Small cohort of patients with epilepsy showed increased activity on *Facebook* before sudden unexpected death

Ian B. Wood [a,1], Rion Brattig Correia [b,c,a,1], Wendy R. Miller [d,*], Luis M. Rocha [e,a,b]

[a] *Center for Social and Biomedical Complexity, Luddy School of Informatics, Computing & Engineering, Indiana University, Bloomington, IN 47408, USA*
[b] *Instituto Gulbenkian de Ciência, Oeiras 2780-156, Portugal*
[c] *CAPES Foundation, Ministry of Education of Brazil, Brasília, DF, Brazil*
[d] *School of Nursing, Indiana University, Indianapolis, IN 46202, USA*
[e] *Department of Systems Science and Industrial Engineering, Binghamton University, Binghamton, NY 13902, USA*

ARTICLE INFO

ABSTRACT

Sudden Unexpected Death in Epilepsy (SUDEP) remains a leading cause of death in people with epilepsy. Despite the constant risk for patients and bereavement to family members, to date the physiological mechanisms of SUDEP remain unknown. Here we explore the potential to identify putative predictive signals of SUDEP from online digital behavioral data using text and sentiment analysis tools. Specifically, we analyze *Facebook* timelines of six patients with epilepsy deceased due to SUDEP, donated by surviving family members. We find preliminary evidence for behavioral changes detectable by text and sentiment analysis tools. Namely, in the months preceding their SUDEP event patient social media timelines show: i) increase in verbosity; ii) increased use of functional words; and iii) sentiment shifts as measured by different sentiment analysis tools. Combined, these results suggest that social media engagement, as well as its sentiment, may serve as possible early-warning signals for SUDEP in people with epilepsy. While the small sample of patient timelines analyzed in this study prevents generalization, our preliminary investigation demonstrates the potential of social media data as complementary data in larger studies of SUDEP and epilepsy.

© 2022 Published by Elsevier Inc.

1. Introduction

Sudden Unexpected Death in Epilepsy (SUDEP) remains a leading cause of death for people with epilepsy (PWE), and includes all epilepsy-related deaths not due to trauma, drowning, status epilepticus, or other identifiable causes. The incidence of SUDEP is about 0.35 cases per 1,000 person-years [1]. While research into the physiological mechanisms underlying SUDEP continues to be thoroughly studied, and new SUDEP-related guidelines for clinicians treating PWE have been published in order to minimize SUDEP risk, SUDEP incidence remains steady [2,3]. To date, the most espoused preventive strategy for SUDEP remains seizure control via appropriate self-management [4], and especially medication adherence, since a clear risk factor for SUDEP is a higher frequency of seizures [5]. While these risk factors have been

disseminated broadly, including to the public, SUDEP remains a leading cause of death for PWE, leading organizations such as The Institute of Medicine, American Epilepsy Society, and Epilepsy Foundation to call for increased study into SUDEP.

Apart from research related to the ways in which providers, patients, and their families discuss SUDEP [2,6], very little behavioral research has been conducted to reveal potential behavioral or social attributes that may precede SUDEP. Should such specific attributes exist, they would provide an area of preventive intervention for SUDEP. In this study, we utilized digital behavioral data and investigated their potential for uncovering behavioral signatures preceding SUDEP that could be leveraged as early-warning signals to inform self-management interventions in PWE. As patients are known to not fully recall important events or even display recognizable behavior change during clinical consultations, digital behavioral data, such as social media data, can offer a complementary view of patient behavior of clinical significance [7]. Specifically, we use text and sentiment analysis to evaluate temporal changes in emotional states and communication patterns of the subjects in the study. The methodology gives us the unique oppor-

tunity to examine longitudinally the emotional states of a cohort of PWE with a known outcome of SUDEP. Our preliminary results show that social media may reveal behavioral experiences leading up to SUDEP, and thus guide areas for SUDEP-preventing interventions. This study also demonstrates the successful use of alternative, real-world data sources in studying SUDEP [7,8].

Psychological stress is known to increase the risk of certain diseases, like the common cold [9]. Directly related to PWE, stress and major life events are known to increase the risk of seizures, which in turn can increase the risk of SUDEP [10,11]. However, direct physiological measurements of stress involves expensive and invasive tools. A compelling alternative is to measure stress and other cognitive states indirectly in self-reported digital behavioral data, such as in social media posting on *Facebook.* This is one of the focuses of the interdisciplinary field of *affective computing*, which has developed methods to measure human emotion (including stress) via linguistic and other computer-based features, such as keystroke dynamics [12]. For instance, Pennebaker [13] found a correspondence between textual features and physiological signals of stress. Similarly, Vizer, Zhou, & Sears [14] found that increased lexical complexity (diversity of words) tends to correspond with increased physical or cognitive stress. However, such studies are often conducted in controlled laboratory conditions, asking participants to write essays with particular prompts. This is not the case with social media, where users write posts spontaneously without being prompted in laboratory settings. Our assumption is that stress and other mood states influence whether and how a social media post is written, and can thus be measured via textual analysis of those posts. A substantial body of literature already reports that social media data enable quantitative measurement and prediction of various behavioral processes of biomedical relevance, i.e. a real-world data source to study "humans as their own model organism" [7]. Indeed, social media data have already been shown to be useful, alone or in combination with other data sources, for a variety of other biomedical problems. For instance, data from Twitter and Instagram help in the detection of health conditions including the spread of flu pandemics [15], warning signals of drug adverse reactions [16], human reproduction [17], and even depression [18]. Social media users who self-reported their diagnosis of depression have been shown to exhibit distorted modes of thinking (cognitive distortions) in their writing, an early warning that can lower the burden of this underdiagnosed condition and leading cause of disability worldwide [19]. A long list of successful applications using social media data for biomedical and health-related problems is discussed in our recent review [7].

