Content analysis of Twitter skin cancer awareness campaign “SunSmart”

Author: Margarite Schindler
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands

ABSTRACT:
Over the past years, social media communication has received a lot of attention with many organizations trying to take advantage of the new ways to understand people and their needs. Online generated content has proven to be extremely useful thanks to all the metadata provided and the few costs associated with such research. Many health organizations joined social media to promote their campaigns, interact with their customers and share their messages. Skin cancer campaign "SunSmart" is one of them. However, little is known about the opinions expressed by individuals during those campaigns. As skin cancer is the most common type of cancer in countries like Australia and New Zealand, it is crucial to understand people's beliefs in order to improve the effectiveness of those campaigns. This paper uses sentiment and content analysis with the goal of analyzing the tweets generated by people in regard to skin cancer. It uses the Health Belief Model as a framework when examining the tweets. Results indicate that the Health Belief Model is difficult to apply in individuals' online activity and few of the tweets fall under the five proposed categories by the model. It nonetheless gives some clear understanding of the informative nature of the tweets and the little personal opinion expressed online. The results are of use for businesses as they help them recognize the main topics generated and could potentially assist in optimizing the way campaigns share their messages. Given the small and specific dataset used in this paper, it is believed that further research should be conducted on broader terms and audiences to see if the results can be generalized across countries and campaigns.

Graduation Committee members:
1st Supervisor: Ms. Anna Priante
2nd Supervisor: Dr. Michel Ehrenhard

Keywords
Online behavior, expressed opinions, content analysis, user-generated content, Twitter, health industry, skin cancer campaign.

11th IBA Bachelor Thesis Conference, July 10th, 2018, Enschede, The Netherlands.

1. INTRODUCTION

The vast amount of online user-generated content created in the recent years has not gone unnoticed by organizations and individuals. The increased use of social media technology made possible the creation of active networks of individuals that share their interests and beliefs, which in turn gives researchers new information to analyze people. Microblogging is a form of social media technology that has become very popular among individuals by allowing them to share small amounts of content and media (Yazdanifard et al., 2011). These microblogging platforms have been used by institutions and health campaigns which are entering the media in an attempt to engage with their target groups and promote their message, but also affect health behaviors in mass populations. Health industries, as well as related businesses, patients and regulatory authorities are increasingly using online technologies to access real-time opinions and gain potentially important information. These actions result in a more knowledgeable understanding of people's needs. Social media has been employed to disseminate health information on different diseases and treatments, and past researchers have shown that the distribution of such information is effective through sources such as Twitter and YouTube (Vance, Howe & Dellavalle, 2009; Scanfeld, Scanfeld & Larson, 2010). The great potential of online campaigns lies in the ability to communicate well-defined and focused messages to the mass population for a long amount of time with very little cost per head.

Twitter has been increasingly used for health campaigns and research purposes (Finfgeld-Connett, 2015). It is used to track emerging diseases (Charles-Smith et al., 2015) and disseminate health-related messages (Lister et al., 2015). It is also used to understand the public's views, knowledge, attitudes, beliefs and behaviors (Hays & Daker-White, 2015). This is possible due to the fact that Twitter has accounts created by hundreds of millions of people. People that continuously publish their thoughts, experiences and product/service reviews. Unlike other social platforms where most of the content is private, Twitter's information is almost all public which provides researchers with real-time data on various issues.

Although Twitter's content has been analyzed to answer many health and business-related research questions, Moorhead et al. argues that health communication research is still in its early stages, and there are a lot of gaps that are yet to be filled (Moorhead et al., 2013). Prior research on online health campaigns has been done by Ji, Chun and Geller in 2013 by developing sentiment classification of Twitter messages to measure the Degree of Concern (DOC) of the Twitter users about different diseases. Massey et al. (2017) quantified HPV vaccine communication on Twitter and developed a methodology to improve the collection and analysis of Twitter data. Furthermore, Wakefield et al. (2010) wrote a review in which it was discussed what the outcomes of mass media campaigns are in the context of various health-risk behaviors such as cancer screening and prevention. Wong et al. (2016) used sentiment analysis to describe the temporal, geospatial and thematic patterns of public sentiment towards breast cancer in the U.S. in order to examine how sentiment was related to screening behavior. However, based on the gathered knowledge about previous research on online health campaigns, a research has not yet been done on exploring the opinions expressed by people in relation to skin cancer campaigns. Therefore, no analysis has been done that gives an insight into what the main opinions regarding skin cancer awareness and protection are, as observed on online platforms such as Twitter. As seen by the past similar research, an insight has been given on other cancer campaigns, but it has been mostly done with predefined measures, instead of having an explorative approach and identifying the opinions as the analysis is being conducted. Having an explorative analysis of skin cancer related opinions is crucial as the goal of health campaigns on Twitter is to spread awareness and promote preventive behavior. Therefore, defining the main topics can build a clear understanding of how effective the existing campaigns are. Such research would be highly beneficial for the business field as it will allow campaigns such as SunSmart and other businesses related to the prevention of skin cancer to understand what their target market thinks and potentially needs.

