Discourse, Health and Well-being of Military Populations Through the Social Media Lens

Umashanthi Pavalanathan[1], Vivek Datla[2], Svitlana Volkova[2], Lauren Charles-Smith[2], Meg Pirrung[2], Josh Harrison[2], Alan Chappell[2], Courtney D Corley[2]
[1]Georgia Institute of Technology, Atlanta, GA, USA
[2]Pacific Northwest National Laboratory, Richland, WA, USA
umashanthi@gatech.edu, {vivek.datla, svitlana.volkova, lauren.charles-smith}@pnnl.gov,
{meg.pirrung, Josh.Harrison, Alan.Chappell, court}@pnnl.gov

Abstract

Social media can provide a resource for characterizing communities and small populations through activities and content shared online. For instance, studying the language use in social media within military populations may provide insights into their health and well-being. In this paper, we address three research questions: (1) How do military populations use social media? (2) What do military users discuss in social media? And (3) Do military users talk about health and well-being differently than civilians? Military Twitter users were identified through keywords in the profile description of users who posted geo-tagged tweets at military installations. The data was anonymized for the analysis. User tweets that belong to military populations were compared to non-military populations. Our results indicate that military users talk more about events in their military life, whereas non-military users talk more about school, work, and leisure activities. Additionally, we identified significant differences in communication behavior between two populations, including health-related language.

Introduction

Social media has become a resource for studying different social, emotional, health and economic conditions of the communities through their online activities and shared content. Recently, there have been studies that seek to understand the emotions and behavior in different groups of people through their social media footprints (Ammari and Schoenebeck 2015; De Choudhury, Counts, and Horvitz 2013). Other studies aim to investigate social issues and phenomena existing in communities through their online activities (Delgado Valdes, Eisenstein, and De Choudhury 2015; Lin 2014).

Social media, such as Twitter, provides publicly available information that provides a resource for potential identification of subcommunities (Ammari and Schoenebeck 2015; De Choudhury, Counts, and Horvitz 2013; Lin 2014). Applying the existing techniques to military populations creates a potential to identify and characterize their health status by analyzing content shared online. For instance, recent studies used signals from social media to study subpopulations online with the goal of detecting food poisoning within certain subpopulations and geographic regions (Harris et al. 2014), identifying subpopulations of smokers and drug addicts (Paul and Dredze 2013).

Military service type (e.g., Army, Navy, Marine, Air Force, active duty, reserves, veterans) may play a role in the health and well-being of military personnel, including the development of specific health conditions. Boehmer et al. (2003) studied the association between military service and health-related quality of life, using a population-based sample of adults in the U.S. They found that the active duty population had more health complaints than either reserve or veteran populations.

In this work, we aim to understand the differences in online behavior and content produced by military populations, which share common characteristics, such as location, work, and culture, and compare them with the surrounding populations. Specifically, we qualitatively and quantitatively estimate the differences in language and communication patterns, and the usage of health-related terms across contrastive populations.

Understanding social media activities and discourse of military personnel and their families may help decision makers gain real-time insights into the mental health, including social and emotional stressors, and other health-related issues of the military population through a minimally invasive and economic approach. The military could use the proposed methods to identify targeted populations quickly and distribute resources effectively.

Below we provide a background about the U.S. military population, describe our data and methods for identifying military users on Twitter. Then, we present our analysis and discuss implications of our findings.

Research Questions

Our motivation to study social media activities and discourse of the military population is to understand better their social interactions, and how it can help decision makers to identify issues specific to military populations. Additionally, social media discourse can also help identify emergent health issues prevalent in the military. Based on these motivations, we are interested in answering the broad questions by addressing the following finer research questions.

- How does the military population use social media?

  - RQ 1: What are the differences in tweeting behavior between military and non-military (control) populations?
- What do military users discuss in social media?
  - RQ 2: What are the linguistic differences between the content produced by the military vs. civilian populations?
- Do military users talk about health differently than others?
  - RQ 3: Are there any differences in the discourse of health-related topics by the military population compared to the control?

Background and Related Work

In this section, we first provide a background and summary of prior research about the U.S. military. Then, we briefly discuss prior work on understanding different populations through social media.

