



Published in final edited form as:

*Stud Health Technol Inform.* 2017 ; 238: 128–131.

## Using Structured and Unstructured Data to Refine Estimates of Military Sexual Trauma Status Among US Military Veterans

Adi V. GUNDLAPALLI<sup>a,b,1</sup>, Emily BRIGNONE<sup>c</sup>, Guy DIVITA<sup>a</sup>, Audrey L. JONES<sup>a</sup>, Andrew REDD<sup>a</sup>, Ying SUO<sup>a</sup>, Warren B. P. PETTEY<sup>a</sup>, April MOHANTY<sup>a</sup>, Lori GAWRON<sup>a</sup>, Rebecca BLAIS<sup>c</sup>, Matthew H. SAMORE<sup>a</sup>, and Jamison D. FARGO<sup>b,c</sup>

<sup>a</sup>VA Salt Lake City Health Care System & University of Utah, Salt Lake City, UT, USA

<sup>b</sup>National Center for Homelessness Among Veterans, Philadelphia, PA, USA

<sup>c</sup>Utah State University, Logan, UT, USA

### Abstract

Sexual trauma survivors are reluctant to disclose such a history due to stigma. This is likely the case when estimating the prevalence of sexual trauma experienced in the military. The Veterans Health Administration has a program by which all former US military service members (Veterans) are screened for military sexual trauma (MST) using a questionnaire. Administrative data on MST screens and a change of status from an initial negative answer to positive and natural language processing (NLP) on electronic medical notes to extract concepts related to MST were used to refine initial estimates of MST among a random sample of 20,000 Veterans. The initial MST positive screen of 15.4% among women was revised upward to 21.8% using administrative data and further to 24.5% by adding NLP results. The overall estimate of MST status in women and men in this sample was revised from 8.1% to 13.1% using both data elements.

### Keywords

Military Sexual Trauma; Veterans; Disclosure of trauma

### Introduction

US military Veterans are at risk for experiencing sexual trauma while serving in the armed forces. This is referred to as military sexual trauma (MST) and is considered an important public health issue, similar to sexual trauma among adults in the general community and in college campuses. The challenge in estimating the true prevalence of MST is that there is under-reporting of this trauma due to social stigma.

In the Veterans Health Administration (VHA), where former US military service members who are now Veterans are provided care, there is a coordinated effort to screen for MST and provide free medical care to those who report it. The screen is administered to all Veterans seeking care in VHA and the results are recorded in the electronic medical record. Here too,

<sup>1</sup> Corresponding Author. Adi V. Gundlapalli, MD, PhD, MS, [Adi.Gundlapalli@hsc.utah.edu](mailto:Adi.Gundlapalli@hsc.utah.edu).

there is a challenge in that not all Veterans disclose their true MST status due to various factors [1-3]. Official estimates of the prevalence of MST are based on a positive screen for MST administered to all Veterans who seek care in VHA medical facilities; approximately 25% of all women Veterans and 1% of all male Veterans report a positive screen for MST [4,5].

Preliminary work has revealed that the MST status of several Veterans who have screened negative at their initial screen is later changed to positive status in administrative or structured data. Furthermore, we have also noted that some Veterans disclose a history of MST to their mental health or medical provider in the course of regular visits to VHA and if the Veteran has no objection, these disclosures are recorded by the provider in the free text of medical notes (unstructured data) [6].

This aim of this study was to determine whether and by how much structured and unstructured data could add to estimating the prevalence of MST positive status as reported by Veterans to providers in VHA.

## **1. Methods**

### **1.1. Administrative data related to military sexual trauma in VHA**

The US Department of Veterans Affairs (VA) serves nearly 6 million unique Veterans through a network of 152 hospitals around the US. Many of these hospitals have emergency departments and urgent care clinics. Administrative data on MST screens and their results (positive or negative) were extracted from the VA corporate data warehouse using Veterans Informatics and Computing Infrastructure (VINCI), a secure research portal [7,8] on a randomly selected sample of 10,000 women and 10,000 men Veterans from recent wars in Afghanistan and Iraq.

### **1.2. Natural language processing on electronic medical notes to extract MST concepts**

All outpatient electronic medical notes in the 12 months following the date of the visit when the MST screen was first completed were extracted using VINCI. Using an NLP pipeline expressly developed for this project, positively asserted concepts related to MST were extracted from these notes. A total of 13,501 Veterans (6,623 women and 6878 men) were found to have electronic medical notes for processing by NLP during the study period. NLP outputs were reviewed by trained human annotators to determine if the concepts were true or false positive and whether they represented concepts related to MST.

### **1.3. Data analysis**

Using administrative data, the number and percentage of women and men Veterans who had an initial positive or negative MST screen was first determined. Then, during a long follow-up period of up to 14 years, the number and percentage of Veterans who had a change of MST status from negative to positive was determined. Using NLP outputs, the number and percentage of Veterans with evidence of positively asserted NLP concepts related to MST were determined.

## 2. Results

### 2.1. MST initial screen results and switch from negative to positive.

From the random sample of 10,000 women and 10,000 men Veterans, 16,847 Veterans had initial MST screen results available for analysis. Of the 8,442 women, 1,297 (15.4%) reported a positive screen; of the 8,405 men, 62 (0.7%) reported a positive screen. During the follow up period ranging from 4 to 14 years, a total of 541 women (6.4%) had a change of MST status from negative to positive. A small number of men (29, 0.3%) had a change of status recorded in structured data. This resulted in an upward estimate of MST positive screen from 15.4% to 21.8% in this group of women Veterans and from 0.7% to 1.0% for men Veterans. When taken together, the MST status by initial positive status in this group of 16,847 women and men Veterans was 8.1% (1,359 positive screens). When adding the status change just from structured data, the MST status for both women and men is revised upward to 11.5% (total 1,929 positive screens).

