An Examination of Potential Misclassification of Army Suicides: Results from the Army Study to Assess Risk and Resilience in Servicemembers

Kenneth L. Cox, MD, MPH, Matthew K. Nock, PhD, Quinn M. Biggs, PhD, MPH, Jennifer Bornemann, MSSW, Lisa J. Colpe, PhD, MPH, Catherine L. Dempsey, PhD, MPH, Steven G. Heeringa, PhD, James E. McCarroll, PhD, MPH, Tszhin Ng, MPH, Michael Schoenbaum, PhD, Robert J. Ursano, MD, Bailey G. Zhang, MS, and David M. Benedek, MD, On behalf of the Army STARRS Collaborators

Debate continues about the accuracy of military suicide reporting due to concerns that some suicides may be classified as accidents to minimize stigma and ensure survivor benefits. We systematically reviewed records for 998 active duty Army deaths (510 suicides; 488 accident, homicide, and undetermined deaths; 2005-2009) and, using research criteria, reclassified 8.2% of the nonsuicide cases to definite suicide (1), suicide probable (4), or suicide possible (35). The reclassification rate to definite suicide was only 0.2% (1/488). This low rate suggests that flagrant misclassification of Army deaths is uncommon and surveillance reports likely reflect the “true” population of Army suicides.

Suicide is now the second-leading manner of death among U.S. military servicemembers, exceeded only by accident deaths (including traffic fatalities; Corr, 2014). The increase in military suicides has been especially prominent in the U.S. Army over the past decade (Nock et al., 2013). Historically, the U.S. Army had enjoyed a considerably
lower suicide rate than that in the general U.S. population, but in 2008, the Army rate surpassed the civilian rate and continued to climb. Seeking ways to reverse this trend, the Army funded the Army Study to Assess Risk and Resilience among Servicemembers (Army STARRS; www.armystarrs.org), a multicomponent epidemiological and neurobiological study of risk and resilience factors for suicidality and its psychopathological correlates among Army personnel. The details of the Army STARRS are described elsewhere (Kessler et al., 2013; Ursano et al., 2014).

Past studies found suicides were underreported or misclassified in various administrative data sources. One study reviewed military deaths from 1998 to 1999 and found 21% more suicides than had been officially reported (Carr, Hoge, Gardner, & Potter, 2004). Accurate and complete identification of suicide cases is essential to generate more complete knowledge about suicidal behaviors, inform policy, and allow the Army to implement effective prevention and treatment programs.

In this article we describe a component of Army STARRS that used criminal investigation reports as a proxy for a psychological autopsy. Psychological autopsy is one systematic method for determining and characterizing the psychological and contextual circumstances of a suicide. Information is obtained from a wide variety of sources including medical records and interviews with people who knew the decedent (Cavanagh, Carson, Sharpe, & Lawrie, 2003; Conner et al., 2011). Such detailed information makes it easier to determine whether the individual intended to die, and facilitates the correct classification of deaths.

This study was primarily descriptive in nature and did not test any a priori hypotheses. The main goals were to: (1) validate Army criminal investigation reports of soldier deaths classified as suicide and (2) apply research criteria to soldier deaths originally classified as accident, undetermined, or homicide to see whether any of these deaths might have been suicides. We discuss differences between the manners of death assigned by Army criminal investigators with those determined by researchers, describe characteristics of those deaths classified as definite suicides by researchers, and discuss any possible differential effects on the overall Army suicide rate.

METHODS

Sample

The sample consisted of all closed, final U.S. Army Criminal Investigation Command (USACIDC, hereafter CID) Reports of Investigation (ROIs) for soldier deaths in the period of 2005-2009 where the assigned manner of death was suicide, accident (including traffic fatalities), homicide, or undetermined. Cases classified as natural death were not included in the sample. Open cases, where a final determination had yet to be made, were not reviewed, although virtually all cases for this historical time period were closed at the time of the data abstraction. As this substudy of Army STARRS consisted of a record review of deceased individuals, it was determined to be an exempt protocol by the Uniformed Services University institutional review board (IRB)–the IRB of record for Army STARRS.