The overall goal of this thesis is to develop a better understanding of how people communicate and what they express on Twitter in relation to skin cancer campaigns. This will be done by answering the following research question:

What are the opinions, observed on Twitter, that people are expressing in regard to health campaigns about skin cancer and its prevention?

The way through which the research question will be answered is by using two methods to analyze tweets from the SunSmart campaign. First, the tweets will be analyzed to find their prevailing sentiment – positive, negative or neutral. Second, a content analysis will concentrate on finding the main topics driving the discussion.

This paper is structured as follows. Section 2 introduces the Health Belief Model as a theoretical framework which will be further adapted to what might be observed when analyzing the Twitter conversation. In section 3, the case study will be introduced, followed by the data processing, and the two types of the analysis used will also be introduced. Section 4 is divided into two main subparts: the results which were observed when using the sentiment analysis tool, followed by results gathered from the content analysis. Section 5 will present a discussion and interpretation of the results observed. Section 6 of the paper will present the limitations associated with this research and will explore the possibilities for future research that can be conducted. Lastly, the paper will conclude with contributions and conclusions.

2. THEORETICAL FRAMEWORK

In the following lines, the paper will introduce the Health Belief Model (HBM). The model was chosen as it presents a framework to understand people's health-related beliefs. It complements the Social Cognitive Theory (SCT) created by Bandura, which has previously been used to analyze health promotion (Bandura, 1998). The Health Belief Model bears in mind the key categories of SCT, namely the outcome expectations about the effects of different lifestyle habits and the way they contribute to health behavior and the beliefs of personal efficacy of whether engaging in a behavior will lead to the anticipated outcomes. It, however, further builds on the theory by also considering the beliefs people have about the health behavior. The Health Belief Model is a useful model to understand health beliefs and opinions of people during health campaigns. The categories that will be presented will serve as a guidance when searching for main topics in the analysis that will be conducted in later stages.

2.1 Health Belief Model

The Health Belief Model (Rosenstock, 1974) is a social-psychological model that tries to predict proactive health behavior (e.g. seeking out medical help, taking medications, engaging in safety behavior), by concentrating on people's attitudes and is extensively used in studies of illness prevention behavior. This model includes a set of concepts about the knowledge and beliefs people have about a certain disease. It believes that preventive health behavior can be understood as a function of six core structures (Figure 1): the perceived **severity of the disease** (belief about the seriousness of the condition or the consequences of it), the perceived **susceptibility** (beliefs about the risks of getting the condition) and the **perceived benefits** to be realized by engaging in particular safety behavior, as opposed to **the barriers of engaging** in this behavior. The last two constructs are also partially covered by the Social Cognitive Theory introduced previously. They include the **self-efficacy**, which is the confidence that the action can be performed. This can be linked to the personal efficacy which also explores the confidence the person has in achieving the desired results by engaging in a specific behavior. Lastly, the **cues to action**, are the factors that activate readiness to change. The HBM is one of the most used theories in health education and promotion (Champion & Skinner, 2008).

