we used to address this problem was limited by the fact that it required the assumption that prevalence ratios across different screening

scales in a single survey could legitimately be generalized across surveys. Future epidemiological studies of depression in the military should address this problem more directly by using the same screening measure. This would be consistent with recent recommendations for use of common data elements in surveillance and operational stress research.⁴⁶

It is noteworthy that none of the screening scales included an exclusion for bipolar disorder. This means that they screened for major depressive episodes (MDE), not for major depressive disorder (MDD), and that some unknown proportion of these cases represent depressive phases of a bipolar disorder (BPD). As bipolar-I (BP-I) leads to military discharge and is so dramatic during the manic phase that it is likely to have a high rate of detection, we would not expect many cases of MD in the military to be associated with BP-I. But BP-II and sub-threshold BPD are together much more common than BP-I³ and often go undetected. Intervention implications are quite different for BPD than MDD, making it important to distinguish between the two. Future epidemiological studies of depression in the military should address this problem by including a BPD screen and the MD screen. This is being done in the Army STARRS study, but we are unaware of any other large-scale military epidemiological survey that has done so.

It would also be useful to include a small clinical reappraisal component in each future major epidemiological survey of military mental health even if a consistent MD screening scale was used. Repeated validity studies are needed because variation in the accuracy of any screening scale across studies can be influenced by survey conditions (e.g., anonymity, rationale, the context created by preceding survey questions and the physical conditions of respondents at the time of survey implementation) that vary across studies.⁴⁷

Another methodological feature that could usefully be added to future military epidemiological studies would be a non-respondent adjustment process. This could include a non-respondent sub-sampling outreach phase in which limited information is obtained from a probability sub-sample of survey non-respondents. Or it could use administrative databases (e.g., information from military electronic medical records about history of diagnoses of mental illness) to weight the survey data for under-representation of personnel with profiles associated with high risk of MD or other mental disorders. Methods of these sorts have been used successfully to address sample bias in other epidemiological surveys.⁴⁸ Weighting seems like an especially attractive approach in military surveys in light of the existence of an extensive series of administrative databases for all military personnel.

A related limitation of our meta-analysis is that the samples, although largely based on probability selection methods within the units studied, often used non-probability methods to select units and, within units, to select critical times in the unit life cycle. This led to over-representation of combat units as well as to over-representation of the months just before deployment and just after redeployment. Although it would theoretically be possible to correct for these sampling biases with weights, the logistical complexities of doing so made this impossible in practice. As a result, caution is needed in drawing inferences from our summary results because of the likely skewed distributions in our samples of military occupation specialties (MOSs), units and timing of deployment histories. Not all of the

samples considered in our meta-analysis shared this last limitation, as some surveys were representative of all military personnel in one or more branches of service. However, in order to use the data from these studies to adjust the results across all studies, we would have needed to work with individual-level data rather than the aggregate data available to us. This highlights another limitation of our meta-analysis: that it was based on summary published results rather than on secondary analysis of individual-level data. More fine-grained analysis could have been carried out in individual-level secondary analysis, including but not limited to, using weights to adjust sample composition for the over-representation of some MOSs, types of units and deployment histories. Pooled secondary analysis of existing survey data has been of great value in advancing our understanding of the epidemiology of depression in the general population.⁴⁹ The same could be true for research on depression (and other mental disorders) in the military if de-identified individual-level data were made available.

Our simulation had only a limited set of variables, some of them relatively coarse, to match the NCS-R national household sample with the characteristics of Army personnel. Failure to control for the many unmeasured selection factors that might influence both enlistment in the Army and depression could have distorted the results. In addition, the simulation was designed to provide data on what the prevalence and age-of-onset distribution of depression would have been expected to be among Army personnel if they had not joined the Army. Although that kind of information is potentially useful in assessing the impact of Army experiences, in the context of the much higher current prevalence estimates in the meta-analysis than the simulation, it tells us nothing about the lifetime prevalence of post-enlistment onset depression or about the persistence of either pre-enlistment or post-enlistment depression.

Our estimates of current MD prevalence in the military are much higher than the 30-day prevalence estimate obtained for socio-demographically comparable civilians in the simulation study. It needs to be noted that the MD prevalence estimates from the meta-analysis were generated based on the parameters for anonymous surveys, whereas the simulation results are based on confidential (but not anonymous) interviews from a general population survey. Previous research indicates that respondents are more likely to provide accurate information on sensitive questions in anonymous versus confidential questionnaires,⁵⁰ but this would explain only a small part of the difference in MD prevalence estimates. The finding that the prevalence estimate was higher for the previously than currently deployed could be an artifact in that the previously deployed personnel considered here over-represented those who had recently returned from deployment. Current prevalence among military personnel was estimated to be higher for women than men, young than old, the unmarried than the married, and those with lower than higher rank and education. These correlates are broadly consistent with those found in general population surveys.^{2, 51}