To infer relevant cognitive states in our cohort of deceased SUDEP subjects we use textual and sentiment analysis of their social media posts. These methods were originally developed to determine the positive or negative feelings expressed in natural language texts toward specific product ratings, often used for marketing purposes [20,21]. However, a number of sentiment analysis tools have been developed from psychological experiments, and can be used to model the emotional states of authors based on their written text [7]. In fact, sentiment analysis has been very useful to track various individual and cohort-specific behaviors of relevance to biomedicine, especially mental health [17,7,19]. Similarly to other domains, these computational methods are likely to be useful to characterize the behavior of SUDEP cohorts, including any possible stress markers hidden in their social media discourse that can be leveraged to inform interventions aimed at improving self-management, a key predictor of epilepsy-related outcomes. Next, we detail the data gathering, textual methods, and three different sentiment analysis tools we apply to our SUDEP cohort.

2. Materials and methods

We began by eliciting families from which a member was known to have died of SUDEP. To do so, we advertised our research goals on the bulletin boards of the Epilepsy Foundation website and epilepsy-related *Facebook* groups. We also distributed information about our study to the Epilepsy Foundation's SUDEP Institute, which passed on the information to members of SUDEP bereavement groups within the Institute. The Epilepsy Foundation website is one of the most popular sites for people with epilepsy. All procedures were evaluated by the Indiana University Institutional Review Board, who ultimately deemed that the study was exempt/not human subjects research. Family members self-referred to our study via email, and were given information about the study, its goals, and were also informed that participation was voluntary. We received about 20 inquiries from families who wanted to donate social media content from their deceased family members. From these, a majority of users had *Facebook* accounts, and only a few had *Twitter* or *Instagram* accounts. Due to data availability, we decided to focus our analysis solely on *Facebook* timelines. This yielded a small cohort of $n = 12$ *Facebook* timelines (four males and eight females) from which we had timelines to collect data from. For six subjects, we obtained full login information, and for the remaining we had varying viewing access to timeline posts, as listed in Table 1.

Data collection for subjects with login information was conducted through an in–house developed application using *Facebook*'s official application programming interface (API). Family members logged into the deceased *Facebook* account and accessed a specific app webpage. The app then collected all of the subject's timeline posts, including text, meta-data (e.g., date, posting device, etc), and the number of likes, comments, and shares. Similarly, when only viewing access to the subject's timeline was available, family members (or a researcher when family was unable/unavailable) were instructed to scroll the deceased timeline, thus loading all posts, and export the subject's timeline content as an HTML file. A script developed in-house was used to process the HTML file, collecting text, available meta-data, and number of likes, comments, and shares from posts. Importantly, unlike the app-collected timelines that made use of subject's login information, timelines collected via the HTML-scraping script may not contain all subject posts, as privacy settings putatively put in place by the subject may have blocked the person collecting the data from viewing them in the first place. In addition, in 2009 *Facebook* made a significant change to their interface: the prompt to the post box changed from "Update your Status", followed by "<Subject name> is…" to "What's on your mind?". Naturally, we believe this interface change may elicit a different response from the user. To avoid any possible interface bias in our analysis, we only consider subject posts that occurred after 2009, when the change took place. All collected data were securely stored within our servers for further analysis. For each subject Table 1 lists basic demographic, subject posting time range, and any notable life event discussed by the subject on their *Facebook* timeline in the month preceding their SUDEP, which was manually annotated by the researchers.

The number of posts collected for each subject varies widely, from only 4 posts written by Subject 12, all the way to 2,271 posts written by Subject 2 (see Table 1). The average number of posts per subject is 726. In total, we collected and processed 8,717 posts with text that were written after 2009, when considering all 12 subjects. However, because some subjects had very little number of posts—as is the case of Subject 12—we opted to limit our analysis to subjects with more than 500 posts that contained text and were written after 2009. In other words, next we only present

Table 1

Demographics and data collection details for study subjects. Six subject timeline posts were collected via a custom-built app accessed using subject's login and password information. Six subject timelines were collected via HTML scraping of pages as visible to the public, to *Facebook* friends, or to friends of friends (FoF), as noted. The number of posts column tallied only posts with written text after 2009 (due to a significant *Facebook* interface change).