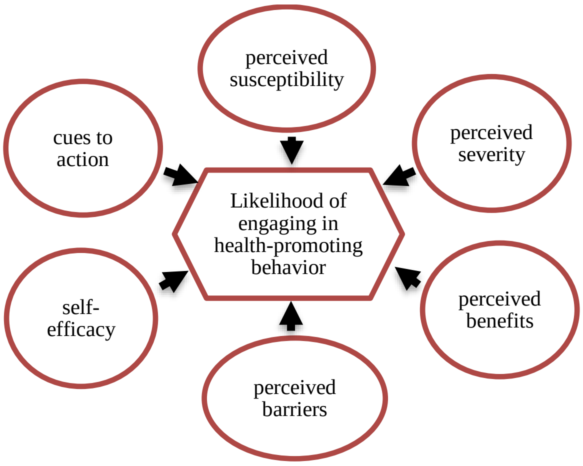

Figure 1. Health Belief Model

Janz & Becker (1984) argued that the most influential factor is the perceived barriers. Furthermore, the article written by Yarbrough and Braden (2001) used this model as their framework to assess the utility of HBM as a theoretical guide for predicting breast cancer screening. After assessing each of the six categories by analyzing previous literature that has used the model, Yarbrough and Braden concluded and further confirmed the statement that the most influential factor was the perceived barriers of engaging in the behavior. What this means is that the biggest explanatory factor for the opinions expressed towards the use of skin protection (in the virtue of this paper) is the potential barriers of using it. By barriers it is meant the perceived cost of engaging in safety behavior or the personal effort needed. Potential outcome of the analysis could show that there is aversion that people have towards using creams and other similar products because of the greasy feeling they leave on the skin or the need to reapply it multiple times a day.

The HBM has been previously studied in connection with skin cancer and harmful tanning intervention. Some of the constructs presented in the model have been linked to individuals' UV protection behavior. One of them is the perceived susceptibility. As people are aware that UV leads to skin cancer and photoaging, they have a motivating factor to use protection (Mahler et al., 1997). Unlike the fourth factor (the barriers), the first two (the perceived severity of the disease and the susceptibility of acquiring the disease) are often determined by the knowledge people have on the issue. That knowledge can be improved exactly by campaigns such as SunSmart which regularly post informative articles on skin cancer prevention and detection. The Health Belief Model first originated in the 1950s to understand why people do not comply with preventive procedures, but its core beliefs are still valid and used nowadays (Hayden, 2014). Despite their importance to skin cancer and tanning-related knowledge and behaviors, it remains unclear the extent to which HBM constructs are addressed in online mass media messages about skin cancer.

2.2 Conceptualization of the Health Belief Model in online skin cancer campaigns

The goal of this paper is to analyze the expressed opinions on Twitter in regard to skin cancer awareness and protection. As such, the HBM will be used as a guidance of what the main topics are. The model presents multiple categories that predict the behavior of people, which can be linked to their opinions online. Ahadzadeh et al. (2015) hypothesizes that individuals which have a higher perceived health risk also have greater motivation to adopt a health-promoting behavior and participate in online communication channels. The article defines perceived health risk as consisting of the different components proposed in the Health Belief Model, and the results prove the hypothesis to be true as there is a strong connection between the perceived risk and the use of internet for health-related purposes. Diddi and Lund (2017) further tested the connection by using content analysis to see how health organizations use their Twitter accounts. All the tweets of the chosen organizations were analyzed for the presence or absence of the theoretical parameters of the Health Belief Model (HBM). The results of the content analysis based on the HBM revealed that the content posted by these organizations reflected the use of varied theoretical constructs of the framework (Diddi & Lund, 2017). Thus, the previous research points that when speaking about skin cancer on Twitter, the components of the model are shaping the opinions of people and can be seen in their messages. In the next paragraphs, each of the categories of the model are introduced and applied to the context of this research.

As no previous research used the model as a framework for individuals' skin cancer opinions, the suggested application of the categories will be based on the Becker et al. (1977) paper where examples are proposed to predict dietary compliance. First, the perceived severity is the belief about the seriousness of skin cancer. Tweets were considered related to the perceived severity when people were writing about survival rates in the different stages of the sickness; about the first signs, the consequential symptoms and how they can be recognized; about the negative impact sun has on the body and the possible ways to minimize the risk. These topics were chosen based on Backer et al. which considered severity topics to be about survival rates and the different symptoms of the sickness. Second, the perceived susceptibility is the belief about how risky it is to get the condition. Messages fell under this category when individuals were talking about how often people get sick from skin cancer (also proposed by Becker et al), possibly giving statistics or when talking about which people are under the most risk of getting sick. Third, the perceived benefits are the person's opinion about the value and usefulness of engaging in the safety behavior as a way to decrease the risk of getting skin cancer. The paper suggested the main opinions to be about the different kinds of protection and about the useful information provided by different parties. Therefore, in this research tweets falling under this category talk about what the benefits of skin protection are and to what extent it can decrease the chances of being diagnosed. They might also talk about the different sun protection alternatives (hats, sunscreen, long sleeves etc.) and their ease of use. More generally, people broadcasting their own sun protective behavior fell under this category. Fourth, the perceived barriers are the individual's own evaluation of the obstacles associated with the adoption of the health behavior. Tweets were considered related if people were stressing on the inconvenience of the need to often reapply SPF protection as this could be difficult under some circumstances. They were also related when individuals were stressing on the positive effects the sun has and/or on the distress coming from wearing long sleeves, a hat etc. when being exposed to high temperatures. This was once again inspired by the Becker et al. research on dietary compliance which suggests barriers' concerns to be about the possible difficulties associated with engaging in healthy behavior.