We estimated that 16.2 percent of respondents in the simulation data had a lifetime history of MD and that the majority (69.7 percent) of these lifetime cases had first onsets prior to expected time of enlistment. We have no comparable lifetime prevalence estimate in the meta-analysis data, although we would expect that lifetime prevalence among military

personnel would be higher due to a presumed larger number of post-enlistment onsets than at comparable ages in the general population. In the absence of a direct estimate of persistence, though, we have no way to know how much higher the prevalence of post-enlistment onsets are in the military than the general population or the proportion of current cases that had first onsets prior to enlistment. However, the high estimated pre-enlistment lifetime prevalence in the simulation data, when coupled with the finding that early onset is positively associated with persistence, lead us to speculate that a substantial minority or perhaps even a majority of military personnel with current depression had first onsets prior to enlistment. To the extent that this is true, secondary preventive interventions with recruits having a pre-enlistment history of depression (or other mental disorders that predict subsequent depression) might be effective in reducing incidence of subsequent depressive episodes among military personnel. More direct data would be needed, though, on lifetime history, age-of-onset, current prevalence and severity of current depression in representative military samples before any such interventions could reasonably be planned. The Army STARRS study is collecting such data for the Army, but we are unaware of any attempt to collect comparable data for other branches of the military.

Acknowledgments

The authors would like to thank MAJ Owen Hill, Craig McKinnon and Monika Wahi of the injury epidemiology research section, military performance division, U. S. Army Research Institute of Environmental Medicine, Natick, Mass., for providing guidance as well as weighting information from the total Army injury and health outcomes database (TAIHOD).

FUNDING

This study was funded under a cooperative agreement between the U.S. Army and the U.S. Department of Health and Human Services, National Institutes of Health, National Institute of Mental Health (NIMH 1U01MH087981-01 and 1U01MH087981-01S1). The contents are solely the responsibility of the authors and do not necessarily represent the views of the Department of Health and Human Services, NIMH, the Department of the Army or the Department of Defense.

References

1. Luppá M, Heinrich S, Angermeyer MC, König HH, Riedel-Heller SG. Cost-of-illness studies of depression: a systematic review. *J Affect Disord.* 2007; 98(1–2):29–43. [PubMed: 16952399]
2. Kessler RC, Berglund P, Demler O, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA.* 2003; 289(23):3095–105. [PubMed: 12813115]
3. Merikangas KR, Ames M, Cui L, et al. The impact of comorbidity of mental and physical conditions on role disability in the US adult household population. *Arch Gen Psychiatry.* 2007; 64(10):1180–8. [PubMed: 17909130]
4. Monroe, SM.; Slavich, GM.; Georgiades, K. The social environment and life stress in depression. In: Gotlib, IH.; Hammen, CL., editors. *Handbook of Depression.* 2. New York, NY: Guilford; 2009. p. 340-60.
5. Prigerson HG, Maciejewski PK, Rosenheck RA. Population attributable fractions of psychiatric disorders and behavioral outcomes associated with combat exposure among US men. *Am J Public Health.* 2002; 92(1):59–63. [PubMed: 11772762]
6. Ramchand, R.; Karney, BR.; Osilla, KC.; Burns, RM.; Caldarone, LB. Prevalence of PTSD, depression, and TBI among returning servicemembers. In: Tanielian, T.; Jaycox, LH., editors. *Invisible Wounds of War: Psychological and Cognitive Injuries, Their Consequences, and Services to Assist Recovery.* Santa Monica, CA: RAND Corporation; 2008. p. 35-85.