Subj.	Collection	Sex	Age	Posts	Window of posts*	Notable life event before SUDEP
1	App	F	23	1,410	2,526	New apartment, job, and city
2	App	M	20	2,271	2,157	Releasing DVD copies of new movie
3	App	F	18	844	2,071	Lonely as new college freshman
4	App	F	24	273	1,865	Graduating a Master's program
5	App	M	14	51	911	Birthday
6	App	F	15	473	843	Return from Europe trip
7	FoF	F	29	62	2,334	n/a
8	Public	F	n/a	2,201	2,315	n/a
9	Public	F	n/a	10	52	Party and writing paper
10	Friend	M	24	984	2,373	Recent concussion and recovery
11	FoF	M	28	134	1,524	Hospitalization
12	Friend	F	16	4	413	Braces Removed

* Column "window of posts" denote the number of days between a subject's first and last posts.

results on subjects 1–3, 6, 8, and 10, a cohort of $n = 6$ subjects. These subjects are highlighted in Table 1.

Textual content of individual posts were processed using the dictionaries of three sentiment analysis tools: *Affective Norms for English Words* (ANEW) [22], *Valence Aware Dictionary for sEntiment Reasoning* (VADER) [23], and *Linguistic Inquiry and Word Count* (LIWC) [24]. These three tools are widely used in the sentiment analysis literature. In fact, VADER and LIWC were consistently among the best tools for 3-class polarity classification (negative, neutral, or positive emotion) across a number of corpora in a benchmark comparison study [25].

Dictionaries were used to match against single words in subject posts. Matched words were then scored over several sentiment and textual dimensions per post. For instance, ANEW includes ratings from 1 to 9 in a dictionary of 1,034 words along three dimensions: *valence*, from unhappy to happy; *arousal* from calm to excited; and *dominance* from controlled to in-control. These ratings were originally collected from surveys given to undergraduates in a psychology class using a 9-point Likert-like scale [22]. We used ANEW to find the mean sentiment along these three dimensions for each post by averaging the sentiments of each word, while neglecting words absent from the dictionary. VADER [23] is a tool for measuring the intensity of positive or negative affect through lexical scores modified by syntactical rules, and is readily available as part of the Natural Language Toolkit for python [26]. In addition to dictionary-based sentiment scores, VADER looks at nearby words and modifies sentiment scores based on 5 simple rules: the presence of exclamations, capitalization, adverbs, negations, and contrasting conjunctions. Using this tool, we computed normalized scores describing the intensity of positive, neutral, and negative emotion present in each subject post. LIWC (pronounced Luke) is the third dictionary-based tool used. It was developed with a well-documented procedure of consistent categorization between a majority of human judges. The latest version of the software, LIWC2015, has dictionaries containing nearly 6,400 words and evaluates text across nearly 90 linguistic and sentiment variables, including summary variables, pronouns, articles, cognitive processes, time focus, personal concerns, and informal language categories [24,27].

3. Results

Assuming some type of stressor prior to SUDEP, which in turn could manifest as a change in the subject's digital verbosity, first we characterize the number of words per subject *Facebook* post (word count) with a simple negative binomial regression. The binomial regression tests whether there was a significant difference in the amount of words per post when comparing posts written in two different epochs of the subject's digital behavior. More specifically, we compare the average number of words per post in the two months (56 days) preceding the subject's SUDEP against the average number of words per post in the rest of the available timeline. We choose the last two months as a conservative time range for a subject behavioral change that at the same time holds enough examples (posts) for a robust statistical analysis—as a 10 samples minimum is a frequently recommended heuristic for an accurate estimation of model parameters [28]. However, we note that posting behavior varies between subjects and we do not know whether, or when, stressors proceedings SUDEP may appear for each subject. We also tested different epochs, ranging from one to twelve weeks prior to SUDEP. Results are consistent for subjects with sufficient data in the last period being considered, and are shown in Fig. S1. From our six analyzed subjects, subjects 1, 2, 6, and 10 had significantly higher word count in the two months preceding their SUDEP. Subject 8 also had a higher word count in the last two months, albeit not significant at $p < 0.05$. Conversely, subject 3 had a significantly lower word count in the last two months. Results are shown in Table 2 and Fig. 1 shows the average word count for each subject timeline. Two regressions are fitted to the data highlighting the slope of the increase (or decrease) in subject verbosity: one considering the complete subject timeline (dotted line) and one only considering the last two months of posts (solid line).

Since digital behavioral changes may be reflected not only in post length but also in how frequent posts are made, next we use a zero-inflated negative binomial regression to examine whether the observed verbosity (word count) and frequency of posting prior to SUDEP was significantly different from subject's previous epochs. A zero-inflated negative binomial regression is an extension of the binomial regression where there is an assumption that a different process governs the likelihood that a subject

Table 2

Significance tests for differences in word counts in posts during the last two months preceding SUDEP compared to other posts. The mean word count for the posts written during the last two months (μ_{last} with n_{last} samples) are compared to the mean word count of all other posts written by the subject before this period (μ_{early} with n_{early} samples). Significance is estimated from a negative binomial regression, with $p < 0.05$ highlighted in bold. Subjects are ordered according to the rank-product of the number of samples during the last month and the number prior to the last month.