Characteristics of the U.S. Military Population

The U.S. military consists of active-duty forces (Army, Air Force, Marine Corps, and Navy) and supporting groups (National Guard, Military Reserves, and Coast Guard). Within the U.S., armed forces density varies by state, and Texas, California, North Carolina, and Virginia have the highest concentrations (Segal and Segal 2004).

In active duty and the reserves, people sign up for a specific length of duty and leave service or retire after that term. Three-quarters of the military population are less than 40 years old, and half of the active duty enlisted personnel are less than 25 years old. More than half of the military personnel are married, and 73% of married personnel have children (Office of the Deputy Assistant Secretary of Defense 2013). The military population is diverse, with non-native speakers of English, nonwhites, and women. The military is vulnerable to physical and mental health problems with nearly 18% of active duty deaths caused by illness, and more than 10% of deaths are caused by suicides (Segal and Segal 2004).

Studies on Military Populations

Since the U.S. armed forces changed from drafting to enlistment in 1973, sociologists have debated whether to study the military as an institution or an occupation (Siebold 2001). In general, the military is becoming oriented as a profession yet retains institutional features (Siebold 2001). The U.S. military population reflects America's racial, ethnic, religious, and socioeconomic diversity (Segal and Segal 2004), however, their military status unifies them as a unique subpopulation.

Understanding Populations through Social Media

Recent studies have attempted to identify and understand different subpopulations through their social media activities. Geotagged social media data can be used to identify specific geographical communities and urban areas (Delgado Valdes, Eisenstein, and De Choudhury 2015; Lin 2014). Other work looking at certain demographic groups use social media profile information and forums to recruit or sample subjects for their studies. Such groups include new mothers (De Choudhury, Counts, and Horvitz 2013), fathers (Ammari and Schoenebeck 2015), and mothers using anonymous social media platforms (Schoenebeck 2013).

In line with recent research, we seek to study the U.S. military population through the lens of their online social media activities.

Data

Identifying subpopulation in social media with certain common characteristics (e.g., profession or location) is a challenging task. For our study, the data collection problem entailed differentiating public social media data from the military population and the surrounding civilian population.

Our initial dataset includes nearly 200 million geo-tagged tweets from November 2011 to June 2015 that originated within a 25-mile radius of 31 U.S. military base locations globally. We used this historical dataset to build a keyword lexicon to identify and sample users who are likely to belong to the military population.

For our analysis, we choose six different U.S. military installations located in California, North Carolina, and Texas (Table 1). We chose locations which have a high population of military to surrounding population. For each of these states, we chose one control location that was at least 50 miles from any military facility, and assumed that at this distance users were less likely to belong to the military population. From tweets that originated within a 25-mile radius of military facilities, we sampled users who were likely to belong to the military using the methodology explained in the next subsection. We sampled the same number of users from non-military locations for our control dataset. We collected up to 3200 of the most recent tweets per user in the military and control samples. Note that this timeline dataset contains anonymized tweets with and without geographic coordinates.

Data Anonymization

We followed a rigorous data anonymizing procedure to ensure privacy of all Twitter users. The data collected from querying the public Twitter API was anonymized specifically for usernames, userids and tweetids. This data was fed into an elasticsearch engine where it was encrypted using the state-of-art encryption algorithms. Our analysis is based only on completely anonymized data and findings are reported on an abstract, aggregate level. Below is the detailed description of our sampling and data collection procedures.

L_1	L_2	L_3	L_4	L_5	L_6
$4,246$	$1,040$	$1,538$	$1,372$	$1,720$	926

Table 1: Military locations $L_1 \ldots L_6$ and the corresponding number of users sampled for both military and control populations together. The total number of users sampled across six locations is $10,814$.

Sampling Military Users on Twitter

While studying social media activities and content shared by the military population, our first challenge was in sampling Twitter users who are likely to belong to the military. The standard practice in identifying specific events or users in social media is to search for specific terms or hashtags in the tweets (Cui et al. 2012; Starbird et al. 2014). This approach was not appropriate for our experiments because we were interested in analyzing the content itself; extracting tweets with such keywords would bias our content analysis.