7. Iversen A, Dyson C, Smith N, et al. 'Goodbye and good luck': The mental health needs and treatment experiences of British ex-service personnel. *Br J Psychiatry*. 2005; 186(6):480–6. [PubMed: 15928358]
8. Wells TS, Miller SC, Adler AB, Engel CC, Smith TC, Fairbank JA. Mental health impact of the Iraq and Afghanistan conflicts: a review of US research, service provision, and programmatic responses. *Int Rev Psychiatry*. 2011; 23(2):144–52. [PubMed: 21521083]
9. Messer SC, Liu X, Hoge CW, Cowan DN, Engel CC Jr. Projecting mental disorder prevalence from national surveys to populations-of-interest. An illustration using ECA data and the U. S. Army Soc Psychiatry *Psychiatr Epidemiol*. 2004; 39(6):419–26.
10. Robins, LN.; Regier, DA. *Psychiatric Disorders in America: The Epidemiologic Catchment Area Study*. New York, NY: The Free Press; 1991.
11. Kessler RC, Merikangas KR. The National Comorbidity Survey Replication (NCS-R): background and aims. *Int J Methods Psychiatr Res*. 2004; 13(2):60–68. [PubMed: 15297904]
12. Riviere LA, Kendall-Robbins A, McGurk D, Castro CA, Hoge CW. Coming home may hurt: risk factors for mental ill health in US reservists after deployment in Iraq. *Br J Psychiatry*. 2011; 198(2):136–42. [PubMed: 21282784]
13. Thomas JL, Wilk JE, Riviere LA, McGurk D, Castro CA, Hoge CW. Prevalence of mental health problems and functional impairment among active component and National Guard soldiers 3 and 12 months following combat in Iraq. *Arch Gen Psychiatry*. 2010; 67(6):614–23. [PubMed: 20530011]
14. Riddle JR, Smith TC, Smith B, et al. Millennium Cohort: the 2001–2003 baseline prevalence of mental disorders in the U.S. military. *J Clin Epidemiol*. 2007; 60(2):192–201. [PubMed: 17208126]
15. Bray, RM.; Kroutil, LA.; Wheelless, SC., et al. 1995 Department of Defense Survey of Health Related Behaviors Among Military Personnel. Research Triangle Park, NC: Research Triangle Institute; 1995.
16. Bray, RM.; Sanchez, RP.; Ornstein, ML.; Lentine, D.; Vincus, AA. 1998 Department of Defense Survey of Health Related Behaviors Among Military Personnel. Research Triangle Park, NC: Research Triangle Institute; 1999.
17. Bray, RM.; Hourani, LL.; Rae, KL., et al. 2002 Department of Defense Survey of Health Related Behaviors Among Military Personnel. Research Triangle Park, NC: Research Triangle Institute; 2003.
18. Bray, RM.; Hourani, LL.; Olmsted, KLR., et al. 2005 Department of Defense Survey of Health Related Behaviors Among Active Duty Military Personnel: A Component of the Defense Lifestyle Assessment Program (DLAP). Research Triangle Park, NC: Research Triangle Institute; 2006.
19. Bray, RM.; Pemberton, MR.; Hourani, LL., et al. 2008 Department of Defense Survey of Health Related Behaviors Among Active Duty Military Personnel: A Component of the Defense Lifestyle Assessment Program (DLAP). Research Triangle Park, NC: Research Triangle Institute; 2009.
20. Mental Health Advisory Team. (MHAT)-II: Operation Iraqi Freedom: US Army Surgeon General. *Army Medicine*; 2005. Available at http://www.armymedicine.army.mil/reports/mhat/mhat_ii/OIF-II_REPORT.pdf [accessed August 24, 2011]
21. Mental Health Advisory Team (MHAT)-III. Operation Iraqi Freedom 04–06: Office of the Surgeon Multinational Force-Iraq and Office of the Surgeon General United States Army Medical Command. *Army Medicine*; 2006. Available at http://www.armymedicine.army.mil/reports/mhat/mhat_iii/MHATIII_Report_29May2006-Redacted.pdf [accessed August 24, 2011]
22. Mental Health Advisory Team (MHAT)-IV. Operation Iraqi Freedom 05–07: Office of the Surgeon Multinational Force-Iraq and Office of the Surgeon General United States Army Medical Command. *Army Medicine*; 2006. Available at http://www.armymedicine.army.mil/reports/mhat/mhat_iv/MHAT_IV_Report_17NOV06.pdf [accessed August 24, 2011]
23. Mental Health Advisory Team (MHAT)-V. Operation Iraqi Freedom 06–08: Iraq; Operation Enduring Freedom 8: Afghanistan: Office of the Surgeon Multi-National Force-Iraq and Office of the Command Surgeon and Office of the Surgeon General United States Army Medical Command. *Army Medicine*; 2008. Available at http://www.armymedicine.army.mil/reports/mhat/mhat_v/MHAT_V_OIFandOEF-Redacted.pdf [accessed August 24, 2011]