### 2.2. NLP on electronic medical notes to extract positively asserted concepts for MST

Electronic medical notes from a total of 13,501 Veterans in the study sample were evaluated using NLP. In this set of 6,623 women Veterans and 6,878 men Veterans, a total of 1,264 Veterans reported a positive screen for MST at their initial screen. This represents a total MST positive screen of 9.4% in this group of Veterans. Using NLP followed by human review of the NLP outputs, a total of 284 (230 women and 54 men) Veterans who were MST screen negative were found to have positively asserted concepts (true positives) related to MST. The overall revised estimate of MST positive status in this group using just the NLP outputs is 11.5% (1,264 + 284 by NLP of 13,501).

### 2.3. Revised estimates from structured and unstructured data mining

Using structured data from MST status changes and evidence from NLP, we were able to revise the estimate of positive MST status in the women in this study sample from 15.4% (1,297 of 8,442) to 24.5% (1,297 + 541 from structured data MST status change + 230 from NLP of 8,442). Similarly, for men in this sample, the revised estimate is 1.7% (62 + 29 from structured data MST status change + 54 from NLP outputs of 8,405). Overall estimate for the study sample is revised from 8.1% to 13.1% using both data elements.

## 3. Discussion

There are inherent challenges in estimating the true prevalence of an extremely sensitive problem such as sexual trauma as victims often are not willing to disclose this history. As such, surveying a large number of Veterans on this topic may not provide any better estimates. Thus, we undertook a study to examine whether mining and analyses of structured and unstructured data would help us refine the estimate of prevalence of MST.

Starting with baseline data on MST positive screen status, mining structured data for MST screen status from negative to positive alone resulted in an upward estimate of MST positive screens by 6.4% in women Veterans, NLP added another 2.7% such that the revised estimate in women Veterans in this study sample was 24.5% (from 15.4%). The change in status

recorded for men Veterans in administrative in this study sample was small. Similarly, a small number of men were noted to have NLP evidence of MST. The revised estimate for men Veterans in this sample is 1.7% based on both structured and unstructured data mining. Using both data elements, the overall estimate for the prevalence of MST positive status among women and men in this study sample was revised upward from 8.1% to 13.1%. This study demonstrates that it is feasible to revise estimates of MST positive status using structured and unstructured data.

The differences in baseline MST positive screen status in women Veterans in our sample (15.4%) versus the 25% reported by VHA sources is likely due to differences in the populations studied. Our study sample consisted of a random sample of 20,000 women and men Veterans from recent conflicts. It is interesting to note that our revised estimate of prevalence of MST status in the women in our sample was 24.5%.

We acknowledge several limitations of this study. It is possible that the structured data elements of the MST screen were not complete and thus we may be underestimating the true prevalence of changes from negative to positive screen. There are false positives and false negatives from the outputs of NLP and this affects the ability to identify relevant concepts. The concordance of MST screen results from administrative data and NLP evidence of MST merits further study. There is also a need for work on determining the feasibility of extracting concepts related to other sexual trauma such as childhood sexual abuse and adult sexual re-victimization using NLP.

## Acknowledgements

We appreciate our colleagues at VA Informatics and Computing Infrastructure (VINCI) for their assistance with accessing VA data. Funding from U.S. Department of Veterans Affairs, Health Services Research & Development grant #IIR12-084. The views expressed are those of the authors and do not necessarily reflect the position or policy of the U.S. Department of Veterans Affairs or the United States Government.

## References

- [1]. Burns B, Grindlay K, Holt K, Manski R, Grossman D. Military sexual trauma among US servicewomen during deployment: a qualitative study. *Am J Public Health*. 2014; 104:345–349. [PubMed: 24328617]
- [2]. Department of Defense, Department of Defense Annual Report on Sexual Assault in the Military Fiscal Year 2016 , in: Department of Defense , ed., Washington DC , 2016.
- [3]. Hoyt T, Klosterman Rielage J, Williams LF. Military sexual trauma in men: a review of reported rates. *J Trauma Dissociation*. 2011; 12:244–260. [PubMed: 21534094]
- [4]. Barth SK, Kimerling RE, Pavao J, McCutcheon SJ, Batten SV, Dursa E, Peterson MR, Schneiderman AI. Military Sexual Trauma Among Recent Veterans: Correlates of Sexual Assault and Sexual Harassment. *Am J Prev Med*. 2015
- [5]. Klingensmith K, Tsai J, Mota N, Southwick SM, Pietrzak RH. Military sexual trauma in US veterans: results from the National Health and Resilience in Veterans Study. *J Clin Psychiatry*. 2014; 75:e1133–1139. [PubMed: 25373123]
- [6]. Divita G, Brignone E, Carter ME, Suo Y, Blais RK, Samore MH. , Fargo JD, , and Gundlapalli AV, Extracting Sexual Trauma Mentions from Electronic Medical Notes Using Natural Language Processing. *Stud Health Technol Inform. MedInfo2017 (In Press)* (2017 ).
- [7]. US Department of Veterans Affairs, VA Informatics and Computing Infrastructure (VINCI), in, US Department of Veterans Affairs , Washington DC , 2017.

- [8]. Abbott DE, Voils CL, Fisher DA, Greenberg CC, Safdar N. Socioeconomic disparities, financial toxicity, and opportunities for enhanced system efficiencies for patients with cancer. *J Surg Oncol.* 2017; 115:250–256. [PubMed: 28105638]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript