

# Combining Natural Language Processing of Electronic Medical Notes With Administrative Data to Determine Racial/Ethnic Differences in the Disclosure and Documentation of Military Sexual Trauma in Veterans

Adi V. Gundlapalli, MS, MD, PhD,\*† Audrey L. Jones, PhD,\*† Andrew Redd, PhD,\*† Guy Divita, MS,\*† Emily Brignone, PhD,\*‡ Warren B.P. Pettey, MPH, CPH,\*† Marjorie E. Carter, MSPH,\*† Matthew H. Samore, MD,\*† Rebecca K. Blais, PhD,\*‡ and Jamison D. Fargo, PhD, MS\*‡

**Background:** Despite national screening efforts, military sexual trauma (MST) is underreported. Little is known of racial/ethnic differences in MST reporting in the Veterans Health Administration (VHA).

**Objective:** This study aimed to compare patterns of MST disclosure in VHA by race/ethnicity.

**Research Design:** Retrospective cohort study of MST disclosures in a national, random sample of Veterans who served in Afghanistan and Iraq and completed MST screens from October 2009 to 2014. We used natural language processing (NLP) to extract MST concepts from electronic medical notes in the year following Veterans' first MST screen.

**Measure(s):** Any evidence of MST (positive MST screen or NLP concepts) and late MST disclosure (NLP concepts following a negative MST screen). Multivariable logistic regressions, stratified by sex, tested racial/ethnic differences in any MST evidence, and late disclosure.

**Results:** Of 6618 male and 6716 female Veterans with MST screen results, 1473 had a positive screen (68 male, 1%; 1405 female, 21%). Of those with a negative screen, 257 evidenced late MST disclosure by

NLP (44 male, 39%; 213 female, 13%). Late MST disclosure was usually documented during mental health visits. There were no significant racial/ethnic differences in MST disclosure among men. Among women, blacks were less likely than whites to have any MST evidence (adjusted odds ratio = 0.75). In the subsample with any MST evidence, black and Hispanic women were more likely than whites to disclose MST late (adjusted odds ratio = 1.89 and 1.59, respectively).

**Conclusions:** Combining NLP results with MST screen data facilitated the identification of under-reported sexual trauma experiences among men and racial/ethnic minority women.

**Key Words:** military sexual trauma, veterans, disclosure of trauma, minority health, natural language processing, informatics

(*Med Care* 2019;57: S149–S156)

Sexual trauma is a major public health issue and has been described in the context of intimate partner relationships, college campuses, clergy, and in workplaces. With the recent increase in news media reports on these topics and a general increase in awareness regarding sexual misconduct in the workplace, which includes both sexual harassment and assault, it is apparent that the problem is widespread.

Differences may exist in risk of experiencing sexual trauma and in subsequent reporting of the abuse among racial/ethnic minorities. For example, studies have described higher rates of occurrence and lower rates of disclosure of sexual trauma among African American women.<sup>1</sup> Because of various sociocultural barriers, including distrust of authorities and victim-blaming myths, the disclosure of sexual trauma by other minorities such as Latina, Asian American, and American Indian women is also likely low (reviewed in Bryant-Davis et al<sup>2</sup>). In addition, the challenges and barriers faced by those who experience sexual trauma are exacerbated for males in terms of reporting and disclosure.<sup>3,4</sup> Most of the knowledge generated on the disclosure of sexual trauma by racial and ethnic minorities is based on small studies of interviews and surveys.

Military sexual trauma (MST) refers to sexual harassment and assault experienced in one of the largest workplaces in the country, the US military. On the basis of a recognition of MST as a significant problem and mandate by Congress,

From the \*Informatics, Decision Enhancement, and Analytic Sciences Center, VA Salt Lake City Health Care System; †Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City; and ‡Department of Psychology, Utah State University, Logan, UT.

Present address: Emily Brignone, PhD, VA Pittsburgh Health Care System. Supported by grant IIR 12-084 (PI: A.V.G.), and VA Center of Innovation Award #150HX001240 from the Health Services Research and Development of the Office of Research and Development of the US Department of Veterans Affairs.

The views expressed in this article are those of the authors and do not necessarily represent the position or policy of the US Department of Veterans Affairs or the US government.

A.L.J. is supported as a VA Office of Academic Affiliations Post-Doctoral Fellow in Medical Informatics at the VA Salt Lake City Health Care System. E.B. is supported as a VA Office of Academic Affiliations Post-Doctoral Fellow in Health Services Research at the VA Center for Health Equity Research and Promotion (CHERP) at the VA Pittsburgh Health Care System. The remaining authors declare no conflict of interest.

Reprints: Adi V. Gundlapalli, MS, MD, PhD, Informatics, Decision Enhancement, and Analytic Sciences (IDEAS 2.0) Center, VA Salt Lake City Health Care System, 500 Foothill Drive, Mail Stop 182, Salt Lake City, UT 84148. E-mail: adi.gundlapalli@hsc.utah.edu.

Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved. ISSN: 0025-7079/19/5706-S149

the Veterans Health Administration (VHA) instituted universal screening for MST in Veterans seeking care in VHA medical facilities in 2004.<sup>5</sup> To our knowledge, this is the only such screen in a large health care system. Given the sensitive and personal nature of sexual trauma, it is possible that Veterans do not disclose their history of MST with a new provider during their initial MST screen.<sup>6</sup> Indeed, some Veterans may deny MST during the initial screening, and later describe sexual trauma experiences once rapport has been developed with their provider in a therapeutic environment. In this case, providers may document MST experiences in medical notes<sup>7,8</sup> without adjusting responses to the initial screening. While prior studies have examined overall rates of MST disclosure and MST-related care, there are no health system-wide evaluations of racial/ethnic differences in disclosure of sexual trauma status among Veterans at the time of an initial screening test or subsequent disclosure to a provider during a medical visit based on race/ethnicity.

Reviewing large sets of medical notes (even electronic) to determine the disclosure and documentation of MST is resource intensive and not practical for research at a population level. Natural language processing (NLP) is a branch of computer science that has been developed to extract relevant concepts from the writings of humans (thus the natural language reference) into a form that can be used for computation and analysis using rules of linguistics. The processing can be automated and scaled to process large sets of documents with efficiency.<sup>9</sup> NLP algorithms have been applied to several clinical disciplines and has shown to be of value in unlocking information from narrative clinical text<sup>10</sup> and for supporting phenotyping of patients when used in conjunction with electronic medical record data.<sup>11</sup> To our knowledge, NLP has not been previously used to understand the differences in the prevalence and patterns of disclosure of sexual trauma experiences in electronic medical notes from large health care systems.

We undertook a proof-of-concept study to determine differences in MST disclosure by racial/ethnic minorities at the time of initial screen and subsequent disclosure during a medical visit (late disclosure) using data from VHA. Specifically our objectives were to (1) assess the feasibility of mining unstructured free text data using NLP and (2) combine NLP concepts with administrative data on MST screen results to perform an innovative analysis in identifying differences among racial/ethnic minorities with regard to sexual trauma disclosure. On the basis of the literature on sexual trauma in racial/ethnic minorities in the community, we hypothesized that black and Hispanic minorities would be less likely than whites to ever disclose MST, but more likely to have a late MST disclosure. We further hypothesized that, among those with a history of MST, mentions of MST would most often occur in mental health visit notes versus primary or specialty care visit notes. Results may help to inform the design of interventions aimed at addressing disparities in mental health.<sup>12,13</sup>

## METHODS

### Setting and Study Population

This retrospective cohort study examined racial/ethnic differences in reporting of MST from a national, random

sample of Veterans who served in recent conflicts as part of Operations Enduring Freedom/Iraqi Freedom/New Dawn (OEF/OIF/OND) and who received care in VHA medical facilities. We selected a random sample of 10,000 female and 10,000 male OEF/OIF/OND Veterans by linking a Department of Defense (DoD) roster of service members who served in the recent conflicts with administrative data stored in the VHA corporate data warehouse. Data were accessed through VA Informatics Computing Infrastructure (VINCI), a secure research portal.<sup>14</sup> We analyzed results of the MST screen that is administered to nearly all Veterans who establish care in the VHA after separation from the military,<sup>5,15,16</sup> as well as electronic medical notes that are available in the VHA corporate data warehouse for use in research.<sup>14</sup> Veterans were included in this study if they received care in VHA medical facilities, had MST screen results from October 2009 to October 2014 (FY2010–2014), and had at least one visit with text notes in the 12 months following their initial MST screen. All study procedures were approved by the University of Utah Institutional Review Board and the Research Review Committee of the VA Salt Lake City Health Care System.

## Measures

### MST Screen

Screening and documentation of MST is implemented through a clinical reminder embedded in VHA's Computerized Patient Record System (CPRS). If a Veteran has not been previously screened, the reminder prompts the clinical provider to administer a 2-item screen: "While you were in the military ... (a) did you receive uninvited and unwanted sexual attention, such as touching, cornering, pressure for sexual favors, or verbal remarks?; and (b) did someone ever use force or threat of force to have sexual contact with you against your will?"<sup>16</sup> With the exception of a "decline" response, in which case the clinical reminder reappears 1 year later, the screen is supposed to be administered only once for a Veteran. The screening items are combined as a single variable in VHA administrative data and cannot be analyzed separately. The screen is considered positive if the Veteran answers in the affirmative to either item. MST status may be updated by a clinician at any subsequent encounter in the event of disclosure of new information.

### Late Disclosure of MST

We followed Veterans that had MST screen results for 12 months after the date of the first MST screen and downloaded all electronic medical notes associated with their outpatient visits in those 12 months. We processed the 12-month set of electronic medical notes using an NLP pipeline, V3NLP, that was developed to extract all sexual trauma concepts such as adverse childhood experiences and mentions of sexual trauma in adulthood (including MST) from VHA electronic medical records.<sup>8</sup> The pipeline is built on the APACHE-Unstructured Information Management Applications (UIMA) framework.<sup>17</sup> In brief, the pipeline first parses the electronic medical note into its component parts such as sections, content headings, lists, sentences, lines, and finally individual concepts (tokens). The individual tokens are

then compared with a look-up dictionary of sexual trauma terms that was developed for this pipeline using terms from standard vocabularies and supplemented with expert opinion. An important feature of the pipeline is to exclude concepts that are negated such as “no evidence of sexual trauma.” Thus, the goal of the pipeline is to extract “positively asserted” concepts. All concepts related to sexual trauma extracted by the NLP pipeline as positively asserted (with clinical text surrounding the extracted concept to provide context, “snippets”) were reviewed by a set of trained human reviewers to determine true and false positivity specifically with regard to MST for this study. The overall positive predictive value (PPV) for identifying positively asserted sexual trauma mentions (not just MST) at the individual concept level has been reported in prior publications as 0.90 (0.95 for females, 0.41 for males)<sup>8</sup>; at the patient level, the overall PPV for sexual trauma mentions is 0.71 (0.82 for females, 0.38 for males). If there were multiple concepts extracted at the document level from either one or multiple documents from a Veteran, a final binary result of Yes/No of evidence of MST by NLP was rolled up to the Veteran level. NLP evidence of MST in electronic medical notes was considered to be a “late disclosure” of MST when the initial MST screen result was nonpositive.