Another approach often used to identify specific users on social media is to use a database or web listings of users belonging to specific groups, e.g., in (Soni et al. 2014), online sources were used to obtain Twitter handles of journalists. However, to the best of our knowledge, there are no such listings available for military users specifically. Extracting tweet handles for some military organizations from their websites (e.g., @USArmyReserve, @camp Lejeune, @Military1Source) provided a way to identify military users. We devised an approach for discovering potential military user Twitter accounts based on publicly provided content in the profile description.

To gather tweets that have a high likelihood of being posted by someone in the military, we extracted tweets that originated within a 0.5-mile radius from four different base locations. These locations were selected based on the highest percentage of military-to-surrounding population ratio obtained from publicly available data.[1] The rationale for choosing a 0.5-mile radius was two-fold; it restricted the area and increased the probability of obtaining tweets from military users, and the resulting number of users per area is nearly 1000, which is a manageable size for faster annotation. The resulting user timelines were then anonymized according to the description above. Expert annotators classified profile descriptions of these anonymized tweets using most keywords from Table 2.

To sample Twitter users who are likely the military population, we extracted tweets from a 25-mile radius of the facilities in chosen military locations, and filtered tweets having most of the keywords from our lexicon in their profile description. Because we used both the geo-location and the appearance of keywords in the profile description to sample users, we expect our approach to perform better than the geo-location based approach used in prior work (Coppersmith, Harman, and Dredze 2014). For the control sample, we identified users from the three different control locations who had never used any of the keywords in their profile description. However, this control sample might include military users if they do not explicitly state their membership in their profile description.

Analysis and Results

RQ1: Differences in Social Media Activities of Military vs. Control

To identify the differences in tweeting behavior between the military personnel and control, we extracted counts for the size of their online social networks (i.e., the number of followers and friends), their interaction with other users (using user mentions as a proxy), their interaction with large groups of virtual communities (using hashtags as a proxy), and to understand their practice of location sharing in social media. We present and contrast the mean counts that represent user activity and online behavior across populations in Table 3. We observe high degree of variability among the military and control populations.

[1] http://www.militaryinstallations.dod.mil/

Group	Keywords
Active Duty	military, national guard, usmc, corporal, sergeant major, hospitalman, sailor, usaf
Family	army wife, usnspouse, military girl, navygirlfriend, army brat, airforce wife
Veteran	veteran, usnveteran, retired army, ex navy

Table 2: Example keywords used to identify military users

Twitter Usage and Frequency We found that tweeting frequency is higher for the control population compared to the military. The differences in status counts and the number of followers per user are not statistically significant for military vs. control populations. Military users write a higher proportion of tweets with geotags. Moreover, it has been reported recently that military personnel are allowed to use smartphones, (Powers 2014; Dave Lee North America Technology 2013), which have the geo-tagging capability.

Size of Social Network and Online Interactions The mean number of favorite counts is higher for the control population and the mean number of friend counts is higher for the military population.

The mean ratio of tweets with mentions and retweets show that military users interact less with others on social media using @-mentions and RT compared to the control, even though they have similar size social networks. However, military users use more hashtags and URLs on average

Counts	μ_{mil}	μ_{con}	p-value
Favorite	1604.1	2113.8	***
Friend	663.7	498.2	**
Follower	955.9	976.0	
Status	8455.2	8268.4	
Tweet freq.	5.434	6.656	***
Geotag	0.155	0.138	***
Hashtag	0.216	0.186	***
Media	0.097	0.112	***
Mention	0.478	0.497	***
Retweet	0.200	0.239	***
Url	0.237	0.183	***

Table 3: Comparing mean values for user activities and online behavior across military vs. control populations (p-value$\leq 0.001^{***}$, p-value$\leq 0.01^{**}$).