Data Source

CID special agents investigate most soldier deaths to determine whether there was any criminal element. Some deaths are outside of CID’s jurisdiction (e.g., Army National Guard or U.S. Army Reserve soldiers who die while not on active duty). Infrequently, some deaths go uninvestigated due to resource shortfalls or other issues. Suspected suicide cases are initially classified as undetermined. A final determination of suicide is made only after taking into account all available evidence (interviews, documents, forensic laboratory results, autopsy, medical examiner or coroner opinion, etc.). Generally, the CID classification
matches the medical opinion; that is, the manner of death reported on the death certificate. When the results of a CID death investigation differ from the medical opinion, CID policy requires the matter be referred to a formal Manner of Death Review Board (MDRB) for resolution. MDRBs are uncommon; for example, there were none in the period of December 10, 2013, through December 9, 2014 (Department of Defense, 2015). In the absence of a medical opinion, the special agent makes a best judgment determination based on all available information.

While not duplicating all aspects of a formal psychological autopsy, military investigating agents often interview a wide range of informants including close relatives, friends, coworkers, supervisors, First Sergeants, commanders, and eye witnesses. The content of these interviews, both questions and answers, is captured verbatim in a ROI. The ROIs often contain demographic, family, health, autopsy, toxicology, and other relevant life-event information, much of which is not available in any other single military administrative data system. ROIs, maintained at the CID Crime Records Center, formed the primary data source for this study.

Data Collection

A team of four study-trained researchers with at least a master’s level education reviewed all available 2005-2009 ROIs (998) involving suicides (510), accident deaths including traffic fatalities (426), homicides (14), and undetermined cases (48). A few of the ROIs involved multiple deaths; for example, a military vehicle rolled over into an irrigation canal and three occupants drowned. In these cases, the researchers abstracted personal information for only the most relevant fatality; for example, the driver. Demographic and military history information not available from the ROI was obtained from the Defense Manpower Data Center and the Department of Defense Suicide Event Report (DoDSER) system. Informant disagreement in the CID case files was not uncommon; for example, one individual reported a history of a suicide attempt, while others did not. Given that individuals do not confide in everyone, the reviewers did not use a preponderance of evidence approach and instead recorded any positive response as well as the number and type of individuals who made the same report; for example, spouse, friend, and coworker.

Data were entered into Blaise® data collection software. A copy of the full data collection instrument is available at www.armystarrs.org. Abstracted information included demographics, military history (rank, occupation, duty status), death specifics (method used if suicide, autopsy findings, toxicology results), past suicidal behavior, emotional and physical problems, risk factors and life stressors in the 12 months preceding death, intervention and treatment history, and a modified version of the Suicide Intent Scale (SIS; Beck, Beck, & Kovacs, 1975). The original SIS questions were designed to collect information that would help determine whether the survivor of a nonfatal suicide attempt had truly intended to die. Consequently, not all questions in the SIS were suitable for postmortem data collection, and the wording of the retained questions was revised. The modified SIS helped reviewers systematically organize information related to intent, but was not used in any quantitative fashion.

Case Classification

The reviewers piloted the data collection instrument using CID 2004 records. This allowed the reviewers to learn the organization of the physical records and improve the instrument. Additionally, the reviewers developed descriptive vignettes to help define two special classifications, suicide probable and suicide possible. Cases classified as suicide probable were those in which there was clear evidence of suicidal intent leading up to the soldier’s death, but insufficient evidence to say with 100% certainty that it was a suicide (e.g., statements about life being over
leading up to the subject approaching police officers while wielding a gun in an apparent “suicide by cop”). Cases classified as suicide possible were those in which there was some evidence of suicidal intent leading up to the death, but the evidence was equivocal (e.g., the presence of depression and statements about wanting to be dead in the weeks before an overdose of oxycodone). Generally, the reviewers found it difficult to identify consistent and reproducible criteria to use for suicide probable cases, resulting in more reclassifications to suicide possible. The distinction between these two additional categories was necessarily subjective, so all suicide probable or possible cases were reviewed by at least two reviewers.

After reviewing all available material, the reviewer assessed whether there was evidence of intent to die, and then assigned a manner of death to the case (natural causes, suicide, suicide probable, suicide possible, homicide, accident, or undetermined). The most compelling component of the modified SIS that influenced the reviewers to classify a case as a definite suicide was the presence of a suicide note. Although there have been reports of faked suicide notes (Ho, Yip, Chiu, & Halliday, 1998; Pestian, Nasrallah, Matykiewicz, Bennett, & Leenaars, 2010), all of the suicide notes in the CID case files appeared genuine.