### Any MST Disclosure

We combined the results of the MST screen with those Veterans having NLP evidence of MST to arrive at the group of Veterans with “any evidence of military sexual trauma.”

### Sociodemographic Predictors

Sociodemographic characteristics were determined from the DoD OEF/OIF/OND roster and VHA clinical records. Specifically, we examined sex, age at the time of MST screen in years (19–30, 31–40, 41–50, 51 y or older), marital status (married, previously married, never married), race/ethnicity (non-Hispanic white, non-Hispanic black, Hispanic, other), and educational status (some college vs. high school or less). Information on race and ethnicity from the DoD OEF/OIF/OND roster and VHA clinical records were combined into a single variable to reduce missingness in each data source.<sup>18</sup> Our preliminary analysis of Veterans from the 2004 to 2016 roster found this method results in 97% classification of race/ethnicity of service members who have established care in VHA (data not shown). We also examined military service characteristics including branch of service (Army, Air Force, Navy, Marines, Coast Guard), Active Duty versus National Guard and Reserve Component, and Officer or Warrant versus enlisted rank from the OEF/OIF/OND roster. All sociodemographic characteristics were treated as categorical (factor) variables in our analyses.

### Statistical Analyses

We examined differences in the distribution of sociodemographic and military service characteristics of Veterans with respect to “any evidence of MST” and “late disclosure” using  $\chi^2$  statistics. The  $\chi^2$  tests were conducted separately for male and female Veterans. Any sociodemographic or military service variable found to be associated with race/ethnicity and

the MST outcomes at the  $P < 0.10$  level in bivariate analyses was selected for inclusion in multivariable models. We then used logistic regression models to estimate the effect of racial/ethnic minority status on the odds of having any evidence of MST, and, among those with who were positive for MST, the odds of having a late disclosure as determined by NLP (with non-Hispanic whites as referent group). The models, stratified by sex, controlled for demographic and military service characteristics from the bivariate analyses. All  $P$ -values were 2-sided and defined to be significant at  $P < 0.05$ . Analyses were conducted using Stata version 15.

## RESULTS

### Evidence of MST

Of the national random sample of 10,000 male and 10,000 female Veterans from recent conflicts in Afghanistan and Iraq, 15,090 had an MST screen result during the study period and 13,334 had at least one electronic medical note from being seen as a patient in VHA in the 12 months following the MST screen (Fig. 1). The analytic sample included 6618 male and 6716 female Veterans, of whom 1473 had a positive MST screen result [68 male (1.0%); 1405 female (21%)].

The V3NLP pipeline was applied to a set of 362,559 documents from the Veterans in the analytic cohort. The pipeline identified 761 individuals with a positively asserted MST concept; 504 following a positive MST screen and 257 following a negative MST screen. The PPV of positively asserted MST concepts was 0.96 overall for those with a positive MST screen (0.99 for females, 0.98 for males). For those with MST negative screens, the PPV at the concept was 0.80 (0.91 for females, and 0.40 for males). The PPV for MST concepts at the *patient* level for those with positive MST screen was 0.95 (0.95 for females, 0.87 for males); for those with negative MST screens, the PPV at the patient level was 0.58 overall (0.70 for females, 0.36 for males).