	L_1			L_2			L_3			L_4			L_5			L_6		
LIWC Category	Δ	t-stat	p	Δ	t-stat	p	Δ	t-stat	p	Δ	t-stat	p	Δ	t-stat	p	Δ	t-stat	p
Linguistic																		
Article	5.4	16.0	***	1.3	1.9		2.2	4.2	**	1.7	3.2		1.6	3.2		5.4	7.4	***
Prepositions	10.0	16.1	***	1.5	1.3		3.3	3.7	*	4.1	4.2	**	3.3	3.7	*	9.4	6.8	***
3rd person plu.	0.3	4.04	**	0.5	3.0		0.1	0.7		0.8	5.8	***	0.3	2.62		0.2	1.74	
Psychological																		
Personal																		
Work	1.2	10.9	***	0.3	1.8		0.2	1.7		0.6	2.4		0.5	3.0		1.3	5.8	***
School	-0.3	-2.8		-2.1	-8.7	***	-1.7	-7.4	***	-1.9	-9	***	-1.4	-6.3	***	-0.4	-1.5	
Money	0.8	6.2	***	0.5	1.9		0.1	0.2		0.5	2.2		0.5	2.7		1.4	5.1	***
Home	0	0.3		0.2	1.2		0.5	3.1		0.2	1.4		0	0.2		0	0.1	
Death	0.1	3.7	*	0	0.6		0.2	3.4		0.3	4.7	***	0.1	1.4		0.1	1.2	
Religion	-0.2	-1.2		-1.1	-2.7		-0.2	-0.9		-0.1	-0.6		-0.1	-0.5		-0.3	-0.9	
Relativity																		
Motion	1.1	5.7	***	0.5	1.1		0.4	1.7		1.1	3.7	*	0.6	2.3		0.1	0.4	
Relative	9.2	11.5	***	1.2	0.8		4.7	4.2	**	3.8	3.0		2.3	2.0		6.6	3.8	*
Space	6.1	14.7	***	1.5	1.8		1.9	3.0		3.0	4.4	**	1.7	2.6		6.5	7	***
Cognitive																		
Inhibition	0.7	10.2	***	0.5	3.7	*	0.5	4.7	***	0.7	6.8	***	0.4	4.4	***	0.7	4.7	***
Causation	0.4	2.8		0.6	2.7		0.7	3.7	*	0.8	3.8	*	0.2	0.9		0.6	2.1	
Perceptual																		
Perception	-0.3	-1.5		-0.5	-1.4		-0.4	-1.7		-0.2	-0.5		-0.6	-2.04		-0.3	-0.8	
Spoken																		
Nonfluencies	0.1	3.2		0.1	1.5		0.2	3.5	*	0.1	2.4		0.1	2.2		0	0.5	

Table 4: Differences Δs in linguistic attributes between military and control populations measured using LIWC. We only present linguistic categories which have the same same directions across populations $\Delta = (\mu_{mil} - \mu_{com}) \times 10^{-3}$ (p-value $\leq 0.001^{***}$, p-value $\leq 0.01^{**}$, p-value $\leq 0.05^{*}$).

but less media content compared to the control population.

RQ2: Differences in Language Use

Psychology literature suggests that language is a reliable way of measuring people's internal thoughts and emotions (Tausczik and Pennebaker 2010). Hence, we are interested in understanding military populations through the language used in their tweets. To identify the differences in the linguistic attributes between the military and control users, we first use a dictionary-based approach applying the Linguistic Inquiry and Word Count lexicon (LIWC) (Pennebaker, Francis, and Booth 2001).

Social media language, specifically in microblogs, is often found to be non-standard (Eisenstein 2013). Although there are a few recent works on normalizing techniques to convert tweets to more standard language (Han and Baldwin 2011; Yang and Eisenstein 2013), their performance has shown an only marginal increase in accuracy. Since these methods are in a very primary stage of development, we did not perform any normalization. Instead, we used an open vocabulary approach for extracting terms that differentiate the language of the military and control populations in complementary to dictionary based (LIWC) analysis.