The primary reviews were not blinded; that is, the reviewers knew the CID classification. Due to the nature of the CID physical records, blinding would have been extremely cumbersome and expensive. To promote consistent abstraction of clinical information and case classification, a subset of cases underwent a second review by another data coder. This allowed the researchers to assess interrater reliability (IRR) among the four primary data coders. There were second reviews of a random selection of approximately 25% of suicide cases and 25% of accident cases. All (100%) of the ambiguous cases underwent secondary review. The ambiguous cases included all undetermined cases and all cases involving reclassification of a CID nonsuicide case to suicide, suicide probable, or suicide possible. Secondary reviews were not blinded. The second reviewer was aware of both CID’s original classification and the decision of the primary reviewer. The secondary reviews consisted of a full review of the ROI, but data abstraction was limited to basic identifying information and six primary domains (prior suicidality, emotional and physical health problems, life events, treatment and intervention for physical or emotional problems, suicidal intent, and the assigned manner of death). These domains contained questions that often required the data coder to make a judgment; for example, was there evidence of depression? The primary and secondary reviewers compared their results for a given case and resolved any differences as described in the next section. Each reviewer’s summary of evidence related to intent was a key part of the reconciliation process. When the two reviewers could not agree, an uncommon situation, one of the other two data coders was asked to render an opinion. Majority rule determined the “best response” for individual item responses other than the manner of death. When STARRS coders did not agree on the manner of death, decision rules favored identifying the case as having been misclassified. For example, if the CID classification was “accident” and one STARRS coder determined the manner of death to be “accident” while a second STARRS coder identified the case as “suicide possible,” the case was reclassified by the research team as “suicide possible” (additional details about the decision rules are available upon request and at www.armystarrs.org).

RESULTS

Classification of Manner of Death

Trained researchers reviewed 998 death investigations by the CID. Of 510 cases classified as suicide by the CID, the researchers assessed the evidence as less than conclusive of definite suicide in 23
cases. In five instances, reviewers felt suicide was probable with the other 18 cases classified as suicide possible. Thus, the down-classification rate was 4.5% (23/510). Of the 488 cases the CID determined to be nonsuicides, reviewers reclassified 35 cases to possible suicide, four cases to probable suicide, and one case to definite suicide. The overall up-classification rate for nonsuicide cases was 8.2% (40/488). This resulted in a total of 488 definite suicide cases (510-23+1), as classified by the reviewers.

The researchers reclassified 26 of the 426 CID accident cases to either suicide \( (n = 1) \) or suicide possible \( (n = 25) \). It was often difficult to determine intent in cases of drug overdose, especially in the absence of a suicide note.

Of 14 CID cases of homicide, the researchers reclassified four as suicide probable \( (n = 2) \) or suicide possible \( (n = 2) \). An additional case was reclassified as an accident. In several of these cases, there was sufficient evidence to suggest the individual acted deliberately in an effort to provoke a lethal response from law enforcement personnel (i.e., “suicide by cop”). Whereas justifiable homicide was the correct classification from a legal standpoint, one could consider these cases as a form of suicide.

The 48 CID undetermined cases presented the greatest challenge. A medical examiner or coroner must classify the case as undetermined “when the information pointing to one manner of death is no more compelling than one or more other competing manners of death when all available information is considered” (Centers for Disease Control, 2003). Neither they nor the CID special agents have the options of suicide probable or suicide possible. The researchers, unconstrained by these guidelines, reclassified 16 undetermined causes as follows: suicide probable \( (n = 2) \), suicide possible \( (n = 8) \), and accident \( (n = 6) \). Many of the undetermined cases involved overdoses. In the absence of a suicide note, it was difficult to delineate between an accidental overdose and intentional self-harm by a known recreational drug user.

The overall kappa \( (\kappa) \) statistic (Landis & Koch, 1977) for the assigned manner of death was 0.95.

**Temporal Patterns of the Classification of Manner of Death**

Classification patterns can vary over time due to new guidelines, increased emphasis on compliance, more exacting quality control measures, closer scrutiny, or other causes. Although no strong time-varying trends emerged from this study, there was a modest increase in the down-classification of suicide to suicide probable or possible for 2005-2009. Similarly, the up-classification of undetermined cases to any type of suicide decreased modestly over time. There is one seemingly anomalous year (2008) where the researchers did not reclassify any accident or undetermined cases to any type of suicide (Table 1).

**DISCUSSION**

Suicides will likely remain underrepresented in administrative databases. This is not surprising given how difficult it can be to determine the manner of death. Past studies suggested deliberate misclassification occurred to ensure benefits to survivors, avoid the perception of dishonor, or for other reasons that may have increased misclassification rates (Bohnert et al., 2013; Carr et al., 2004; Claassen et al., 2010; Eaton, Messer, Garvey Wilson, & Hoge, 2006; Palmer et al., 2014; Tøllefsen, Hem, & Ekeberg, 2012). There would seem to be less need now for such deliberate misclassification given various changes over the past decade in business practices (e.g., Service-members Group Life Insurance is paid to survivors regardless of the manner of death), societal views, and other factors.