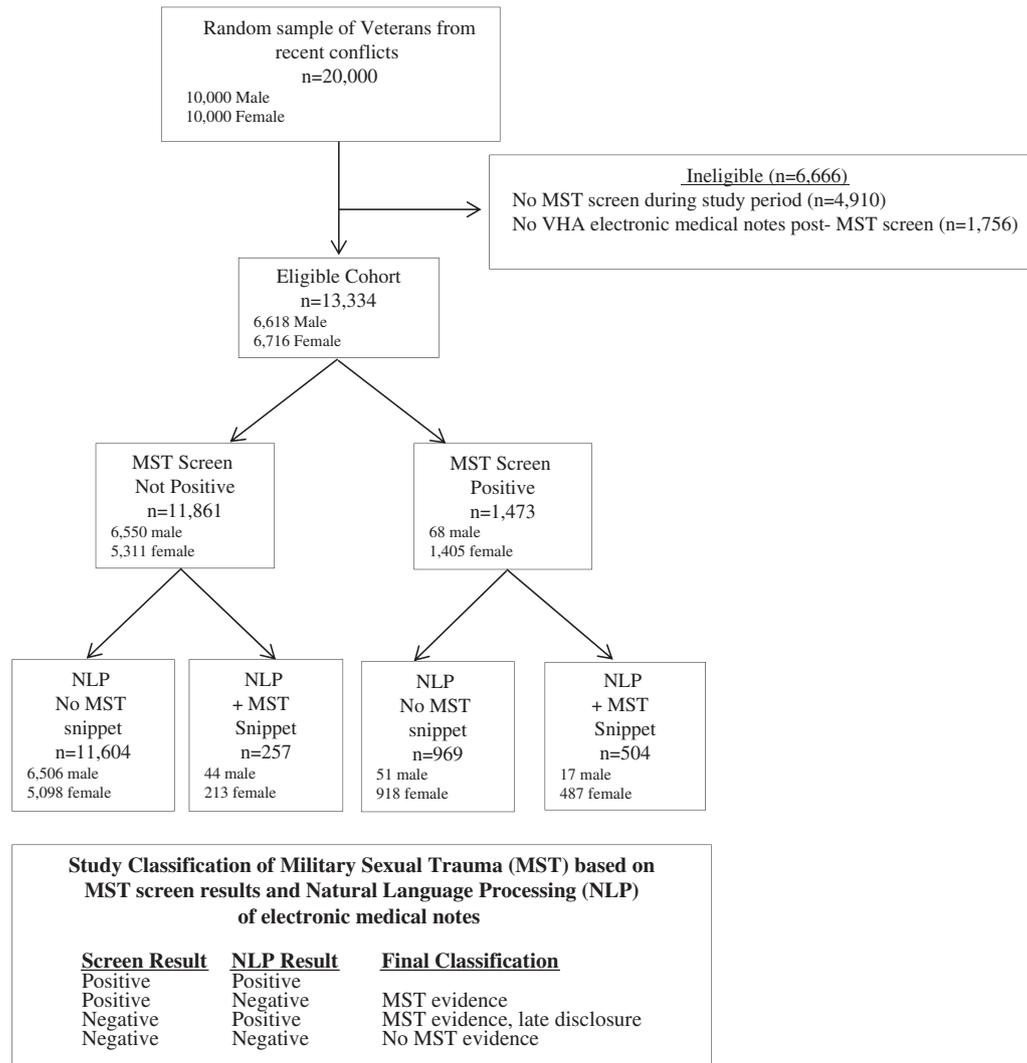
The 257 Veterans with evidence of MST by NLP only were designated as having “late disclosure of MST.” Combining the initial positive MST screen results with the NLP evidence yielded a group of 112 male and 1,618 female Veterans with “any evidence of MST.” Of these, 44 (39%) males and 213 (13%) females had a late disclosure of MST.

### Documentation of Late Disclosure

Late disclosure was most often documented in mental health-related visit notes for both male and female Veterans (Fig. 2). A manual review of notes revealed a range of concepts related to MST: direct references such as “reports being sexually assaulted by a member of her unit” to general references such as “being treated for PTSD as a result of MST” or “reports a history of MST during his last deployment.”

### Characteristics Associated With Any MST Evidence

As shown in Table 1, many of the sociodemographic variables were associated with MST evidence in females, but not males. Among males, the MST evidence group had a greater percentage of Veterans who served Active Duty compared with the no MST evidence group (64% vs. 54%).



**FIGURE 1.** CONSORT style flow diagram of study cohorts and classification of MST based on MST screen results and NLP of electronic medical notes. MST indicates military sexual trauma; NLP, natural language processing; VHA, Veterans Health Administration.

No other sociodemographic variables were associated with MST evidence in males (all *P*'s > 0.05).

Among women, the MST evidence group had a smaller percentage of non-Hispanic black Veterans compared with the no MST group (29% vs. 32%). The MST group had greater percentages of Veterans who were age 31–40 years, married, and who served Active duty, compared with the no MST evidence group; they had smaller percentages of Veterans with more than a high school education, and those serving in an Officer or Warrant rank.