24. Cabrera OA, Hoge CW, Bliese PD, Castro CA, Messer SC. Childhood adversity and combat as predictors of depression and post-traumatic stress in deployed troops. *Am J Prev Med.* 2007; 33(2):77–82. [PubMed: 17673093]
25. Hoge CW, Castro CA, Messer SC, McGurk D, Cotting DI, Koffman RL. Combat duty in Iraq and Afghanistan, mental health problems, and barriers to care. *N Engl J Med.* 2004; 351(1):13–22. [PubMed: 15229303]
26. Kline A, Falca-Dodson M, Sussner B, et al. Effects of repeated deployment to Iraq and Afghanistan on the health of New Jersey Army National Guard troops: implications for military readiness. *Am J Public Health.* 2010; 100(2):276–83. [PubMed: 20019304]
27. Luxton DD, Skopp NA, Maguen S. Gender differences in depression and PTSD symptoms following combat exposure. *Depress Anxiety.* 2010; 27(11):1027–33. [PubMed: 20721909]
28. Milliken CS, Auchterlonie JL, Hoge CW. Longitudinal assessment of mental health problems among active and reserve component soldiers returning from the Iraq war. *JAMA.* 2007; 298(18): 2141–8. [PubMed: 18000197]
29. Hoge CW, Auchterlonie JL, Milliken CS. Mental health problems, use of mental health services, and attrition from military service after returning from deployment to Iraq or Afghanistan. *JAMA.* 2006; 295(9):1023–32. [PubMed: 16507803]
30. Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: validity of a two-item depression screener. *Med Care.* 2003; 41(11):1284–92. [PubMed: 14583691]
31. Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. *J Affect Disord.* 2009; 114(1–3):163–73. [PubMed: 18752852]
32. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med.* 2001; 16(9):606–13. [PubMed: 11556941]
33. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire.* *JAMA.* 1999; 282(18):1737–44. [PubMed: 10568646]
34. Hoge CW, Castro CA, Messer SC, McGurk D, Cotting DI, Koffman RL. Combat duty in Iraq and Afghanistan, mental health problems and barriers to care. *US Army Med Dep J.* 2008 Jul-Sep;:7–17. [PubMed: 20088060]
35. Rost K, Burnam MA, Smith GR. Development of screeners for depressive disorders and substance disorder history. *Med Care.* 1993; 31(3):189–200. [PubMed: 8450677]
36. Burnam MA, Wells KB, Leake B, Landsverk J. Development of a brief screening instrument for detecting depressive disorders. *Med Care.* 1988; 26(8):775–89. [PubMed: 3398606]
37. Radloff LS. The CES-D Scale: a self-report depression scale for research in the general population. *Appl Psychol Measur.* 1977; 1(3):385–401.
38. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials.* 1986; 7(3):177–88. [PubMed: 3802833]
39. Shao, J.; Tu, D. *The Jackknife and Bootstrap.* New York, NY: Springer-Verlag; 1995.
40. Kessler RC, Berglund P, Chiu WT, et al. The US National Comorbidity Survey Replication (NCS-R): design and field procedures. *Int J Methods Psychiatr Res.* 2004; 13(2):69–92. [PubMed: 15297905]
41. Kessler RC, Üstün TB. The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *Int J Methods Psychiatr Res.* 2004; 13(2):93–121. [PubMed: 15297906]
42. Haro JM, Arbabzadeh-Bouchez S, Brugha TS, et al. Concordance of the Composite International Diagnostic Interview Version 3.0 (CIDI 3.0) with standardized clinical assessments in the WHO World Mental Health surveys. *Int J Methods Psychiatr Res.* 2006; 15(4):167–80. [PubMed: 17266013]
43. Rogan WJ, Gladen B. Estimating prevalence from the results of a screening test. *Am J Epidemiol.* 1978; 107(1):71–6. [PubMed: 623091]
44. Arroll B, Goodyear-Smith F, Crengle S, et al. Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. *Ann Fam Med.* 2010; 8(4):348–53. [PubMed: 20644190]

45. Whooley MA, Avins AL, Miranda J, Browner WS. Case-finding instruments for depression. Two questions are as good as many. *J Gen Intern Med.* 1997; 12(7):439–45. [PubMed: 9229283]
46. Nash WP, Vasterling J, Ewing-Cobbs L, et al. Consensus recommendations for common data elements for operational stress research and surveillance: report of a federal interagency working group. *Arch Phys Med Rehabil.* 2010; 91(11):1673–83. [PubMed: 21044711]
47. Groves, RM.; Fowler, JFJ.; Couper, MP.; Lepkowski, JM.; Singer, E.; Tourangeau, R. *Survey Methodology.* Hoboken, NJ: John Wiley & Sons; 2004.
48. Kessler RC, Little RJ, Groves RM. Advances in strategies for minimizing and adjusting for survey nonresponse. *Epidemiol Rev.* 1995; 17(1):192–204. [PubMed: 8521937]
49. Klerman GL, Weissman MM. Increasing rates of depression. *JAMA.* 1989; 261(15):2229–35. [PubMed: 2648043]
50. Ong AD, Weiss DJ. The impact of anonymity on responses to sensitive questions. *J Appl Soc Psychol.* 2000; 30(8):1691–708.
51. Hasin DS, Goodwin RD, Stinson FS, Grant BF. Epidemiology of major depressive disorder: results from the National Epidemiologic Survey on Alcoholism and Related Conditions. *Arch Gen Psychiatry.* 2005; 62(10):1097–106. [PubMed: 16203955]
52. Lapierre CB, Schwegler AF, Labauve BJ. Posttraumatic stress and depression symptoms in soldiers returning from combat operations in Iraq and Afghanistan. *J Trauma Stress.* 2007; 20(6): 933–43. [PubMed: 18157882]
53. Reger MA, Gahm GA, Swanson RD, Duma SJ. Association between number of deployments to Iraq and mental health screening outcomes in US Army soldiers. *J Clin Psychiatry.* 2009; 70(9): 1266–72. [PubMed: 19689917]
54. Warner CM, Warner CH, Breitbach J, Rachal J, Matuszak T, Grieger TA. Depression in entry-level military personnel. *Mil Med.* 2007; 172(8):795–9. [PubMed: 17803068]
55. Mental Health Advisory Team (MHAT)-VI. Operation Enduring Freedom 2009: Afghanistan: Office of the Command Surgeon US Forces Afghanistan (USFOR-A) and Office of the Surgeon General United States Army Medical Command. *Army Medicine*; 2009. Available at http://www.armymedicine.army.mil/reports/mhat/mhat_vi/MHAT_VI-OEF_Redacted.pdf [accessed August 24, 2011]
56. Mental Health Advisory Team (MHAT)-VI. Operation Iraqi Freedom 07–09: Office of the Surgeon Multi – National Corps - Iraq and Office of the Surgeon General United States Army Medical Command. *Army Medicine*; 2009. Available at http://www.armymedicine.army.mil/reports/mhat/mhat_vi/MHAT_VI-OIF_Redacted.pdf [accessed August 24, 2011]
57. Joint Mental Health Advisory Team (MHAT)-7. Operation Enduring Freedom 2010: Afghanistan: Office of The Surgeon General United States Army Medical Command and Office of the Command Surgeon HQ, USCENTCOM and Office of the Command Surgeon US Forces Afghanistan (USFOR-A). *Army Medicine*; 2011. Available at http://www.armymedicine.army.mil/reports/mhat/mhat_vii/J_MHAT_7.pdf [accessed August 24, 2011]
58. Schell, TL.; Marshall, GN. Survey of individuals previously deployed for OEF/OIF. In: Tanidian, T.; Jaycox, LH., editors. *Invisible Wounds of War: Psychological and Cognitive Injuries, Their Consequences, and Services to Assist Recovery.* Santa Monica, CA: RAND Corporation; 2008. p. 87-115.
59. Wittkamp KA, Naeije L, Schene AH, Huyser J, van Weert HC. Diagnostic accuracy of the mood module of the Patient Health Questionnaire: a systematic review. *Gen Hosp Psychiatry.* 2007; 29(5):388–95. [PubMed: 17888804]
60. Fechner-Bates S, Coyne JC, Schwenk TL. The relationship of self-reported distress to depressive disorders and other psychopathology. *J Consult Clin Psychol.* 1994; 62(3):550–9. [PubMed: 8063981]
61. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. *Psychiatr Ann.* 2002; 32(9):1–7.
62. Kroenke K, Spitzer RL, Williams JB, Lowe B. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. *Gen Hosp Psychiatry.* 2010; 32(4):345–5. [PubMed: 20633738]