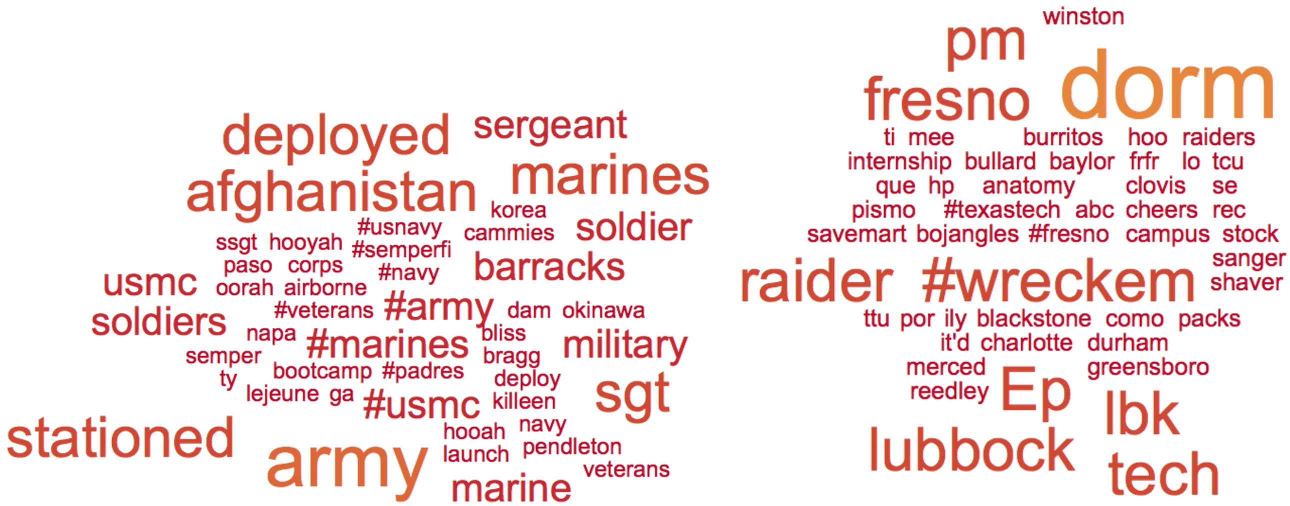
Differences in Linguistic Attributes We used the psycholinguistic lexicon[2] to measure the differences in linguistic attributes. LIWC consists of several categories of linguistic attributes, such as linguistic or psychological processes, personal concerns, and spoken categories.

To measure the differences in LIWC linguistic categories, we aggregate all tweets per user, then count the number of LIWC terms in each category, and normalize these counts by the total number of tokens per user. We compare the differences in LIWC terms for the military and control populations using an independent sample t-test. We report the results, showing only the measures that exhibit the same direction in the t-test for all military locations in Table 4.

Linguistic Processes Our results show that the military population uses more *articles* (e.g., a, an, the) and *prepositions* (e.g., to, with, above) compared to the control. Military users talk more about others by using more *third person plural words* (e.g., they, their) in comparison with the control.

Psychological Processes Military populations talk more about work and financial issues compared to the control populations in social media (estimated over the sample of the data we collected), as indicated by higher mentions of *work* (e.g., job, majors, labor) and *money* (e.g., bank, income, loan) related terms. In all of the six locations military users use more *home* (e.g., family, leasing, housing) related words compared to the control, although none of the differences are statistically significant.

[2]LIWC: `http://www.liwc.net`



(a) Military

(b) Control

Figure 1: Example keywords used to identify military vs. control users aggregated over all six locations, extracted using SAGE (Eisenstein, Ahmed, and Xing 2011).

Military personnel in certain locations use the significantly higher number of *death* (e.g., buried, died, kill) related terms. Compared to control users, military users talk significantly less about *school* related terms; they also talk less about *religion* (e.g., church, mosque, prayer) although the differences are not statistically significant. Military users in all of the six locations use *inhibition* (e.g., block, constrain, stop) related words in a significantly higher rate than respective control users.

Keyword Extraction To find keywords that are specific for military and control populations, we extracted terms that differentiate language between these populations. We used a regularized log-odds ratio based method (Eisenstein, Ahmed, and Xing 2011), which compares the base word distribution of each group and outputs terms that are specific for each group. We show the top terms for the military and control samples as word clouds in Figure 1.