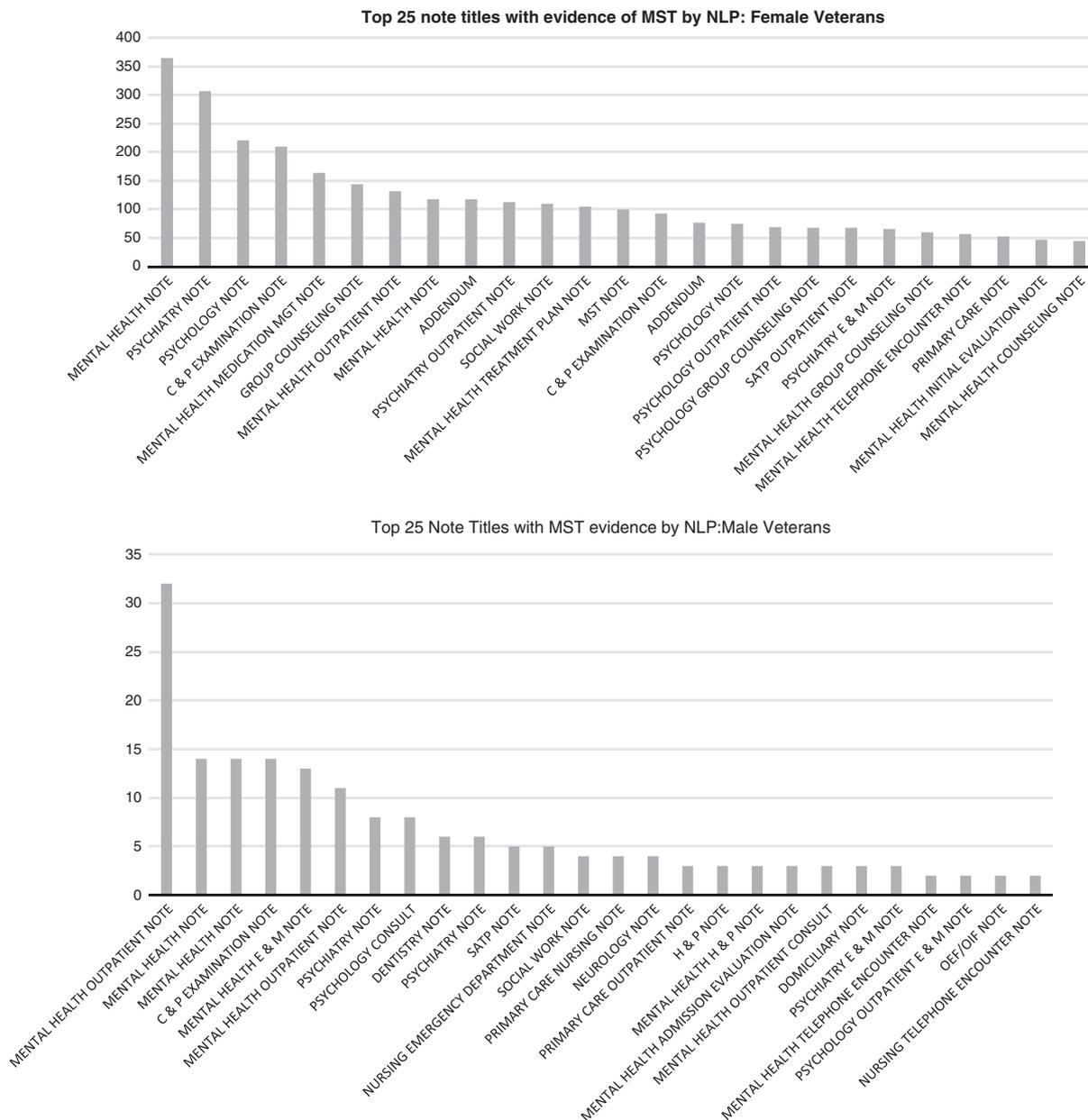
### Adjusted Differences in Any MST Evidence

The patterns of results were similar when controlling for potential confounders (Table 2). Men who served Active Duty were more likely than men who served in the Reserves to have any MST evidence [adjusted odds ratio (AOR) = 1.53]. Black women were less likely than white women to have any MST evidence (AOR = 0.75). Women over age 30 years, and those

who served Active Duty, were more likely than women ages 19–30 years and/or women serving in the National Guard or Reserve components to have any MST evidence (AORs = 1.40, 1.51, 1.66, and 1.53, respectively). More than a high school education and Officer or Warrant rank was associated with lower likelihood of having any MST evidence among women (AOR = 0.83 and 0.70, respectively).

### Characteristics Associated With Late Disclosure of MST

In the subsample of respondents with any evidence of MST, several sociodemographic variables were associated with delayed disclosure (Table 3). Among men, the late disclosure group had a smaller percentage of Veterans who served Active Duty (vs. National Guard or Reserve component) compared with the initial disclosure group (46% vs. 77%); no other sociodemographic variables were associated with delayed disclosure of MST (all *P*'s > 0.05).



**FIGURE 2.** Graph of top 25 note titles with evidence of MST by NLP for female and male Veterans. C&P indicates Compensation and Pension; E&M, evaluation & management; H&P, history and physical; MST, military sexual trauma; NLP, natural language processing; OEF/OIF, Operations Enduring Freedom/Iraqi Freedom; SATP, substance abuse treatment program.

Among women, the late MST disclosure group had a greater percentage of black and Hispanic Veterans than the initial MST disclosure group (39% and 13% vs. 27% and 11%). The late MST disclosure group had a greater percentage of Army Veterans than the initial MST screen positive group, but a smaller percentage of Active Duty Veterans.

Hispanic women were more likely than white women to evidence MST through delayed disclosure (AOR = 1.89 and 1.59, respectively). Women who served Active Duty were less likely than those in the National Guard or Reserves to disclose MST late (AOR = 0.66).

**Adjusted Differences in Late Disclosure of MST**

In multivariable models that accounted for potential confounding (Table 4), men who served Active Duty were less likely than men serving in the National Guard or Reserves to have a delayed MST disclosure (AOR = 0.29). Black and

**DISCUSSION**

Building upon the natural experiment of the formal screening of all Veterans seeking medical care in VHA for MST and offering care to sexual trauma survivors, we demonstrate that combining NLP results with traditional administrative data allowed us to demonstrate differences

**TABLE 1.** Demographic and Military Characteristics of Persons With “Any Evidence of Military Sexual Trauma” Among a National Random Sample of Male and Female US Military Veterans From Recent Conflicts in Afghanistan and Iraq