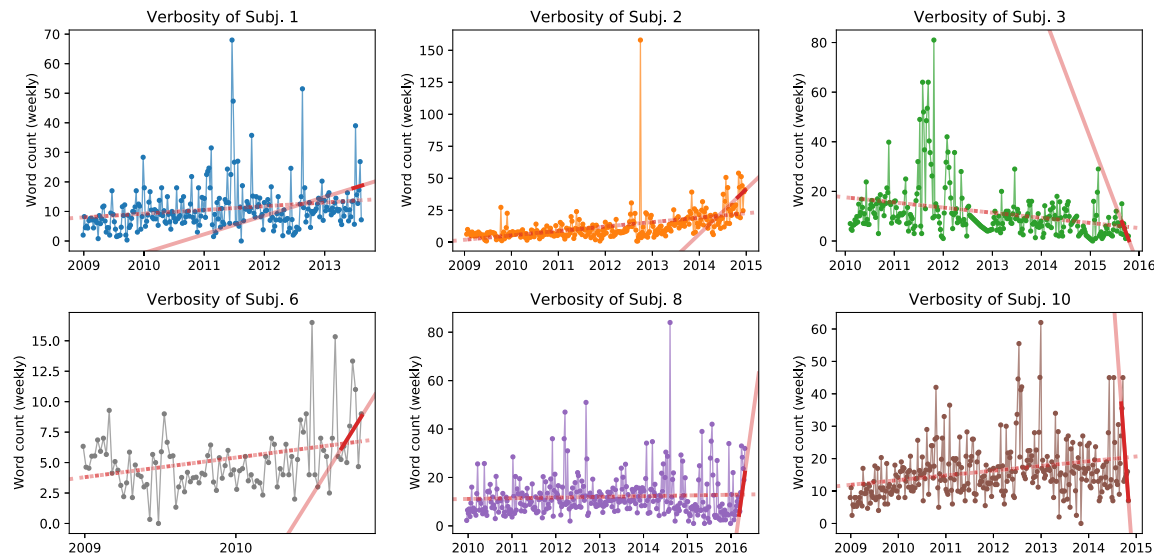| subject	n_{early}	n_{last}	μ_{early}	μ_{last}	$time_p$
2	2,162	109	12.431	**34.413**	**1.197e-32**
1	1,547	54	9.592	**17.889**	**4.146e-06**
8	2,185	16	12.070	**18.375**	0.081
6	7,17	23	5.252	**7.304**	**0.021**
10	1147	7	13.983	**23.571**	**0.048**
3	834	10	**11.125**	4.100	**0.001**

Fig. 1. Subject verbosity measured by word count. Values are shown as weekly average to improve readability. Dashed red line shows the trend over the entire range of subject's posts. Solid red line is the trend over the last two months of data with darker color denoting the period length.
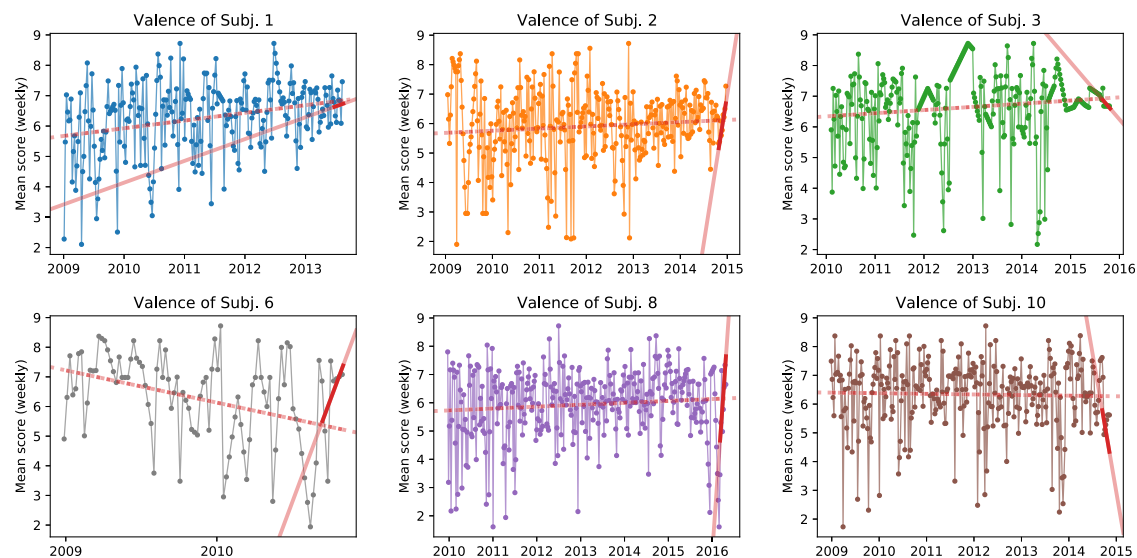

