

Investigating the Documentation of Electronic Cigarette Use in the Veteran Affairs Electronic Health Record: A Pilot Study

Danielle L. Mowery

Dept. of Biomedical Informatics
University of Utah
Salt Lake City, UT 84108, USA
danielle.mowery@utah.edu

Brett R. South

Dept. of Biomedical Informatics
University of Utah
Salt Lake City, UT 84108, USA
Brett.South@hsc.utah.edu

Shu-Hong Zhu

Dept. of Family Medicine & Public Health
University of California San Diego
La Jolla, CA 92093, USA
szhu@ucsd.edu

Olga V. Patterson

Division of Epidemiology
University of Utah
Salt Lake City, UT 84132, USA
olga.patterson@utah.edu

Mike Conway

Dept. of Biomedical Informatics
University of Utah
Salt Lake City, UT 84108, USA
mike.conway@utah.edu

Abstract

In this paper, we present pilot work on characterising the documentation of electronic cigarettes (e-cigarettes) in the United States Veterans Administration Electronic Health Record. The Veterans Health Administration is the largest health care system in the United States with 1,233 health care facilities nationwide, serving 8.9 million veterans per year. We identified a random sample of 2000 Veterans Administration patients, coded as current tobacco users, from 2008 to 2014. Using simple keyword matching techniques combined with qualitative analysis, we investigated the prevalence and distribution of e-cigarette terms in these clinical notes, discovering that for current smokers, 11.9% of patient records contain an e-cigarette related term.

1 Introduction

Electronic cigarettes — e-cigarettes — were developed in China in the early 2000s and first introduced to the US market in 2007. Once established in the US, the product experienced explosive growth, with the number of e-cigarette users doubling every year between 2008 and 2012 (Grana et al., 2014). In 2012 it was estimated that 75% of US adults had heard of e-cigarettes, and

8.1% had tried them (Zhu et al., 2013). By 2014, the proportion of adult Americans who had tried e-cigarettes increased to 12.6% (Schoenborn and Gindi, 2015).

Public health practitioners, government regulatory authorities, professional associations, the media, as well as individual clinicians and health workers are divided as to whether e-cigarettes represent an exciting new smoking cessation opportunity (Green et al., 2016; McNeill et al., 2015; Caponnetto et al., 2013) or are an untested, potentially dangerous technology that risks undermining recent successes in “denormalising” smoking (Choi et al., 2012; Etter et al., 2011; Gornall, 2015; U.S. Department of Health and Human Services, 2016; Department of Health and Human Services, 2014).

Currently, little is known about how clinicians “on-the-ground” advise patients who use, or are considering using, e-cigarettes. While Winden et al. (2015) has gone some way to describing e-cigarette Electronic Health Record (EHR) documentation behaviour in the context of a medical system in Vermont, national patterns in e-cigarette documentation have not been explored. In this paper, we present pilot work on characterising the documentation of e-cigarettes in the United States Veterans Administration Electronic Health Record. The Veterans Health Administration (VA) is the largest health care system in the

United States with 1,233 health care facilities nationwide, serving 8.9 million veterans per year. VA EHR data provides the opportunity for nationwide population-health surveillance of e-cigarette use.

The remainder of this document consists of five sections. Following a discussion of related work in Section 2, Section 3 describes both our cohort selection procedure, and our method of identifying e-cigarette documentation in clinical notes, while Sections 4 and 5 present the results of our analysis, and some discussion of those results. The final section outlines some broad conclusions.

2 Background

The VA collects data about patient smoking history and status using several approaches at the time of a patient encounter. Most patient clinical encounters have an associated *health factor* (i.e. semi-structured data that describes patient smoking status or smoking history (Barnett et al., 2014)). In addition, if the veteran has received dental care, the VA dental data contains descriptions of patient smoking status as a coded database field. However, neither of these data sources can be used to define what type of tobacco the patient uses and more specifically, if the patient uses e-cigarettes. This information is only found embedded in clinical text.