	Male [n (%)]		P	Female [n (%)]		P
	No MST Evidence (n = 6506)	Any MST Evidence (n = 112)		No MST Evidence (n = 5098)	Any MST Evidence (n = 1618)	
Race/ethnicity			0.90			0.015
Non-Hispanic white	4386 (69.2)	72 (67.3)		2564 (51.7)	881 (55.8)	
Non-Hispanic black	970 (15.3)	18 (16.8)		1577 (31.8)	450 (28.5)	
Hispanic	749 (11.8)	14 (13.1)		568 (11.5)	184 (11.7)	
Other	232 (3.7)	3 (2.8)		253 (5.1)	65 (4.1)	
Age (y)			0.16			<0.001
19–30	1772 (27.2)	23 (20.5)		1394 (27.3)	348 (21.5)	
31–40	2566 (39.4)	54 (48.2)		2171 (42.6)	767 (47.4)	
41–50	1303 (20.0)	18 (16.1)		931 (18.3)	310 (19.2)	
51 +	865 (13.3)	17 (15.2)		602 (11.8)	193 (11.9)	
More than high school	805 (12.6)	8 (8.2)	0.17	984 (19.5)	264 (16.5)	0.007
Marital status			0.49			0.038
Married	2837 (44.3)	56 (50.0)		1667 (33.4)	585 (36.4)	
Previously married	253 (3.9)	4 (3.6)		483 (9.7)	166 (10.3)	
Never married	3312 (51.7)	52 (46.4)		2841 (56.9)	857 (53.3)	
Branch of service			0.14			0.71
Army	4170 (64.1)	69 (61.6)		3253 (63.8)	1043 (64.5)	
Air force	587 (9.0)	7 (6.3)		789 (15.5)	255 (15.8)	
Navy	776 (11.9)	14 (12.5)		847 (16.6)	249 (15.4)	
Marines	965 (14.8)	21 (18.8)		206 (4.0)	69 (4.3)	
Coast guard	8 (0.1)	1 (0.9)		3 (0.1)	2 (0.1)	
Active Duty	3497 (53.8)	72 (64.3)	0.027	2664 (52.3)	999 (61.7)	<0.001
Officer or Warrant rank	386 (5.9)	4 (3.6)	0.28	414 (8.3)	99 (6.1)	0.005

Any evidence of MST = positive MST screen result or evidence of MST by natural language processing.  
 P-values obtained from  $\chi^2$  test of differences.  
 MST indicates military sexual trauma.

among racial/ethnic minority Veterans with respect to initial versus late disclosure of a history of MST. Specifically, we found that black women were less likely than white women to have any evidence of any MST, and if they disclosed, they were more likely than white women to disclose at a later date

as evidenced by documentation in the electronic medical note. Hispanic women were also more likely than white women to disclose MST late.

To our knowledge, this is the first report on racial/ethnic differences in MST disclosures by Veterans who establish and

**TABLE 2.** Adjusted Racial/Ethnic Differences in “Any Evidence of MST” Among a National Random Sample of US Military Veterans From Recent Conflicts in Afghanistan and Iraq

	Male		P	Female		P
	AOR (SE)	95% CI		AOR (SE)	95% CI	
Race/ethnicity						
Non-Hispanic white (ref.)	1.0	—	—	1.0	—	—
Non-Hispanic black	1.11 (0.30)	0.66–1.88	0.69	0.75 (0.05)	0.65–0.86	<0.001
Hispanic	1.13 (0.33)	0.63–2.01	0.68	0.92 (0.09)	0.76–1.10	0.36
Other	0.76 (0.45)	0.24–2.43	0.64	0.80 (0.12)	0.60–1.08	0.14
Age (y)						
19–30 (ref.)	—	—	—	1.0	—	—
31–40	—	—	—	1.40 (0.10)	1.21–1.62	<0.001
41–50	—	—	—	1.51 (0.15)	1.24–1.83	<0.001
51 +	—	—	—	1.66 (0.20)	1.32–2.09	<0.001
More than high school	—	—	—	0.83 (0.08)	0.69–0.99	0.04
Marital status						
Married (ref.)	—	—	—	1.0	—	—
Never married	—	—	—	0.92 (0.06)	0.81–1.05	0.20
Previously married	—	—	—	1.03 (0.11)	0.84–1.27	0.78
Active Duty	1.53 (0.31)	1.04–2.27	0.03	1.53 (0.09)	1.36–1.73	<0.001
Officer or Warrant rank	—	—	—	0.70 (0.09)	0.54–0.90	0.006

Any evidence of MST = positive MST screen result or evidence of MST by NLP. Results obtained from multivariable logistic regression models of any MST, run separately for male and female Veterans.  
 AOR indicates adjusted odds ratio; CI, confidence interval; MST, military sexual trauma; NLP, natural language processing; ref., reference.

**TABLE 3.** Demographic and Military Characteristics of Persons With Late MST Disclosure, Among a National Random Sample of Male and Female US Military Veterans With Any Evidence of MST

	Male [n (%)]		P	Female [n (%)]		P
	Initial MST Screen Positive (n = 68)	Late MST Disclosure (n = 44)		Initial MST Screen Positive (n = 1405)	Late MST Disclosure (n = 213)	
Race/ethnicity			0.52			0.001
Non-Hispanic white	41 (65.1)	31 (70.5)		789 (57.6)	92 (44.0)	
Non-Hispanic black	11 (17.5)	7 (15.9)		369 (26.9)	81 (38.8)	
Hispanic	8 (12.7)	6 (13.6)		156 (11.4)	28 (13.4)	
Other	3 (4.8)	0		57 (4.2)	8 (3.8)	
Age (y)			0.95			0.18
19–30	14 (20.6)	9 (20.5)		313 (22.3)	35 (16.4)	
31–40	32 (47.1)	22 (55.0)		659 (46.9)	108 (50.7)	
41–50	12 (17.7)	6 (13.6)		263 (18.7)	47 (22.1)	
51+	10 (14.7)	7 (15.9)		170 (12.1)	23 (10.8)	
More than high school	6 (9.1)	3 (6.8)	0.67	227 (16.3)	37 (17.7)	0.62
Marital status			0.81			0.53
Married	33 (48.5)	23 (52.3)		515 (36.8)	70 (33.5)	
Previously married	3 (4.4)	1 (2.3)		146 (10.4)	20 (9.6)	
Never married	32 (47.1)	20 (45.5)		738 (52.6)	119 (56.9)	
Branch of service			0.25			0.047
Army	38 (55.9)	31 (70.5)		887 (63.1)	156 (73.2)	
Air force	6 (8.8)	1 (2.3)		229 (16.3)	26 (12.2)	
Navy	11 (16.2)	3 (6.8)		222 (15.8)	27 (12.7)	
Marines	12 (17.7)	9 (20.5)		65 (4.6)	4 (1.9)	
Coast guard	1 (1.5)	0		2 (0.1)	0	
Active Duty	52 (76.5)	20 (45.5)	0.001	888 (63.2)	111 (52.1)	0.002
Officer or Warrant rank	2 (2.9)	2 (4.6)	0.66	86 (6.1)	13 (6.1)	0.97