Looking at the top population-specific terms, we find that terms relevant to the events in military life (e.g., Semper [motto of U.S. marine corps], barracks [buildings in military & facilities], boot camp, deployed, stationed, Sergeant) are more prevalent in the social media content of the military population. On the other hand, terms related to school, work and leisure (e.g., ep [episode], tix [tickets], dorm, campus, Raiders [sports], Savemart, Blackstone [stores or businesses], Greensboro, Winston, Sanger [place names]) are more prevalent in the control population's social media content.

We also find that military slang words are widespread in the social media content of military users (e.g., oorah [battle cry of marine corps], hooyah [battle cry of the navy], chow [food]); where as the control users have widespread usage of Internet slang words (e.g., ep [episode], tix [tickets], ik [I know], tbh [to be honest]) and entity names (e.g., Greensboro, Fresno [place names], Bojangles [food chain], ttu [university], Raiders [sports]). These results show that social media language of the military population is different than the control population.

RQ3: Health Discourse

Military populations are considered to be more vulnerable to infectious diseases, such as influenza, and mental health issues, because of overcrowding and a high degree of physical and mental stress (Gray et al. 1999; Pflanz 2001; Russell et al. 2006). To identify differences in the way military members and their families talk about health conditions compared to the general population, we created a lexicon of health terms and possible misspellings (e.g., "influenza" or "influlenza" for the correct spelling "influenza") and grouped them into six categories as shown in Table 5.

We calculated the total counts of terms appearing in user tweets and compared the average term count per token using a t-test (Table 6). After Bonferroni correction (Dunn 1961), we find that the direction and significance level of these health measures differ across health-related categories for the military and the civilian populations.

Overall the mean frequency of health-related tweeted complaints from the civilian population is slightly higher than the military population across all health-related categories. These results are obtained through comparisons of military and civilian populations across different geolocations and military service types. In Table 6, we show that civilians use disease-related words more frequently than the military.

Category	Example Keywords and stems
Self-related health experience	suffer*, struggl*, fatigue, weak
ILI-specific symptoms	fever*, cough*, shiver*, runny nose
Disease names and related terms	influenza, sick*, flu, asthma
Health entities	hospital*, doctor*, ER, clinic*
Parts of body & related	lung*, throat, stomach, platelet
Non-ILI-specific symptoms	breath*, diarrhea, dehydrat*, sneez*

Table 5: The example of health category keywords (ILI: Influenza like illnesses). * indicates a regular expression, for example fever* indicates words that have a prefix fever such as fevers, feverish and fevered.

Health Category	$\mu_{mil} \times 10^{-3}$	$\mu_{con} \times 10^{-3}$	t-stat	p-value
Self-related health experience	3.04	3.35	-5.246	9.49×10^{-7}
ILI-specific symptoms	71.5	79.0	-3.701	1.29×10^{-3}
Disease names and related terms	1.06	1.17	-4.781	1.06×10^{-5}
Health entities	1.02	1.07	-1.576	6.90×10^{-1}
Parts of body & related	28.5	33.6	-5.134	1.73×10^{-6}
Non-ILI-specific symptoms	49.5	54.5	-3.623	1.75×10^{-3}

Table 6: Comparing the counts of health words for military vs. control populations.

Discussion

In this paper, we analyzed social media data collected from military sites and corresponding control populations of users surrounding military locations. We explored the language and metadata inside the tweets from both populations in the following dimensions: behavior, language, and the discourse related to health topics.

Through the analysis of tweeting activities, we found that military populations use fewer retweets and @-mentions compared to the control group. As the usage of retweets and @-mentions are usually considered a measure of social interaction on Twitter (Macskassy 2012), similar to comments and likes on Facebook, these findings indicate that the military users are less interactive on Twitter compared to others.

We found differences in linguistic patterns of military users compared to the control: tweets from military users have a higher usage rate of articles, propositions, third person plural pronouns, and inhibition words; military users talk more about work and death, and less about school related terms in social media. The increased use of articles suggests that military users use more concrete nouns, and they are interested in objects and things (Tausczik and Pennebaker 2010) compared to the control, while the increased use of propositions suggest that military population is concerned with precision (Tausczik and Pennebaker 2010). Inhibition words are used to suppress strong emotions (Rand, Kraft-Todd, and Gruber 2015). Therefore, increased usage of inhibition words by military users may suggest that they suppress the expression of strong emotional content in social media compared to the control population. Military specific terms and slang words are prevalent in the tweets of the military users while the control users talk more about school and leisure activities.