The fifth and last factor considered in this analysis is the self-efficacy. That is one's own belief of their ability to do something and the outcome of this effort. For self-efficacy, the tweets considered to be part of this category were about how big the risk of skin cancer is even when implementing all the safety behavior in one's daily regime. It also included expressed concerns that cancer is unpredictable and there is no way to guarantee full protection from it as no one is immune. In terms of the expressed opinion people have towards the use of skin protection, the efficacy can explain why people might not protect themselves even though they are aware of the consequences UV has on their health. If a person believes that using skin care protection regularly might still lead to skin cancer diagnosis in the future, their motivation to take on that behavior drops. Table 1 presents examples of how the five components of the model will be used when analyzing the content of the tweets.

Table 1. Conceptualization of Health Belief Model

Category	Definition	Examples
Severity	Belief about the seriousness and consequences of skin cancer	"Symptoms you should look for...", "ways to get a skin check and recognize skin cancer are...", "Some of the effects Sun has on people..."
Susceptibility	Beliefs about the risk of getting cancer	"High risk behaviors include…", "Pale people have a higher chance of getting sunburnt".
Benefits	Beliefs about the benefits coming from engaging in sun protective behavior	"Why you should use protection…?", "Sun protection is readily available for everyone", "I love being Sun Smart… (may or may not give a further explanation why)".
Barriers	Potential inconveniences or risks from being sun protective.	"Sun has some health benefits such as (e.g. Vitamin D)", "No one around me is being Sun Smart", "The feeling of cream on the skin is very unpleasant", "Why is sunscreen so expensive?"
Efficiency	Belief about ease to engage in behavior and probability of achieving desired results.	"Why should I be using protection if so many people get sick regardless of their behavior"

The sixth category, cues to action, presents the readiness to change. It is said to be stimulus, that could be external or internal that motivates a person to engage in a behavior. Unlike the other five factors that are defined by a person's own belief or opinion on the matter, this sixth category is said to be a "factor". It is a psychological process that is often left uncommunicated. The first five categories show different beliefs that people hold and might express, whereas cues to action are triggers, or events that prompted engagement in health-promoting behaviors Therefore, cues to action does not represent any opinion on skin cancer prevention. Given this category does not give insight on the expressed opinions, it was excluded from the analysis.

3. METHODS

3.1 Case

Skin cancer is by far the most common form of cancer (Skin cancer, n.d.). Mass media campaigns have aimed at the prevention of skin cancer by concentrating on reducing sun exposure. Sun exposure is the main risk factor for skin cancer, making it crucial to understand what people's sentiment towards the use of protection is, in order for unhealthy behavior to be minimized. In countries like the United States of America, Australia and the United Kingdom, campaigns such as "Slip, Slap, Slop" and later on "SunSmart" were created to foster conversations about the topic and build awareness. In this report, the campaign "SunSmart" will be investigated. It is an annual health campaign found on various social media channels. The main focus is on raising awareness of the different types of skin cancer and promoting safe behaviors through the use of skin protection and clothing. SunSmart is a global campaign but each of its campaigns is localized through the creation of multiple social media accounts depending on the location. The research is concentrated on Australia and New Zealand (NZ). The original focus was solely on Australia as the SunSmart campaign is active there and has social media profiles for each of their six states – Western Australia, Northern Territory, South Australia, Queensland, New South Wales and Victoria. Another reason for this preference was due to the fact that approximately two in three Australians are diagnosed with skin cancer by the time they turn 70, and skin cancer accounts for about 80% of all diagnosed cancers (Cancer Council Australia, 2018). Nonetheless, after additional investigation, New Zealand was also included in the analysis as melanoma is a significant problem there due to high levels of UV exposure caused by the region's proximity to the ozone hole over the Antarctic. Australia and New Zealand both have one of the highest rates of skin cancer in the world, according to the largest international melanoma foundation, the Texas-based AIM at Melanoma. Both countries have more than double the incidence rates found in North America. This leads to the belief that there is a big conversation on this topic going on in these two countries.