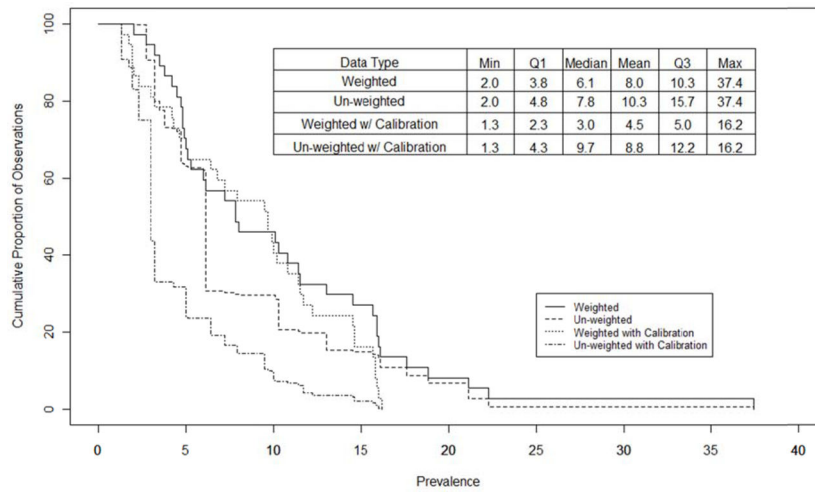


Figure 1. Inverse cumulative distribution function of major depression prevalence based on studies included in the analyses (based on 37 observational records containing a total of 712,698 assessments)

Table 1

Description of studies included in the analyses

Study ID	Branch ^d										Deployment status ^b			RR ^f %	Prevg %	Years ^h	(n) ^j	Citation
	A	N	AF	M	CG	PD	CD	ND	Anonymous ^c	Measure ^d	RP ^e							
1a	X					X	X		Yes	5	3	54 ^j	6.0	2003	4,529	Cabrera 2007 ²⁴		
1b	X					X			Yes	5	3	54 ^j	7.8	2004	2,392	Cabrera 2007 ²⁴		
2a	X					X	X		Yes	4	3	77	11.4	2003	2,418	Hoge 2004 ²⁵		
2b	X			X		X			Yes	4	3	53	14.5	2003	3,500	Hoge 2004 ²⁵		
3a ^k	X			X		X		No	No	1	2	+ ^j	3.5	2003/04	16,318	Hoge 2006 ²⁹		
3b ^k	X			X		X		No	No	1	2	+ ^j	6.1	2003/04	222,620	Hoge 2006 ²⁹		
3c ^k	X			X		X		No	No	1	2	+ ^j	2.7	2003/04	64,967	Hoge 2006 ²⁹		
4	X					X		Yes	Yes	5	3	59	5.3	2006	2,464	Hoge 2008 ³⁴		
5a	X					X	X		Yes	4	2	89	2.0	2007/08	1,910	Kline 2010 ²⁶		
5b	X					X		Yes	Yes	4	2	89	5.1	2007/08	625	Kline 2010 ²⁶		
6	X					X		No	No	6	1	48	37.4	2005	4,089	Lapierre 2007 ⁵²		
7a	X					X	X		No	3	2	+ ^j	4.5	2004-08	6,943	Luxton 2010 ²⁷		
7b	X					X		No	No	3	2	+ ^j	10.1	2006-09	6,943	Luxton 2010 ²⁷		
8a ^m	X					X		No	No	1	2	+ ^j	4.7	2004-06	56,350	Milliken 2007 ²⁸		
8b ^m	X					X		No	No	1	3	+ ^j	10.3	2005/06	56,350	Milliken 2007 ²⁸		
8c ^m	X					X		No	No	1	2	+ ^j	3.8	2004-06	31,885	Milliken 2007 ²⁸		
8d ^m	X					X		No	No	1	3	+ ^j	13.0	2005/06	31,885	Milliken 2007 ⁴¹		
9	X					X		No	No	4	2	+ ^j	4.2	2005-07	1,301	Reger 2009 ⁵³		
10	X	X	X	X	X	X	X	No	No	4	2	36	3.2	2001-03	76,476	Riddle 2007 ¹⁴		
11a ⁿ	X					X		Yes	Yes	5	3	58	5.0	2005-07	2,454	Riviere 2011 ¹²		
11b ⁿ	X					X		Yes	Yes	5	3	71	7.3	2005-07	1,415	Riviere 2011 ¹²		
12a ^o	X					X		Yes	Yes	4	3	62	16.0	2004-07	4,723	Thomas 2010 ¹³		
12b ^o	X					X		Yes	Yes	4	3	62	15.7	2004-07	3,749	Thomas 2010 ¹³		
12c ^o	X					X		Yes	Yes	4	3	62	11.5	2004-07	2,607	Thomas 2010 ¹³		