Late disclosure = MST evidence obtained with NLP of electronic medical notes in the 12 months following an initial nonpositive MST screen result; Any evidence of MST = positive MST screen result or evidence of MST by NLP.

P-values obtained from  $\chi$  test of differences.

MST indicates military sexual trauma; NLP, natural language processing.

seek care in VHA. In addition, there have been no studies that have combined NLP extractions with administrative data to evaluate the prevalence of disparities in diagnoses or management of health conditions among Veterans from racial/ethnic minority groups. Our finding, that 39% of men and 13% of women with any evidence of MST disclosed late, support prior research highlighting the nondisclosure of MST by male Veterans.<sup>19,20</sup> Prevalence estimates of MST derived from the

national screen are likely low, especially for men. Our finding that black women with any MST evidence were more likely than white women to disclose late, is consistent with qualitative studies describing a reluctance to disclose sexual trauma among African American women in the community. Combined, our findings of late MST disclosure among men, as well as black and Hispanic women Veterans using NLP methods support the notion that sex, racial/ethnic, and culturally tailored

**TABLE 4.** Adjusted Racial/Ethnic Differences in “Late Disclosure of MST” Among a National Random Sample of Male and Female US Military Veterans With Any Evidence of MST

	Male		P	Female		P
	AOR (SE)	95% CI		AOR (SE)	95% CI	
Race/ethnicity						
Non-Hispanic white (ref.)	1.0	—	—	1.0	—	—
Non-Hispanic black	1.03 (0.59)	0.34–3.13	0.95	1.89 (0.32)	1.36–2.62	<0.001
Hispanic	1.34 (0.83)	0.40–4.53	0.64	1.59 (0.37)	1.01–2.53	0.047
Other	—	—	—	1.34 (0.53)	0.61–2.91	0.46
Branch of service						
Army or coast guard (ref.)	—	—	—	1.0	—	—
Air force	—	—	—	0.76 (0.17)	0.48–1.19	0.23
Navy	—	—	—	0.80 (0.18)	0.51–1.26	0.34
Marines	—	—	—	0.46 (0.24)	0.16–1.29	0.14
Active Duty	0.29 (0.13)	0.13–0.68	0.004	0.66 (0.10)	0.48–0.90	0.006

Late disclosure = MST evidence obtained with NLP of electronic medical notes in the 12 months following an initial nonpositive MST screen result. Results obtained from multivariable logistic regression models of delayed MST disclosure versus initial disclosure on an MST screen, stratified by sex. Because, the number of female Veterans serving in the Coast guard was too small to analyze uniquely (n=2), the Coast Guard was combined with the Army referent group (n=255) in multivariable analyses.

AOR indicates adjusted odds ratio; CI, confidence interval; MST, military sexual trauma; NLP, natural language processing; ref., reference.

interventions may be needed to facilitate the early disclosure and treatment of sexual trauma experiences in VHA.<sup>21</sup>

As hypothesized, the documentation of MST in the clinical narrative was most often seen in notes from mental health-related visits to psychiatrists, psychologists, counselors, and social workers. A manual review of the snippets revealed graphic references to the type and nature of sexual assault and the recurrence of the assaults. Some Veterans refer to the lack of support from those in authority and negative responses received by the victim after disclosure. These descriptions, though not quantitative, support the general notion that disclosure of sexual trauma is challenging for many victims. It is of interest to note that there were fewer references to MST in primary care and specialty medical care notes. A visit and electronic medical note that is unique to Veterans in VHA is the Compensation and Pension Examination (C&P) note which was a particularly high-yield note title for both females and males. This visit and note are to document exposures and traumatic experiences during deployment that have resulted in a compensable disability or condition; MST is one such condition.

The results of our study highlight several avenues for future research. First, it will be important to replicate these pilot findings of delayed disclosure with larger samples. It may be fruitful to review text notes over longer durations following the MST screen to account for longer delays in MST disclosure. Second, prospective studies are needed to better understand barriers to reporting MST in primary care setting for male and non-Hispanic black Veterans. It would be important to know, for example, whether the sex and racial/ethnic concordance of providers administering the screen are factors in patients' initial versus late disclosure of MST and documentation in the medical note. With a steady increase seen in racial/ethnic minority military service members who subsequently become Veterans and seek care in VHA, this is of importance at present and in future years. With regard to the NLP pipeline, the applications extend beyond VHA records as the principles and terminologies used to represent sexual trauma in other health care system records are likely to be similar. Thus, our work is generalizable to other health care system electronic medical records in identifying disclosure of sensitive topics, such as sexual trauma.