Given the rapid rise in popularity of e-cigarettes, and the lack of adequate public health surveillance systems currently focussing on these novel tobacco products, various methods and data sources have been used to understand changes in e-cigarette prevalence and usage patterns, including analysing search engine queries relevant to e-cigarettes (Ayers et al., 2011), mining social media data (Myslín et al., 2013; Chen et al., 2015), and — the focus of this paper — analysing EHR data for e-cigarette related documentation (Windén et al., 2015).

Previous work on smoking status identification in the EHR context has focussed on *structured data* (e.g. Wiley et al. (2013) used ICD-9 codes successfully to identify current smokers in the Vanderbilt Medical Center EHR), *semi-structured data* (e.g. McGinnis et al. (2011) used VA EHR *health factors* to reliably identify current smokers), and *unstructured data* (e.g. Clark et al. (2008); Savova et al. (2008); Da Silva et al. (2011) applied natural language processing methods to

EHR clinical notes to identify smoking status).

EHR corpus analysis has been the focus of several research efforts in the tobacco domain. For example, Chen et al. (2014) investigated the documentation of general tobacco use in clinical notes from Vermont’s Fletcher Allen Health Center, discovering that free-text clinical notes are frequently used to document amount of tobacco used, tobacco use frequency, and start and end dates of tobacco use (i.e. important clinical information that is difficult to represent with structured data). In follow-up work focussing specifically on e-cigarettes rather than general tobacco use, Windén et al. (2015), again using EHR data from Fletcher Allen Health Center, developed a sophisticated annotation scheme to code e-cigarette documentation, with categories including *dose*, *device type*, *frequency*, and *use for smoking cessation*. One result of particular note garnered from this research is the observation that less than 1% of patients had e-cigarette mentions in their note.

In this pilot study, our aim is to complete an initial corpus analysis of VA patient record data with the goal of quantifying the frequency with which e-cigarette usage is documented within the VA patient record.

3 Materials and Methods

We queried the VA dental record data found in the VA Corporate Data Warehouse to identify a national cohort of all Veterans Affairs patients with a coded history of current (or current and past) smoking between the years 2008-2014. Dental records were chosen as a data source as they are believed to be the most reliable indicators of smoking status in the VA context. From these data we identified 87,392 unique patients (77,491 current smokers, 9,901 current and past smokers). We then selected a random sample of 2,000 patients and extracted their associated clinical notes yielding 154,991 clinical notes. Note types include progress notes, consultation notes, consent documents, instructions, triage notes, history and physical notes, amongst others.

Based on an iterative process of corpus exploration, along with insights gleaned from previous work on e-cigarette related natural language processing (Myslín et al., 2013; Windén et al., 2015), we identified twenty e-cigarette related terms (listed in Table 1), and — using these terms — performed a keyword search within the pa-

tient clinical notes. We reviewed each e-cigarette term instance in its context to ascertain whether the e-cigarette term instance actually referred to e-cigarette usage.

We report the precision of each e-cigarette term defined as the proportion of term match instances actually referencing e-cigarette usage of all term matches.

4 Results

Term	Total	TP	FP	Precision
ecig	14	14	0	100.0
electronic cig	10	10	0	100.0
liquid nicotine	5	5	0	100.0
Ecig	4	4	0	100.0
E CIG	4	4	0	100.0
electric cig	1	1	0	100.0
ecigarette	1	1	0	100.0
Ecigarette	1	1	0	100.0
E Cig	1	1	0	100.0
vape	19	18	1	0.947
e cig	7	6	1	0.857
Vape	6	5	1	0.833
VAPOR	9	7	2	0.778
VAPE	2	1	1	0.500
vapor	241	81	160	0.336
vaporizer	192	36	156	0.188
Vapor	73	4	69	0.055
Vaporizer	3	0	3	0.000
VAPORIZER	3	0	3	0.000
ECIG	5	0	5	0.000
Total	601	199	402	–

Table 1: Proposed e-cigarette related terms ranked by precision

We analysed notes from 2,000 VA patients. From these notes, we observed 238 patients (11.9%) with one or more e-cigarette mentions within their notes (see **Figure 1**). In total, there were 601 mentions, with 436 notes containing more than one mention. Of these 601 mentions, 199 (33.1%) mentions described true e-cigarette usage (**Table 1**) as ascertained by manual inspection. The most frequent e-cigarette term matches included variants of the term *vapor* (*vapor*: 241, *vaporizer*: 192, *Vapor*: 73). These terms were also the most frequent sources of false positives (*vapor*: 160, *vaporizer*: 156, and *Vapor*: 69). Thirteen of the twenty terms yielded precision scores greater than 0.500. Of these high-precision terms,

the most prevalent terms included *vape*: 19, *ecig*: 14, and *electronic cig*: 10.