We found misclassification rates lower than those previously reported in 1998-1999 military populations (Carr et al., 2004), suggesting that misclassification of military suicides may have substantially diminished in
recent years. However, the earlier study used data from the Armed Forces Medical Examiner System (AFMES) and DoD casualty reports, not criminal investigation data. Also, the earlier study included all service-member deaths, whereas this study involved only deaths of U.S. Army soldiers.

Differences in the classification of cases should not be construed as errors by either CID or the involved medical examiners and coroners. The researchers used a unique classification system reflecting clinical and academic perspectives. CID and the medical examiners are bound by different rules and constraints. The option of down-classifying suicide cases to suicide probable or possible is not available to medical examiners or law enforcement professionals. The up-classification of other manners of death to possible or probable suicide is of interest as it helps evaluate the possibility of underestimating suicides due to existing laws, policies, and practices pertaining to civilian law enforcement, Army CID investigations, and medical examiner activities (civilian and military).

Although no strong time-varying trends emerged from this study, there was a slight increase in the down-classification of suicide to suicide probable or possible for 2005-2009. This could indicate that over time CID has become more likely to classify difficult or borderline cases as suicide. Similarly, the up-classification of undetermined cases to any type of suicide decreased slightly over time. This, too, could reflect an increased tendency to classify some of these difficult cases as suicide.

The primary strength of this study was the large number of suicides and other deaths available for review. Additionally, most CID case files often contained large amounts of detailed, relevant information. There were some challenges in using law enforcement records as a proxy for a psychological autopsy. Law enforcement interviews seek factual information to decide whether a crime has occurred and do not always produce an exhaustive list of personal risk or protective factors and precipitating stressors. Questions were more likely to address previous suicidal thoughts, the victim’s intent and personal actions at the time of death, substance misuse, and access to mental health clinical services. Information about social support, resilience, coping strategies, or specific mental health diagnoses was less common. To make things more difficult, it was not uncommon for informants to contradict each other. Another limitation to consider is that the reviewers were not blinded to either CID’s or the paired reviewer’s classification of the case. Lastly, this study involved only U.S. Army soldiers and the results may not reflect patterns among other branches of the U.S.

### Table 1

<table>
<thead>
<tr>
<th>Type of Reclassification</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>Overall (2005–09)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suicide to Suicide Probable or Suicide Possible</td>
<td>3.8</td>
<td>2.4</td>
<td>3.9</td>
<td>5.6</td>
<td>5.4</td>
<td>4.5</td>
</tr>
<tr>
<td>Suicide to Accident or Traffic</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Accident or Traffic to Any Suicide</td>
<td>6.3</td>
<td>4.1</td>
<td>6.8</td>
<td>0.0</td>
<td>10.7</td>
<td>6.1</td>
</tr>
<tr>
<td>Homicide to Any Suicide</td>
<td>33.3</td>
<td>0.0</td>
<td>0.0</td>
<td>40.0</td>
<td>20.0</td>
<td>28.6</td>
</tr>
<tr>
<td>Homicide to Accident</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>20.0</td>
<td>0.0</td>
<td>7.1</td>
</tr>
<tr>
<td>Undetermined to Any Suicide</td>
<td>37.5</td>
<td>37.5</td>
<td>27.3</td>
<td>0.0</td>
<td>10.0</td>
<td>20.8</td>
</tr>
<tr>
<td>Undetermined to Accident or Traffic</td>
<td>0.0</td>
<td>0.0</td>
<td>27.3</td>
<td>18.2</td>
<td>10.0</td>
<td>12.5</td>
</tr>
<tr>
<td>Any NonSuicide to Any Suicide</td>
<td>10.0</td>
<td>7.3</td>
<td>9.0</td>
<td>2.2</td>
<td>11.0</td>
<td>8.2</td>
</tr>
</tbody>
</table>

*a* Any Suicide includes suicide, suicide probable, and suicide possible.