A study from Australia has provided convincing evidence of improvements in attitudes and behavior relating to sun protection in the presence of variable amounts of media campaign exposure (Dobbinson et al., 2008). Reductions in the incidence of melanoma have also been observed over the decades of the SunSmart campaign (Hill & Marks, 2008). These past studies have shown the effectiveness of media campaigns, further stressing the need and opportunity to identify people's main conversations regarding skin protection in order to understand what their opinions are and whether the campaigns foster healthy behaviors.

The social media used in this research for collecting data is Twitter. A factor that distinguishes Twitter from the other social networks and makes it great for analyzing behavior is the age of its users. Usually, younger people are the ones that drive the growth of social networks (Miller, 2009). However, with Twitter it was observed that people from all ages use the network. The fact that it also attracts adults could be explained by the information-based nature of the system. This makes it different from the other networks which are mostly used only for connecting to acquaintances. Having people from all ages communicating on Twitter is a big advantage when doing research because it gives a representation of a wider subset of the population and does not exclude whole age groups from the analysis. Tweets also contain a wealth of data; analyzing and mining this data can provide insights into public opinion and behavior.

Though Twitter is beneficial for analysis due to its public nature and the users wide range of demographics, a drawback might be that Twitter is not in the top three most popular social media platforms of Australia and NZ. This can result in problems finding enough data to make any concrete conclusions. However, an attempt to prevent this was by collecting data for a large time period, therefore allowing for a wide timespan for tweets to be generated. The tweets extracted were only the ones that contained the phrase SunSmart, either as part of the sentence or used as a hashtag (#SunSmart). This decision is justified by the fact that the research is concentrated on the content generated within this campaign and using only this keyword will limit the content analysis only to tweets generated strictly in relation to the SunSmart topic.

3.2 Data collection and processing

The tweets with the user's Twitter activity and profile description information were originally obtained from the Twitter #datagrant project on large online cancer awareness campaigns. A database of approximately 11,700 English tweets sent by 3,147 different users was extracted for the SunSmart campaign. The tweets have been gathered for the period between the 30th of April 2014 and the 1st of February 2015 and, as previously mentioned, all included the phrase SunSmart in them either as a hashtag (#) or as part of the sentence. This time frame also included May, which is the international skin cancer awareness month (Skin Cancer Foundation, n.d.). Although data published on Twitter is publicly available, a confirmation from the university ethics committee was further obtained that the research has met the ethical standards and the data can be used without requiring the consent of each of the users whose data was captured.

The original dataset was not limited to a specific region. To ensure that the tweets are within the target group of the research, the first step was to take out all the data coming from geographical areas not located in Australia or New Zealand. This was done via Excel by filtering and extracting from the sheet all the unnecessary data. For Australia, cities such as Sidney, Melbourne, Perth, Brisbane, Adelaide, Canberra, Hobart and New Castle were considered, as well as phrases like Oz, Aussie, Queensland, Victoria and Western Australia (WA). This was due to the fact that through a brief search of the data, it was found that many of the profiles put such phrases in their location section and it was important to not exclude any of the country's residents. Furthermore, the same step was undertaken for New Zealand where phrases such as "kiwi" (a common word used as a self-reference by people from NZ) as their location or place of birth were also included. In the cases where a person has put more than one location, one of which being an Australian/New Zealand location, those tweets were being added to the data as the focus of this study is everyone that lives or has lived there and has suffered the consequences of the UV levels in this climate. The final count of tweets after this step of the cleaning process was 1,570.

As the research question of this study addressed, the individuals are the target of this research. For this purpose, all of the tweets that were not sent by individual accounts had to be filtered out. To separate the data that will be further analyzed from the organizational data, a manual approach was used. The tweets were classified into two main categories –individual and organizational. The sample of 1,570 tweets each has an ID number specific to the user. Based on that, the bio of the user was derived and conclusions on whether it is an organization or a person were drawn. The biography of each Twitter ID was used to categorize the tweets using multiple guidelines.