5 Discussion

We observed a variety of linguistic contexts describing e-cigarette usage. Patients report use of e-cigarettes with other tobacco products (e.g., “smokes 10 tobacco cigs per day and uses *vape*”). Similar to tobacco cessation, clinicians report providing encouragement and counselling for patients to stop e-cigarette use. Patients often contemplate e-cigarettes as an alternative to tobacco usage (e.g., “thinking about switching to *ecig*”) or as an approach to tobacco cessation (e.g., “uses *nicotine vaporizer* and hasn’t smoked tobacco in 6 mos”). This was not a surprising finding given that, according to the Centers for Disease Control, “among current cigarette smokers who had tried to quit smoking in the past year, more than one-half had ever tried an e-cigarette and 20.3% were current e-cigarette users” (Schoenborn and Gindi, 2015). Patients reported differing experiences of using e-cigarettes as a smoking cessation aid, with one patient stating directly that e-cigarettes were an ineffective tool in his struggle to quit smoking. Consistent with current uncertainty regarding the safety of e-cigarettes and their utility as a smoking cessation aid, not all clinicians support the use of e-cigarettes as a safe alternative to tobacco usage (e.g., “I do not recommend *ecig/vapor*”).

Analogous to the “packs-per-day” metric used by clinicians to document volume of combustible tobacco use, patients report their frequency of e-cigarette use in volume over time (e.g., “6mg/day”). E-cigarette usage goals are often set by both clinicians (“reducing consumption from 9 grams to 3 with goal of quitting”) and patients (“using *e cig* and cutting back by half”) alike. One clinician reported a patient’s use of e-cigarettes with “no side effects with current meds” suggesting that clinicians are aware that known side effects with medication use is a possibility.

Although most of the twenty e-cigarette terms used in this study yielded precision scores greater than 0.500, we also observed a substantial proportion of term matches that did not indicate actual e-cigarette usage. Many false positives occurred due to the ambiguous nature of the word *vaporizer* and its variants. For example, the domestic use of a vaporizer to increase room humidity, the treatment of patients with over-the-counter sinus