*b* NonSuicide cases include accident or traffic, homicide, and undetermined.
CONCLUSIONS

In this study we found a relatively low rate of potential misclassification of definite suicides in USACID reports of investigation. The results suggest that misclassification of suicides may have diminished in recent years. A number of factors may have contributed to a decline in misclassification, and further study of the degree to which these factors influenced classification has implications for the training of law enforcement personnel, medical examiners, and coroners. Most importantly, the identified reclassification rate would have had a negligible effect on the overall Army suicide rate and certainly could not account for the increasing Army suicide rate from 2005 to 2009. The low rate of misclassification provides reassurance that manner of death classification by medical examiners and military criminal investigators likely reflect the “true” population of military suicides. Despite some variation in quality and completeness, information contained within military criminal investigation records revealed consistencies with other psychological autopsies that suggest that such records can be a useful and cost-efficient proxy for a formal psychological autopsy.

REFERENCES


review of psychosocial risk and protective factors. Psychiatry, 76, 97–125.


APPENDIX 1: ON BEHALF OF THE ARMY STARRS COLLABORATORS:
FUNDING SUPPORT

Army STARRS was sponsored by the Department of the Army and funded under cooperative agreement number U01MH087981 with the U.S. Department of Health and Human Services, National Institutes of Health, National Institute of Mental Health (NIH/NIMH).

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 Veterans Administration, Department of the Army, or the Department of Defense.

ROLE OF THE SPONSORS

Although a draft of this manuscript was submitted to the Department of Defense, the Department of the Army, and NIMH for review and comment prior to submission, this was with the understanding that comments would be no more than advisory.

ADDITIONAL CONTRIBUTIONS

The Army STARRS Team consists of coprincipal investigators: Robert J. Ursano, MD (Uniformed Services University of the Health Sciences) and Murray B. Stein, MD, MPH (University of California San Diego and VA San Diego Healthcare System); Site Principal Investigators: Steven Heeringa, PhD (University of Michigan) and Ronald C. Kessler, PhD (Harvard Medical School); National Institute of Mental Health (NIMH) collaborating scientists: Lisa J. Colpe, PhD, MPH and Michael Schoenbaum, PhD; Army liaisons/consultants: COL Steven Cersovsky, MD, MPH (USAPHC (Provisional)) and Kenneth Cox, MD, MPH (USAPHC (Provisional)); other team members: Pablo A. Aliaga, MA (Uniformed Services University of the Health Sciences); COL David M. Benedek, MD (Uniformed Services University of the Health Sciences); K. Nikki Benevides, MA (Uniformed Services University of the Health Sciences); Paul D. Bliese, PhD (University of South Carolina); Susan Borja, PhD (NIMH); Evelyn J. Bromet, PhD (Stony Brook University School of Medicine); Gregory G. Brown, PhD (University of California San Diego); Laura Campbell-Sills,
PhD (University of California San Diego); Catherine L. Dempsey, PhD, MPH (Uniformed Services University of the Health Sciences); Carol S. Fullerton, PhD (Uniformed Services University of the Health Sciences); Nancy Gebler, MA (University of Michigan); Robert K. Gifford, PhD (Uniformed Services University of the Health Sciences); Stephen E. Gilman, ScD (Harvard School of Public Health); Marjan G. Holloway, PhD (Uniformed Services University of the Health Sciences); Paul E. Hurwitz, MPH (Uniformed Services University of the Health Sciences); Sonia Jain, PhD (University of California San Diego); Tzu-Cheg Kao, PhD (Uniformed Services University of the Health Sciences); Karestan C. Koenen, PhD (Columbia University); Lisa Lewandowski-Romps, PhD (University of Michigan); Holly Herberman Mash, PhD (Uniformed Services University of the Health Sciences); James E. McCarroll, PhD, MPH (Uniformed Services University of the Health Sciences); James A. Naifeh, PhD (Uniformed Services University of the Health Sciences); Tsz Hin Hinz Ng, MPH (Uniformed Services University of the Health Sciences); Matthew K. Nock, PhD (Harvard University); Rema Raman, PhD (University of California San Diego); Holly J. Ramsawh, PhD (Uniformed Services University of the Health Sciences); Anthony Joseph Rosellini, PhD (Harvard Medical School); Nancy A. Sampson, BA (Harvard Medical School); LCDR Patcho Santiago, MD, MPH (Uniformed Services University of the Health Sciences); Michael L. Thomas, PhD (University of California San Diego); Leming Wang, MS (Uniformed Services University of the Health Sciences); Christina L. Wassel, PhD (University of Vermont); Simon Wessely, FMedSci (King’s College London); Christina L. Wryter, BA (Uniformed Services University of the Health Sciences); Hongyan Wu, MPH (Uniformed Services University of the Health Sciences); LTC Gary H. Wynn, MD (Uniformed Services University of the Health Sciences); and Alan M. Zaslavsky, PhD (Harvard Medical School).