Fig. 2. Subject happiness measured by ANEW's Valence score. Values are shown as weekly average to improve readability. Dashed red line shows the trend over the entire range of subject's posts. Solid red line is the trend over the last two months of data with darker color denoting the period length.

makes no posts in a day (zero word count), which is then modeled by a logistic regression. Results are consolidated and presented in Table S2; different epochs considered are shown in Fig. S2. In general, we see that both subjects 1 and 2 were more likely to post in the two months preceding SUDEP, as well as writing longer posts. Perhaps due to increased model complexity, changes in subject 6's posting behavior are less significant, being less likely to post in weeks preceding SUDEP with little difference in the number of words written per day. Subjects 8 and 10 were significantly less likely to post in the final weeks before SUDEP, with a non-significant increase in words per day when they did. Lastly, subject 3 did not have a significant change in the number of days with a post, but did write significantly fewer words.

Having analyzed subject verbosity, we now turn to the sentiment of the text they wrote. We remember each sentiment dimension is calculated by averaging per-word sentiment scores

calculated for ANEW, LIWC, and VADER, three independent sentiment tools. In the following Figs. 2–4, line plots denote the average of a specific sentiment dimension measure over all posts each week. Some particular sentiment trends can be observed in these figures. For instance, four of the six subjects show an overtime increase in happiness sentiment, as measured by ANEW's valence dimension (see dotted lines in Fig. 2). Only two subjects, 3 and 10, show a decrease in happiness in the last two months (solid line). Importantly, Subject 3 has an overall happiness increase but a sharp sentiment shift in the last two months, reflected by her described feelings of loneliness of being a college freshmen. On the other hand, subject 6 has an overtime happiness decrease, but a sharp happiness increase in the last two months, reflecting a sentiment shift due to her European travels. Overall, despite some subjects having reversed valence sentiment, when their complete timeline sentiment is compared to the sentiment in the last
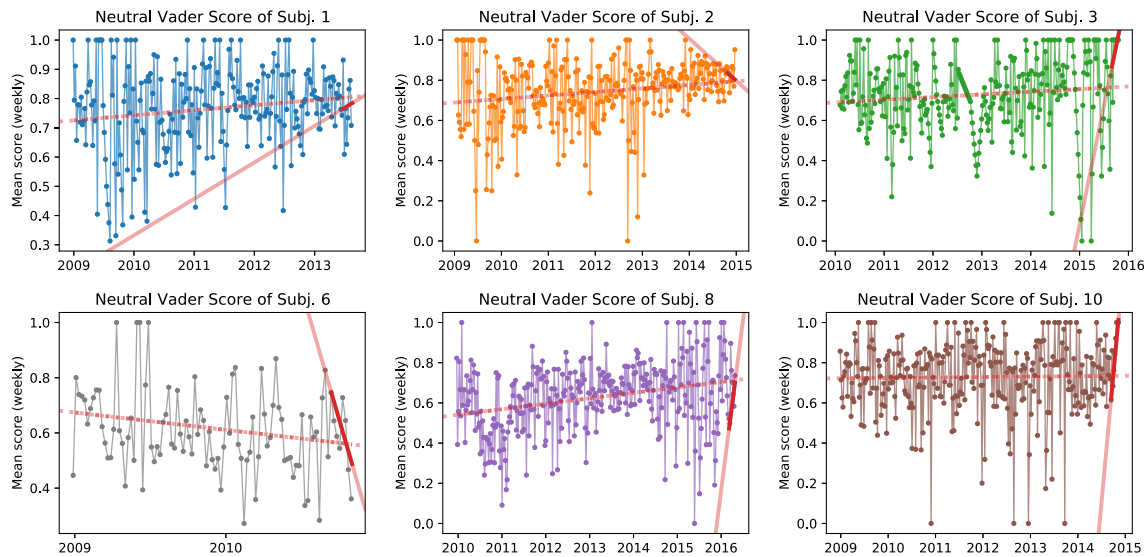
Fig. 3. Subject use of neutral words measured by VADER's Neutral score. Values are shown as weekly average to improve readability. Dashed red line shows the trend over the entire range of subject's posts. Solid red line is the trend over the last two months of data with darker color denoting the period length.
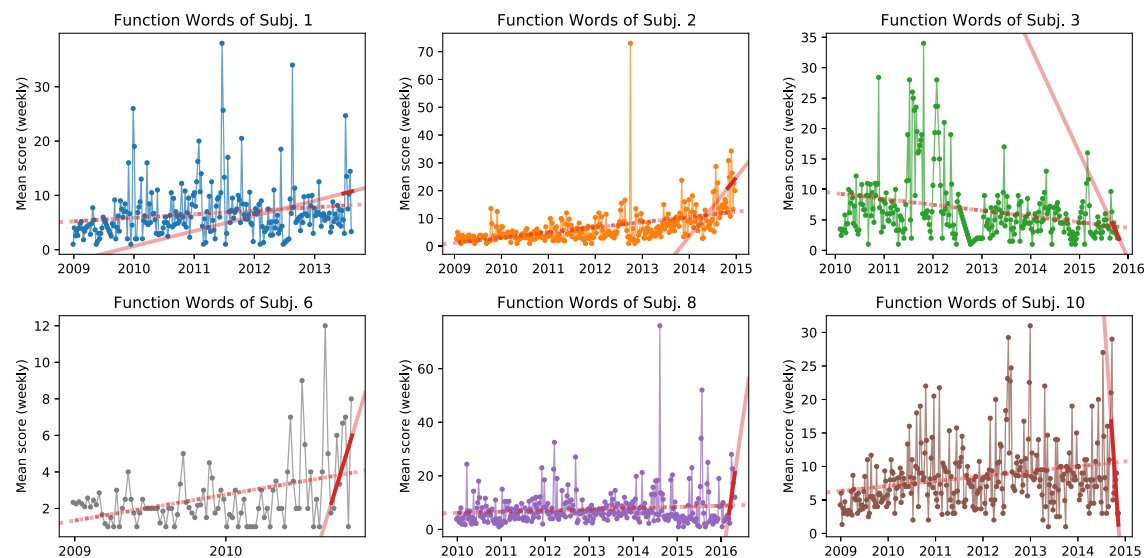