From our analysis, we observed significant differences in online behavior and discourse of the U.S. military when compared to civilian users in social media. Below, we discuss the implications of our findings on life and health of military populations.

Implications for Military Social Life

Our study offers novel and interesting findings on social media activities and the discourse of the U.S. military population. This is an early work towards understanding the role of social networks for improving lives of military populations. From our findings from RQ1, military populations have lower social interactions on Twitter when compared to the control users. This suggests that military users are socially less active than others in social media. Findings from RQ2 show a significantly higher usage rate of inhibition words, which suggests the self-consciousness expressed in the military populations' messages.

Implications for Military and Public Health

Our findings for RQ3 show significant differences in the usage of medically-related terms between military vs. civilian populations on Twitter. Overall, civilian populations tend to use more health-related terms (diseases, symptoms, etc.) than military populations. Although, we cannot conclude that military populations are healthier compared to the civilian populations, as further study is needed to explore the usage of colloquial language or military jargon instead of the standard disease terminology. However, the direction of the variables that indicate health-related terms shows that there are significant differences in the way military and civilian populations talk about health in social media.

The discovered health-related expressions of military personnel on Twitter suggest that it is possible to utilize social media content from military users to identify emerging health issues that are prevalent in the military population due to the nature of their job and living conditions. Faster and better identification of health-related issues have implications on public health.

Limitations and Future Work

First, we relied on social media content from Twitter alone to study our research questions. Using only one social media source is a limitation and future work can expand this to other sources such as Facebook and Reddit. Second, we relied on the geotagged tweets for the initial identification of military users. However, recent work shows biases in the geotagged Twitter data regarding text content (Pavalanathan and Eisenstein 2015), and suggests considerations of these biases when generalizing research findings. Third, we relied on geo-origins of the tweets, and keywords in Twitter profile descriptions to extract users belonging to the military, but better identification methods can be explored. Fourth, we did not take into account demographics for military and control populations.

There are several directions for future work. Complementing this analysis with an interview study about social media usage of the military users will help researchers and decision makers to understand the limitations in using social media among military personnel.

Moreover, linguistic differences between military and civilian users would enable the construction of classification models to automatically identify military users in social media. Expanding the analysis on health discourse and deriving cues about military health issues to predict disease outbreaks is another possible direction.

Finally, understanding the discourse of military when compared to the civilians helps to identify and prevent social issues affecting them un-evasively. For instance, differences in discourse between military and non-military populations have been effectively used in other studies to identify emotional stress, depression, and PTSD related illness. In future, we would like to study fine-grained emotional differences between military and non-military populations over time, and model language variations among populations more effectively.

Conclusion

In this paper, we studied language and online behavior of military populations in comparison to civilians within the same geographic region through social media. We observed significant differences in tweeting behavior between these populations. We further analyzed language inside the tweets and observed that there are significant linguistic differences in emotion and psychological words used between military and civilian populations. Finally, we found that linguistic differences in health discourse between two populations.

References

Ammari, T., and Schoenebeck, S. 2015. Understanding and Supporting Fathers and Fatherhood on Social Media Sites. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 1905–1914. ACM.

Coppersmith, G.; Harman, C.; and Dredze, M. 2014. Measuring Post Traumatic Stress Disorder in Twitter. In *Proceedings of ICWSM*.

Cui, A.; Zhang, M.; Liu, Y.; Ma, S.; and Zhang, K. 2012. Discover Breaking Events with Popular Hashtags in Twitter. In *Proceedings of CIKM*, 1794–1798. New York, NY, USA: ACM.

Dave Lee North America Technology. 2013. US military approves Android phones for soldiers.

De Choudhury, M.; Counts, S.; and Horvitz, E. 2013. Predicting postpartum changes in emotion and behavior via social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3267–3276. ACM.

Delgado Valdes, J. M.; Eisenstein, J.; and De Choudhury, M. 2015. Psychological Effects of Urban Crime Gleaned from Social Media. In *Proceedings of ICWSM*.