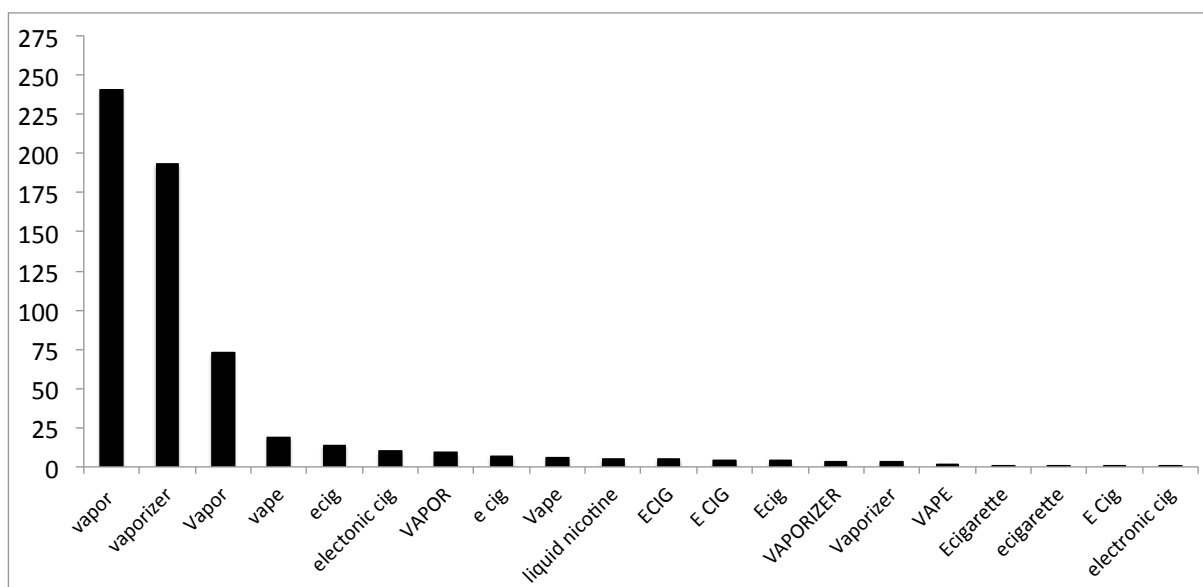


Figure 1: Frequency of e-cigarette terms in the clinical notes of a cohort of 2,000 VA smokers

relief (“Vicks *vapor* rub”), or the use of vaporizers to ingest medical marijuana (“MJ via *vaporizer*”). These non-relevant mentions of e-cigarette related terms are most frequently found in anaesthesia consent notes (n=185 mentions).

From notes containing matched e-cigarette variants, we discovered several co-occurring terms which could improve the term’s precision, with examples including *nicotine vaporizer*, *vapor-nicotine*, *vapor cig*, *vapor cigarettes*, *vapor pens*, *vapor cigarets*, *methonol vapor*, and *vapor nicotine*.

The pilot work described in this short paper has several limitations. First, our list of e-cigarette related keywords was limited to twenty. As indicated above, there may well be additional high precision e-cigarette related terms that we did not use in this work. Second, unlike Winden et al. (2015) we have not conducted a large scale annotation effort or mapped to an annotation scheme. Finally, while the VA is the largest integrated medical system in the United States, and the only nationwide system, VA patients are not necessarily representative of the general population. It is particularly important to note that approximately 92% of veterans are male (National Center for Veterans Analysis and Statistics, 2013).

6 Conclusion

In conclusion, we have demonstrated that for current smokers, e-cigarette terms are present in 11.9% (238) of VA patient records. Of this 11.9%

of patients, it is estimated that around two thirds of e-cigarette mentions are false positives, suggesting that around 4% of smokers have e-cigarette use documented in their clinical notes.

Acknowledgments

This research was supported by the National Library of Medicine of the National Institutes of Health under award numbers R00LM011393 & K99LM011393, and University of Utah funds.

Ethics Statement

This study was approved by the University of Utah Institutional Review Board (IRB_00088382).

References

- John W. Ayers, Kurt M. Ribisl, and John S. Brownstein. 2011. Tracking the rise in popularity of electronic nicotine delivery systems (electronic cigarettes) using search query surveillance. *American Journal of Preventive Medicine* 40(4):448–453.
- Paul Barnett, Adam Chow, and Nicole Flores. 2014. Using health factors data for VA health service research. Technical Report 28, Health Economic Resource Center. https://www.herc.research.va.gov/files/RPRT_768.pdf.
- Pasquale Caponnetto, Cristina Russo, Cosimo Bruno, Angela Alamo, Maria Amaradio, and Ricardo Polosa. 2013. Electronic cigarette: a possible substitute for cigarette dependence. *Monaldi Arch Chest Dis* 79(1):12–9.