Fig. 4. Subject use of functional words measured by LIWC. Values are shown as weekly average to improve readability. The dashed red line shows the trend over the entire range of a subject's posts, while the solid red line is the trend over the last two months of data.

two months of posting, they all have something in common: a significant sentiment shift, as measured by the difference in slope of the two regressions.

To show this phenomenon is not simply an effect of the sentiment tool of choice, Figs. 3 & 4 show subject use of emotion-neutral words and functional words, measured by VADER and LIWC, respectively. Functional words include a broad category of words such as pronouns ('him', 'she'), articles ('the', 'a'), conjunctions ('and', 'but'), interjections ('oh', 'ah'), pro-sentences ('yes', 'no', 'okay'), and others. We observe an overtime increase in the average number of such words used per post for 5 of the 6 subjects (see Fig. 4). In addition, for 4 subjects the amount of functional words used increases substantially in the last two months of posting. In regard to emotion-neutral words, five of the six subjects show an increase use of emotion-neutral words—a sentiment dimension that other tools, such as ANEW, ignores (see Fig. 3). However, similarly to subject verbosity, all subjects have a drastic

shift in the analyzed sentiment categories when their complete timeline is compared to the last two months, again as measured by the regression slope (see red lines in aforementioned plots).

4. Discussion

First, we would like to emphasize that we cannot claim SUDEP causation, or the predictive accuracy of these tools applied to the social media posts of living individuals. However, the noticeable increase in functional words and the overall verbosity preceding SUDEP for a number of subjects is particularly suggestive of some detectable changes in the digital behavior of subjects, and that may serve as early-warning signals correlating with SUDEP. It is known that stress and major life events are likely to increase the risk of epilepsy [10,11], and that in turn may increase the risk of SUDEP. Several of our subjects had major life changes in the weeks preced-

ing their death, from concussions, moving to another city, returning from an overseas trip, or feeling lonely as a new college freshmen. In addition, the misuse of functional words has been associated with Aphasia, a language impairment attributed to the Wernicke's area, a brain area in the left (dominant) temporal-parietal region characterized by EEG abnormalities in patients with epilepsy [29–31]. Unlike impairment to Broca's area where patients speak slow, in hesitating ways, and phrases are devoid of functional words, impairment to the Wenicke's area cause patients to speak warmly and fluidly but using functional words with no content at all [29]. We manually checked sentence construction in the last two months of posting for our subjects and found no trace of functional word misuse aside from their increased occurrence. Nonetheless, if an increase in verbosity or changes in functional word use is indicative of stress or major life changes, the use of textual and sentiment tools may allow for a predictive, quantitative measure in larger studies, complementing current qualitative analyses. However, we do stress that the lack of appropriate sample size and a rigorous case-control in our current study hinders generalization of our findings at this point. Nonetheless, our preliminary results serve to invite additional research into this problem, especially to encourage attention to social media and other digital behavior data, thus contributing to better prediction of warning signals of SUDEP.

One possible avenue to evaluate the potential of sentiment analysis for predicting SUDEP is to employ statistical machine learning models using the text and sentiment analysis tools we described above. We attempted to build such models to predict changes in the last day or week of posts in a subject's timeline—instead of the last two months of posts shown in regressions above. However, we encountered two common machine-learning problems, especially in shorter window scenarios. The first was overfitting and the subsequent false positive prediction. Since sentiment tools possess many sentiment variables (dimensions), it is easy to perfectly fit posts used in training the algorithm. Yet, the resulting prediction/classification models do not generalize to predicting subject posts left out for testing. Stricter model regularization and dimensionality reduction methods can help, but in the end, using shorter prediction windows results in a classification scenario with a very large class imbalance with very few positive instances (i.e., posts preceding SUDEP) which does not allow automatic machine learning classification. This is because most posts occur when subjects are deemed healthy, and only very few instances can be safely set as being SUDEP related—those that happened right before death. Given this problem of class imbalance, classifiers for automatic prediction are not possible with our current dataset.