Study ID	Branch ^a							Deployment status ^b				RR ^f %	Preval ^g %	Years ^h	(n) ⁱ	Citation
	A	N	AF	M	CG	PD	CD	ND	Anonymous ^c	Measured ^d	RP ^e					
12 ^d	X					X			Yes	4	3	62	15.9	2004-07	1,501	Thomas 2010 ³³
13	X						X		Yes	3	2	91	15.9	2004	1,090	Wamer 2007 ⁵⁴
14	X	X	X	X		X	X		Yes	7	1	70	17.6	1995	16,193	Bray 1995 ¹⁵
15	X	X	X	X		X	X		Yes	7	1	59	16.1	1998	17,264	Bray 1999 ¹⁶
16	X	X	X	X		X	X		Yes	8	1	56	18.8	2002/03	12,756	Bray 2003 ¹⁷
17	X	X	X	X		X	X		Yes	7	1	52	22.3	2005	16,146	Bray 2006 ¹⁸
18	X	X	X	X	X	X	X		Yes	7	1	72	21.1	2008	28,546	Bray 2009 ¹⁹
19	X						X		Yes	5	3	++ ^p	5.0	2004	2,064	MHAT-II 2005 ²⁰
20	X					X	X		Yes	5	3	++ ^p	8.0	2005	1,124	MHAT-III 2006 ²¹
21	X			X		X	X		Yes	5	3	++ ^p	7.7	2006	1,767	MHAT-IV 2006 ²²
22	X					X	X		Yes	5	3	++ ^p	7.2	2007	3,114	MHAT-V 2008 ²³
23	X					X	X		Yes	5	3	++ ^p	4.8	2009	1,360	MHAT-VI 2009 ^{24,25}
24	X					X	X		Yes	5	3	++ ^p	4.9	2008/09	2,442	MHAT-VI 2009 ^{26,27}
25	X			X		X	X		Yes	5	3	++ ^p	4.8	2010	1,246	MHAT-VII 2011 ²⁷
26 ^g	X	X	X	X		X		No	No	2	2	44	10.8	2007/08	1,041	Schell 2008 ²⁸

^a A: Army; N: Navy; AF: Air Force; M: Marines; CG: Coast Guard.

^b PD: Previously deployed; CD: Currently deployed; ND: Never deployed. 83.8% of assessments across studies were from respondents who had previously deployed, 14.4% never deployed and 1.8% currently deployed.

^c See the text for the definition of anonymity.

^d Patient Health Questionnaire (PHQ)-2 = 1, PHQ-8 = 2, PHQ-9 = 3, PHQ-9/DSM-IV = 4, PHQ-9/DSM-IV + Functional impairment (FI) = 5. Center for Epidemiologic Studies Depression Scale (CES-D) = 6, three-item version A Burnam depression screen = 7, eight-item Burnam depression screen = 8. (Information on the cut-points used for these scales is available on request.)

^e RP: Recall period: one week = 1, two weeks = 2, one month = 3.

^f RR: Response rate. Although RR is reported based on the information provided in the publications, it should be noted that there is wide variation in the ways response rates are reported in the literature. Even though standards for reporting response rates exist (http://www.aapor.org/Response_Rates_An_Overview1.htm), few of the studies included in our analysis reported information about their response rate calculation methods in enough detail to tell which of these definitions they used.

^g Prev: Prevalence of major depression.

^h Years: Year(s) when data were collected.