- Annie T Chen, Shu-Hong Zhu, and Mike Conway. 2015. What online communities can tell us about electronic cigarettes and hookah use: A study using text mining and visualization techniques. *J Med Internet Res* 17(9):e220.
- Elizabeth S Chen, Elizabeth W Carter, Indra Neil Sarkar, Tamara J Winden, and Genevieve B Melton. 2014. Examining the use, contents, and quality of free-text tobacco use documentation in the electronic health record. *AMIA Annu Symp Proc* 2014:366–74.
- Kelvin Choi, Lindsey Fabian, Neli Mottey, Amanda Corbett, and Jean Forster. 2012. Young adults' favorable perceptions of snus, dissolvable tobacco products, and electronic cigarettes: findings from a focus group study. *Am J Public Health* 102(11):2088–93.
- Cheryl Clark, Kathleen Good, Lesley Jezierny, Melissa Macpherson, Brian Wilson, and Urszula Chajewska. 2008. Identifying smokers with a medical extraction system. *J Am Med Inform Assoc* 15(1):36–9.
- Lalindra Da Silva, Thomas Ginter, Tyler Forbus, Neil Nokes, Brian Fay, Ted Mikuls, Grant Cannon, and Scott DuVall. 2011. Extraction and quantification of pack-years and classification of smoker information in semi-structured medical records. In *Proceedings of the 28th International Conference on Machine Learning*, Bellevue, WA.
- Department of Health and Human Services. 2014. The health consequences of smoking – 50 years of progress. Technical report, Surgeon General.
- Jean-François Etter, Chris Bullen, Andreas D Flouris, Murray Laugesen, and Thomas Eissenberg. 2011. Electronic nicotine delivery systems: a research agenda. *Tob Control* 20(3):243–8.
- Jonathan Gornall. 2015. Public Health England's troubled trail. *BMJ* 351:h5826.
- Rachel Grana, Neal Benowitz, and Stanton A Glantz. 2014. E-cigarettes: a scientific review. *Circulation* 129(19):1972–86.
- Sharon H Green, Ronald Bayer, and Amy L Fairchild. 2016. Evidence, policy, and e-cigarettes—will England reframe the debate? *N Engl J Med* 374(14):1301–3.
- Kathleen A McGinnis, Cynthia A Brandt, Melissa Skanderson, Amy C Justice, Shahida Shahrir, Adeel A Butt, Sheldon T Brown, Matthew S Freiberg, Cynthia L Gibert, Matthew Bidwell Goetz, Joon Woo Kim, Margaret A Pisani, David Rimland, Maria C Rodriguez-Barradas, Jason J Sico, Hilary A Tindle, and Kristina Crothers. 2011. Validating smoking data from the Veteran's Affairs health factors dataset, an electronic data source. *Nicotine Tob Res* 13(12):1233–9.
- Ann McNeill, Leonie Brose, Robert Calder, and Sara Hitchman. 2015. E-cigarettes: an evidence update - report commissioned by Public Health England. Technical report, Public Health England.
- Mark Myslín, Shu-Hong Zhu, Wendy Chapman, and Mike Conway. 2013. Using Twitter to examine smoking behavior and perceptions of emerging tobacco products. *J Med Internet Res* 15(8):e174.
- National Center for Veterans Analysis and Statistics. 2013. Women veteran profile. Technical report, United States Department of Veterans Affairs. https://www.va.gov/vetdata/docs/SpecialReports/Women_Veteran_Profile5.pdf.
- Guergana K Savova, Philip V Ogren, Patrick H Duffy, James D Buntrock, and Christopher G Chute. 2008. Mayo clinic NLP system for patient smoking status identification. *J Am Med Inform Assoc* 15(1):25–8.
- Charlotte Schoenborn and Renee Gindi. 2015. Electronic cigarette use among adults: United States, 2014. Technical Report 217, NCHS. <https://www.cdc.gov/nchs/data/databriefs/db217.pdf>.
- U.S. Department of Health and Human Services. 2016. E-cigarette use among youth and young adults: a report of the Surgeon General. Technical report.
- Laura K Wiley, Anushi Shah, Hua Xu, and William S Bush. 2013. ICD-9 tobacco use codes are effective identifiers of smoking status. *J Am Med Inform Assoc* 20(4):652–8.
- Tamara J Winden, Elizabeth S Chen, Yan Wang, Indra Neil Sarkar, Elizabeth W Carter, and Genevieve B Melton. 2015. Towards the standardized documentation of e-cigarette use in the electronic health record for population health surveillance and research. *AMIA Jt Summits Transl Sci Proc* 2015:199–203.
- Shu-Hong Zhu, Anthony Gamst, Madeleine Lee, Sharon Cummins, Lu Yin, and Leslie Zoref. 2013. The use and perception of electronic cigarettes and snus among the US population. *PloS One* 8(10):e79332