The second problem pertains to the labeling of posts as SUDEP-relevant. Assuming that only the last posts before SUDEP are relevant, may miss prior days and posts (positive instances) that may have been close calls for SUDEP. Without the proper labeling of these instances, our algorithms are potentially missing several learning opportunities. The two-month window prior to SUDEP we used in the regression analysis is reasonable for the observed cohort, allowing a reasonable amount of positive posts for most subjects (see Table S2). However, the regression serves as an observation tool more than an automatic predictor. Indeed, at the current stage, social media analysis can only enhance and provide a different perspective to other health data, such as electronic health records, personal diaries, epilepsy warning devices, service animals, etc. A more systemic and complete picture of SUDEP may emerge by combining these seemingly heterogenous data sources.

Going forward, our goal is to combine clinical (e.g., physician notes, laboratory exams, genetic profiling, questionnaire responses, electronic health records) with non-clinical digital behavioral data (e.g., electronic diaries, discussion boards, email exchange, phone usage patterns, social media posting, and consumption) into research design. This is planned via recruitment of patients with epilepsy who consent to the collection of their digital behavioral data, such as social media IDs [7]. Our own work with focus groups of patients with epilepsy and their caretakers have demonstrated willingness to donate digital behavioral data for studies. Indeed, as shown in the work we report here, this can be even done postmortem to avoid an observer bias—patients changing their behavior by knowing they are being observed. With enough subjects to account for the increase in variables, the next step is to validate the predictive power of social media signals in case-control experiments. We intend to focus on specific questions such as: why are subjects writing or using certain words more often prior to their death? Can this be statistically correlated with an increased risk of SUDEP? Can we pinpoint a behavioral phase shift to inform self- and caretaker-management as an early warning? The preliminary results we now report demonstrate the feasibility of extracting such signals. As we recruit additional subjects in planned larger studies, it will be possible to answer these questions more quantitatively and conclusively.

To compile additional digital behavioral data sources, our team is currently developing myAura [32], a personalized web service for epilepsy management. MyAura will include self-reported patient diaries, such as seizure tracking, food and water intake, medication adherence, physician encounters, among others. One of its goals is to test a variety of clinical and non-clinical temporal variables that may be proven useful in epilepsy management. The use of patient-donated social media timelines, as we have shown here, can prove to be the next frontier in informing our understanding of SUDEP and other epilepsy outcomes. MyAura will include the option for users to donate their social media timelines, thus allowing the recruitment of larger patient cohorts. Findings from analysis of the data of larger cohorts are likely to inform self-management recommendations for PWE, including allowing for SUDEP-predicting behaviors to be identified. For instance, patients with epilepsy could be monitored for an increased risk of SUDEP. In addition, our text and sentiment analysis could be used to inform individualized self-management interventions based on patient's posts and behaviors. At the same time, behavioral results can help direct physiologic studies, as cellular-level or biomarker changes can, for example, ultimately be correlated with behavioral experiences (e.g., cortisol and physiologic or psychological stress).

As a small pilot, our study has demonstrated the feasibility of mining social media data for SUDEP (and other epilepsy-related) research, as well as very preliminary findings regarding increased social media activity preceding SUDEP. While the sample size of this study is too small to render generalizations in terms of SUDEP prediction, our work here demonstrates the feasibility of a novel way of investigating epilepsy-related phenomena, including SUDEP. This work also demonstrates the value in the interdisciplinary collaboration between clinical/behavioral epilepsy researchers and informatics/complex systems scientists.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.yebeh.2022.108580.

References

[1] Bagnall RD, Crompton DE, Semsarian C. Genetic basis of sudden unexpected death in epilepsy. Front Neurol 2017;8:348.

[2] Miller WR, Young N, Friedman D, Buelow JM, Devinsky O. Discussing sudden unexpected death in epilepsy (SUDEP) with patients: Practices of health-care providers. Epilepsy Behavior 2014;32(Supplement C):38–41.

[3] Harden C, Tomson T, Gloss D, Buchhalter J, Cross JH, Donner E, et al. Practice guideline summary: sudden unexpected death in epilepsy incidence rates and risk factors: report of the guideline development, dissemination, and implementation subcommittee of the american academy of neurology and the american epilepsy society. Neurology 2017;88(17):1674–80.

[4] Smithson WH, Colwell B, Hanna J. Sudden unexpected death in epilepsy: addressing the challenges. Current Neurol Neurosci Rep 2014;14(12):502.