- ⁱIf available, the sample size is reported as the number of respondents actually screened for depression (which in some cases is different from the total number of respondents included in analyses in the studies).
- ^jThis is an estimate based on personal communication with Dr. Hoge (September 20, 2011).
- ^k3a is a subsample of respondents surveyed after return from Operation Enduring Freedom, 3b from Operation Iraqi Freedom and 3c from other locations.
- ^lStudies 3 and 7–9 are based on the mandatory Post-deployment Health (Re)Assessment or equivalents and response rates were not reported.
- ^m8a and 8b are longitudinal assessments of a sample of active duty soldiers (approximately six months between assessments); 8c and 8d are equivalent for a sample of National Guard and Reserve soldiers.
- ⁿThe data used in Riviere et al.¹² (collected at 3 months [11a] and 12 months [11b] post-deployment) are a subsample of the data used in Thomas et al.,¹³ but the former is used to calculate ORs because it reports additional information about prevalence by demographic characteristics.
- ^oCross-sectional samples of Active Component and National Guard soldiers were assessed at 3 months (12a and 12c, respectively) and 12 months (12b and 12d, respectively).
- ^pNo response rates are available for studies 19–25 (Dr. Bliese, personal communication, August 29, 2011).
- ^qRespondents who were discharged or retired were excluded.

Note: With the exception of studies 19–22, the studies used probability sampling methods to select samples.

Table II

Sensitivity and specificity of screening measures compared to diagnostic interviews

Measures ^a	Sensitivity %	Specificity %	(n)	Criterion ^d	Reference
PHQ-2	83	90	580 ^b	Overview of SCID-DSM-III-R and diagnostic questions from PRIME-MD, major depressive disorder, past month	Kroenke et al. (2003) ³⁰
PHQ-9	88	88	580 ^b	Overview of SCID-DSM-III-R and diagnostic questions from PRIME-MD, major depressive disorder, past month	Kroenke et al. (2001) ³²
PHQ-9/DSM-IV	77	94	3,001 ^b	Formal interview based on DSM-IV; SCID-DSM-IV; Overview of SCID-DSM-III-R and diagnostic questions from PRIME-MD, major depressive disorder ^c	Wittkamp et al. (2007) ⁵⁹
Burnam-3A	81	95	3,116 ^d	DIS-DSM-III, Major depressive episode or dysthymia, past year	Rost et al. (1993) ³⁵
Burnam-8 ^c	77	97	3,116 ^d	DIS-DSM-III, Major depressive episode or dysthymia, past year	Rost et al. (1993) ³⁵
20 item CES-D	80	71	425 ^b	SCID-DSM-III-R, major depressive disorder, past month	Fechner-Bates et al. (1994) ⁶⁰

^aPHQ: Patient Health Questionnaire; CES-D: Center for Epidemiologic Studies Depression Scale; PRIME-MD: Primary Care Evaluation of Mental Disorders; SCID: Structured Clinical Interview for DSM-IV/DSM-III-R; DIS: Diagnostic Interview Schedule. The PHQ-8 and PHQ-9 with severity coding scheme have very similar operating characteristics⁶¹ and are highly correlated⁶² and are thus treated as equivalent. To our knowledge, there are no studies that examined the sensitivity and specificity of the PHQ-9/DSM-IV plus self-reported functional impairment at the cut point used in this study.

^bData were collected in primary care clinics.

^cPooled results from four studies that compared the PHQ-9/DSM-IV with various structured interviews using a random effects model. Only one study provided the time frame of the criterion, which was past month.

^dData were collected in a community sample.

Association between anonymity of survey, deployment status and type of measure and prevalence of major depression using random effects models with logit links

Table III

	Model A. DV: Prevalence as reported in articles ^a		Model B. DV: Prevalence adjusted based on calibration of measures ^a		Model C. DV: Prevalence adjusted based on calibration of measures ^a	
	OR	(95% CI)	OR	(95% CI)	OR	(95% CI)
Intercept	0.0*	(0.0–0.0)	0.0*	(0.0–0.0)	0.0*	(0.0–0.0)
Anonymous						
Yes	3.1*	(1.7–5.6)	2.9*	(1.6–5.3)	3.0*	(1.9–4.8)
No	1.0	--	1.0	--	1.0	--
Measure						
3-item Burnam	5.9*	(3.0–11.5)	1.0	(0.5–1.9)		
8-item Burnam	6.3*	(2.5–16.2)	0.9	(0.6–3.8)		
CES-D	24.9*	(8.2–75.4)	3.2*	(1.1–9.6)		
PHQ-2	2.9*	(1.1–7.5)	0.6	(0.2–1.6)		
PHQ-8	5.0*	(1.6–15.6)	0.9	(0.3–2.8)		
PHQ-9	5.8*	(2.5–13.3)	0.9	(0.4–2.1)		
PHQ-9/DSM-IV	1.9	(1.0–3.8)	0.8	(0.4–1.6)		
PHQ-9/DSM-IV+FI	1.0	--	1.0	--	1.0	--
Deployment status						
Currently	2.7*	(1.4–5.2)	2.2*	(1.2–4.1)	2.3*	(1.4–3.6)
Previously	3.2*	(2.6–3.9)	2.5*	(2.0–3.1)	2.5*	(2.0–3.1)
Never	1.0	--	1.0	--	1.0	--

^aThe 8 coefficients associated with type of measure differ significantly from each other in Model A ($\chi^2_7 = 37.1, p < 0.001$) but not Model B ($\chi^2_7 = 2.2, p = 0.95$), suggesting that the calibration, which was based on data independent of the studies analyzed here, succeeded in correcting for between-scale differences in concordance with clinical diagnoses.