We acknowledge several limitations. While our random sample is representative of the larger set of OEF/OIF/OND Veterans with regard to the prevalence of MST in this group (~1% in males and 20% in females), our sample included small numbers of male Veterans with any MST evidence. This could have limited our ability to observe effects that would have been more apparent with a larger sample size. The PPV for evidence of MST by NLP on electronic medical notes at the concept and patient level for male Veterans was low for those with negative MST screens. This may have affected our ability to detect late disclosure of MST among men in particular. Finally, in both men and women, the small numbers of Veterans with any MST evidence precluded us from examining MST disclosure patterns for Asian, Pacific Island, and American Indian/Alaska Native Veterans, or assessing regional variations in documentation of sexual trauma in VHA. These topics merit further study.

In conclusion, this pilot study demonstrates the feasibility of analyzing unstructured free text data in phenotyping Veterans with regard to their disclosure of MST. Researchers engaged in health disparities research may consider adding NLP of electronic medical notes to their toolkit.

## REFERENCES

1. Tillman S, Bryant-Davis T, Smith K, et al. Shattering silence: exploring barriers to disclosure for African American sexual assault survivors. *Trauma Violence Abuse*. 2010;11:59–70.
2. Bryant-Davis T, Chung H, Tillman S, et al. From the margins to the center: ethnic minority women and the mental health effects of sexual assault. *Trauma Violence Abuse*. 2009;10:330–357.
3. Turchik JA, Wilson SM. Sexual assault in the US military. A review of the literature and recommendations for the future. *Aggress Violent Behav*. 2010;15:267–277.
4. Turchik JA, Edwards KM. Myths about male rape: a literature review. *Psychol Men Masc*. 2012;13:211–226.
5. Kimerling R, Gima K, Smith MW, et al. The Veterans Health Administration and military sexual trauma. *Am J Public Health*. 2007;97:2160–2166.
6. Blais RK, Brignone E, Fargo JD, et al. Assailant identity and self-reported nondisclosure of military sexual trauma in partnered women Veterans. *Psychol Trauma*. 2018;10:470–474.
7. Gundlapalli AV, Brignone E, Divita G, et al. Using structured and unstructured data to refine estimates of military sexual trauma status among US Military Veterans. *Stud Health Technol Inform*. 2017;238:128–131.
8. Divita G, Brignone E, Carter ME, et al. Extracting sexual trauma mentions from electronic medical notes using natural language processing. *Stud Health Technol Inform*. 2017;245:351–355.
9. Divita G, Carter M, Redd A, et al. Scaling-up NLP pipelines to process large corpora of clinical notes. *Methods Inf Med*. 2015;54:548–552.
10. Friedman C, Rindflesch TC, Corn M. Natural language processing: state of the art and prospects for significant progress, a workshop sponsored by the National Library of Medicine. *J Biomed Inform*. 2013;46:765–773.
11. Liao KP, Cai T, Savova GK, et al. Development of phenotype algorithms using electronic medical records and incorporating natural language processing. *BMJ*. 2015;350:h1885.
12. Office of the Surgeon General, Center for Mental Health Services, National Institute of Mental Health. *Mental Health: Culture, Race, and Ethnicity: A Supplement to Mental Health: A Report of the Surgeon General*. Rockville, MD: Substance Abuse and Mental Health Services Administration; 2001. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK44243/>.
13. Miranda J, McGuire TG, Williams DR, et al. Mental health in the context of health disparities. *Am J Psychiatry*. 2008;165:1102–1108.
14. US Department of Veterans Affairs. *VA Informatics and Computing Infrastructure (VINCI)*. Washington DC: US Department of Veterans Affairs; 2016. Available at: [www.hsrd.research.va.gov/for\\_researchers/vinci/](http://www.hsrd.research.va.gov/for_researchers/vinci/). Accessed September 10, 2018.
15. Kimerling R, Street AE, Pavao J, et al. Military-related sexual trauma among Veterans Health Administration patients returning from Afghanistan and Iraq. *Am J Public Health*. 2010;100:1409–1412.
16. Brignone E, Gundlapalli AV, Blais RK, et al. Differential risk for homelessness among US male and female veterans with a positive screen for military sexual trauma. *JAMA Psychiatry*. 2016;73:582–589.
17. The Apache Software Foundation. Apache UIMA: the Apache Software Foundation. 2018. Available at: <https://uima.apache.org/>. Accessed September 12, 2018.
18. Koo KH, Hebenstreit CL, Madden E, et al. Race/ethnicity and gender differences in mental health diagnoses among Iraq and Afghanistan veterans. *Psychiatry Res*. 2015;229:724–731.
19. Eckerlin DM, Kovalesky A, Jakupcak M, CE. military sexual trauma in male service members. *Am J Nurs*. 2016;116:34–43.
20. Sheppard SC, Hickling EJ, Earleywine M, et al. Preliminary data suggest rates of male military sexual trauma may be higher than previously reported. *Psychological services*. 2015;12:344–347.
21. Roberts ST, Watlington CG, Nett SD, et al. Sexual trauma disclosure in clinical settings: addressing diversity. *J Trauma Dissociation*. 2010;11:244–259.