Dunn, O. J. 1961. Multiple comparisons among means. *Journal of the American Statistical Association* 56(293):52–64.

Eisenstein, J.; Ahmed, A.; and Xing, E. P. 2011. Sparse Additive Generative Models of Text. In *Proceedings of ICML*, 1041–1048.

Eisenstein, J. 2013. What to do about bad language on the internet. In *Proceedings of NAACL*, 359–369. Stroudsburg, Pennsylvania: ACL.

Gray, G. C.; Callahan, J. D.; Hawksworth, A. W.; Fisher, C. A.; and Gaydos, J. C. 1999. Respiratory diseases among US military personnel: countering emerging threats. *Emerging Infectious Diseases* 5(3):379.

Han, B., and Baldwin, T. 2011. Lexical normalisation of short text messages: Makn sens a# twitter. In *Proceedings of HLT-Volume 1*, 368–378. ACL.

Harris, J.; Mansour, R.; Choucair, B.; Olson, J.; Nissen, C.; Bhatt, J.; and Brown, S. 2014. Health department use of social media to identify foodborne illness-chicago, illinois, 2013-2014. *MMWR. Morbidity and mortality weekly report* 63(32):681.

Lin, Y.-R. 2014. Assessing Sentiment Segregation in Urban Communities. In *Proceedings of the 2014 International Conference on Social Computing*, SocialCom '14, 9:1–9:8. New York, NY, USA: ACM.

Macskassy, S. A. 2012. On the study of social interactions in twitter. In *Proceedings of ICWSM*.

Office of the Deputy Assistant Secretary of Defense. 2013. Demographics report.

Paul, M. J., and Dredze, M. 2013. Drug extraction from the web: Summarizing drug experiences with multi-dimensional topic models. In *Proceedings of HLT-NAACL*, 168–178.

Pavalanathan, U., and Eisenstein, J. 2015. Confounds and consequences in geotagged twitter data. In *Proceedings of EMNLP*, 2138–2148. Lisbon, Portugal: ACL.

Pennebaker, J. W.; Francis, M. E.; and Booth, R. J. 2001. Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates* 71:2001.

Pflanz, S. 2001. Occupational stress and psychiatric illness in the military: investigation of the relationship between occupational stress and mental illness among military mental health patients. *Military Medicine* 166(6):457.

Powers, R. 2014. Can I Use My Cell Phone During Basic Training?

Rand, D. G.; Kraft-Todd, G.; and Gruber, J. 2015. The collective benefits of feeling good and letting go: Positive emotion and (dis) inhibition interact to predict cooperative behavior. *PloS One* 10(1):e0117426.

Russell, K. L.; Broderick, M. P.; Franklin, S. E.; Blyn, L. B.; Freed, N. E.; Moradi, E.; Ecker, D. J.; Kammerer, P. E.; Osuna, M. A.; Kajon, A. E.; and others. 2006. Transmission dynamics and prospective environmental sampling of adenovirus in a military recruit setting. *Journal of Infectious Diseases* 194(7):877–885.

Schoenebeck, S. Y. 2013. The Secret Life of Online Moms: Anonymity and Disinhibition on YouBeMom. com. In *Proceedings of ICWSM*.

Segal, D. R., and Segal, M. W. 2004. *America's military population*, volume 59. Population Reference Bureau Washington, DC.

Siebold, G. L. 2001. Core issues and theory in military sociology. *Journal of Political and Military Sociology* 29(1):140–159.

Soni, S.; Mitra, T.; Gilbert, E.; and Eisenstein, J. 2014. Modeling Factuality Judgments in Social Media Text. In *Proceedings of ACL*.

Starbird, K.; Maddock, J.; Orand, M.; Achterman, P.; and Mason, R. M. 2014. Rumors, false flags, and digital vigilantes: Misinformation on Twitter after the 2013 Boston Marathon Bombing. In *Proceedings of iConference*.

Tausczik, Y. R., and Pennebaker, J. W. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology* 29(1):24–54.

Yang, Y., and Eisenstein, J. 2013. A Log-Linear Model for Unsupervised Text Normalization. In *Proceedings of EMNLP*, 61–72. ACL.