[5] Sudden unexpected death in epilepsy (sudep), https://www.cdc.gov/epilepsy/about/sudep/index.htm, accessed: 2018-08-02 (2018). .

[6] Stevenson MJ, Stanton TF. Knowing the risk of sudep: two family's perspectives and the danny did foundation. Epilepsia 2014;55(10):1495–500.

[7] Correia RB, Wood IB, Bollen J, Rocha LM. Mining social media data for biomedical signals and health-related behavior, Annual Review of Biomedical Data. Science 2020;3:433–58. https://doi.org/10.1146/annurev-biodatasci-030320-040844.

[8] Ramagopalan SV, Simpson A, Sammon C. Can real-world data really replace randomised clinical trials? BMC Med 2020;18(1):13. https://doi.org/10.1186/s12916-019-1481-8.

[9] Cohen S, Tyrrell DA, Smith AP. Psychological stress and susceptibility to the common cold. New England J Med 1991;325(9):606–12.

[10] McConnell H, Valeriano J, Brillman J. Prenuptial seizures: a report of five cases. J Neuropsychiatry Clinical Neurosci 1995;7(1):72–5.

[11] McKee HR, Privitera MD. Stress as a seizure precipitant: Identification, associated factors, and treatment options. Seizure 2017;44:21–6. https://doi.org/10.1016/j.seizure.2016.12.009.

[12] Zhai J, Barreto A. Stress detection in computer users based on digital signal processing of noninvasive physiological variables. In: Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE. IEEE; 2006. p. 1355–8.

[13] Pennebaker JW. Putting stress into words: Health, linguistic, and therapeutic implications. Behaviour Res Therapy 1993;31(6):539–48.

[14] Vizer LM, Zhou L, Sears A. Automated stress detection using keystroke and linguistic features: An exploratory study. Int J Human-Computer Studies 2009;67(10):870–86.

[15] Christakis NA, Fowler JH. Social network sensors for early detection of contagious outbreaks. PLOS ONE 2010;5: . https://doi.org/10.1371/journal.pone.0012948e12948.

[16] Correia RB, Li L, Rocha LM. Monitoring potential drug interactions and reactions via network analysis of instagram user timelines. Pacific Symposium on Biocomputing, vol. 21. p. 492–503.

[17] Wood IB, Varela PL, Bollen J, Rocha LM, Gonçalves-Sá J. Human sexual cycles are driven by culture and match collective moods. Sci Rep 2017;7(1):1–11.

[18] Choudhury MD, Counts S, Horvitz E. Social media as a measurement tool of depression in populations. In: Proc. 5th Annual ACM Web Science Conf., WebSci'13. ACM; 2013. p. 47–56.

[19] Bathina KC, ten Thij M, Lorenzo-Luaces L, Rutter LA, Bollen J. Individuals with depression express more distorted thinking on social media. Nature Human Behaviour 2021;5:458–66. https://doi.org/10.1038/s41562-021-01050-7.

[20] Pang B, Lee L. Opinion mining and sentiment analysis, Foundations and Trends. Inform Retrieval 2008;2(1–2):1–135.

[21] Liu B. Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers; 2012.

[22] Bradley MM, Lang PJ, Affective norms for english words (anew): Instruction manual and affective ratings, Tech. Rep. C-1, The Center for Research in Psychophysiology, University of Florida., Gainesville, FL; 1999. .

[23] Hutto C, Gilbert E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: Eighth International Conference on Weblogs and Social Media (ICWSM-14), Ann Arbor MI. p. 216–25.

[24] Tausczik YR, Pennebaker JW. The psychological meaning of words: Liwc and computerized text analysis methods. J Language Social Psychol 2010;29(1):24–54.

[25] Ribeiro FN, Araújo M, Gonçalves P, Gonçalves MA, Benevenuto F. Sentibench - a benchmark comparison of state-of-the-practice sentiment analysis methods. EPJ Data Sci 2016;5(1):1–23.

[26] Bird EL, Steven, Klein E. Natural Language Processing with Python. O'Reilly Media Inc.; 2009.

[27] Pennebaker J, Boyd R, Jordan K, Blackburn K. The development and psychometric properties of liwc2015 Tech. rep. Austin, TX: University of Texas at Austin; 2015.

[28] Harrell Jr FE. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. Springer; 2015.

[29] Chung C, Pennebaker J. The psychological functions of funcion words. In: Fiedler K, editor. Social Communication. New York: Psychology Press; 2007. p. 343–59.

[30] Acharya AB, Wroten M, Wernicke aphasia, in: StatPearls [Internet], StatPearls Publishing, Treasure Island, FL, 2020, available from: https://www.ncbi.nlm.nih.gov/books/NBK441951/. .

[31] Epilepsy Foundation of America, Types of language problems in epilepsy, Online, accessed on Aug 18, 2021 (2021). .

[32] Rocha LM, Börner K, Miller WR, myaura: personalized web service for epilepsy management, Available from https://hsrproject.nlm.nih.gov/view_hsrproj_record/20191123, accessed Nov 29 (2019). .