Table IV

Socio-demographic correlates of major depression: Weighted average and range across studies of within-study odds-ratios^a

	Weighted average	Minimum	Maximum	Number of studies	(n)
Gender^b				8	
Female	1.6	1.1	1.9		42,982
Male	1.0				135,194
Race-ethnicity^c				7	
African-American, non-Hispanic	1.1	0.7	1.4		26,617
Hispanic	1.1	0.9	1.3		23,572
Other	1.2	1.0	1.3		7,433
White, non-Hispanic	1.0				116,289
Education				6	
High school or less	3.0	2.0	3.6		63,616
Some college	1.8	1.6	2.1		60,529
College graduate or higher	1.0				43,236
Age^d				8	
24/25 or younger	2.0	1.0	2.2		53,022
25/26 or older	1.0				125,135
Marital status^e				7	
Not married	1.8	1.0	2.0		65,596
Married	1.0				105,028
Rank				6	
Enlisted	2.8	1.9	3.6		129,648
Warrant officer	1.1	0.9	1.2		4,319
Commissioned officer	1.0				33,414
Service^f				6	
Army	2.0	1.6	2.1		58,279
Navy	1.7	1.3	1.8		36,942
Marine Corps	2.0	1.4	2.3		22,985
Air Force	1.0				45,319

^aBetween six and eight studies were used to examine each of the seven correlates. Six studies were the same for all seven correlates,^{14–19} The other studies used varied across correlates, as described in the following notes.

^bFor gender, we used the post-deployment sample of Luxton et al.²⁷ and the combined three and 12-month assessment of Riviere et al.¹²

^cThe race-ethnicity categories reported here are the ones used in the largest set of studies.^{15–19} One other study used the categories 'Caucasian', 'African-American', 'Hispanic' and 'Asian' (we coded the latter as 'other'),²⁴ while another used the categories 'White non-Hispanic', 'Black non-Hispanic' and 'other'.¹⁴ In this case, we coded the category 'other' as 'Hispanic' because the majority of this group is assumed to be Hispanic.

^dThe largest set of studies used here^{15–19} provided information on age ≤ 25 and ≥ 26 , whereas the other studies^{12, 14, 24} provided information on age ≤ 24 and ≥ 25 .

^eAlso based on Riviere et al.¹² 'Not married' includes 'never married', 'divorced/widowed' and 'single' (and 'separated' in the largest set of studies,^{15–19} this is unclear for the other studies). In one study,¹⁵ 'married' includes living in a marriage-like relationship, whereas in others^{16–19} only legally married personnel were included as 'married' (this is unclear for the rest of the studies).

^fIn Riddle et al.¹⁴ Navy and Coast Guard are combined. Note. Within each study ORs were calculated and then a weighted average OR was calculated across studies.

Simulated lifetime, 12-month and past 30 days prevalence estimates of DSM-IV/CIDI major depressive episode and dysthymic disorder in the sub-sample of NCS-R respondents weighted to approximate the population of active duty Army personnel (n = 1785)¹

Table V

	Prevalence		Prevalence of disorder that began prior to enlistment		Prevalence of disorder that began only after enlistment		Proportion of prevalence due to disorder that began prior to enlistment	
	%	(SE)	%	(SE)	%	(SE)	%	(SE)
I. Lifetime prevalence	16.2	(3.1)	11.3	(2.7)	5.5	(1.7)	69.7	(8.5)
II. 12-month prevalence	6.3	(2.1)	5.0	(2.1)	1.4	(0.3)	79.8	(7.9)
III. Past 30 days prevalence	1.3	(0.6)	1.0	(0.5)	0.3	(0.1)	77.9	(11.3)

¹ Respondents are all ages 18 to 65 years, have at least a high school education and are employed with health insurance to match the broad socio-demographic profile of Army personnel. All NCS-R respondents who were ever convicted of a felony or served at least one year in prison were excluded from the sample. All NCS-R respondents with handicaps or physical or mental disorders that would normally lead to rejection from Army enlistment or discharge were excluded from the sample. The handicaps included deafness, blindness, paralysis (of one or both arms, legs or sides of the body) and a missing limb (hand, foot, arm or leg). The physical disorders included cardiovascular disorders (heart attack, stroke, hypertension, heart disease), respiratory disorders (COPD, asthma), diabetes, ulcer, HIV or AIDS, epilepsy or seizure disorder, Crohn's disease, cancer (except skin cancer), severe migraines and extreme obesity. The mental disorders included schizophrenia and other non-affective psychoses, bipolar I disorder, and serious suicide attempts that occurred before the imputed age of enlistment.