The first category is tweets of individuals. They are the ones sent by people and are the unit of analysis here. In those tweets, key topics and opinions will later be identified by using sentiment and content analysis. Individuals' accounts were characterized by the use of descriptions about who they are and what their personality is (e.g. "mom", "painter", "photographer", "marketer", etc.), what their interests are (e.g. "love bands Muse and Metallica", "love jogging", "vegan eater") and often used inspirational quotes (e.g. "Life is a school of learning, so live, dance, play, and enjoy ☺").

The second category is organizational tweets and it has two subcategories – informative and promotional. Informative tweets are defined as tweets that are neutral in their nature and their message consists of strictly non-personal information such as UV levels or dangerous hours to be outside without protection. Promotional tweets are again neutral but promote an event or a product. For the purpose of this article, no difference was made between the two when categorizing because no further analysis will be done on this category. Promotional and Informative tweets were both classified as organizational tweets and were identified based on their biography section which was concentrated on describing the nature of the business and often used their mission statements (e.g. "SAHMRI's vision is to transform research into health", "We are Queensland's leading independent organization in cancer control. We are to support all Queenslanders affected by cancer"). Where no bio was available, the content of the tweet itself was used to draw conclusions. This decision was preferred over the username as a way to classify the tweets because the data is from 2015 which means that some of the accounts do not exist anymore. Out of the whole set, about 50 tweets did not have a bio. Tweets that used the word "I" in a sense of experience or expressed opinion were often categorized into the personal section, whereas the tweets promoting an event or a product were put into the organizational section. When there was uncertainty about the sender of the message, the tweet was automatically excluded from the dataset, so the final set consists only of individual tweets. Once the classification was done, only 567 out of the 1,570 tweets fell under the individual tweets category. This makes up for 36.1% of the country-specific data set. A flowchart of the different steps undertaken to obtain the final dataset is presented in Figure 2.

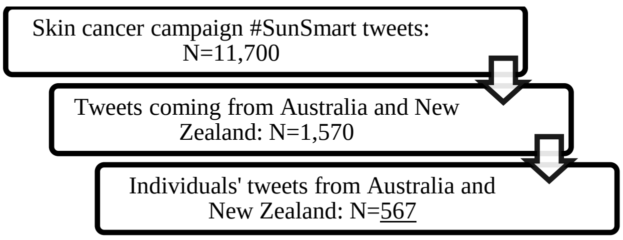

Figure 2. Flowchart of data processing

The final phase of the data processing was making the data readable for the sentiment and content classifiers. Tweets, when extracted, originally transform symbols into special characters (e.g. \x92; \n), which content tools cannot deal with. Sentiment classifiers are also unable to read some symbols such as "@". Therefore, it was necessary to exchange such symbols to readable text – "@" to "at", "\x92" to ";", etc. The final cleaned and processed document was run by the VADER classifier, followed by manual annotation and LIWC analysis.

3.3 Analytical approach

To answer our research question "What are the opinions, observed on Twitter, that people are expressing in regard to health campaigns about skin cancer and its prevention?", two methods were used: sentiment analysis to identify the prevailing polarity of opinion shown online and if it is positive or negative and content analysis to further understand the content of the tweets and identify the opinions expressed.

3.3.1 Sentiment analysis

Sentiment can be defined as an opinion that is held or expressed (Cambridge Dictionary, 2018). Sentiment analysis is a very useful tool for researchers, marketers, sociologists and psychologists as it can resolve many research problems by organizing qualitative data (Hutto & Gilbert, 2014). It is being used by businesses and also by public health organizations (Liu, 2012; Ji et al., 2015). The objective of this analysis is to translate opinions and expressions into meaningful and easily readable data that can further be quantified in order to determine attitudes towards various topics, services, products and events. It is part of the natural language processing that translates the subjective information from a post into different sentiment scores or classifications (for example: positive, negative or neutral) (Hutto & Gilbert, 2014). Thus, the sentiment analysis has turned out to be a valuable tool for businesses when evaluating opinions online. The sentiment analysis that will be used is descriptive analysis (visualizing sentiment orientation (polarity-based) of skin protection tweets).

Some of the existing classifiers are the SentiWordNet (SWN), General Inquirer (GI) and SenticNet (SCN). However, many of those classifiers were not created specifically for microblogging platforms. Hutto and Gilbert (2014) developed the classifier that will be used which is tailored for microblogging platforms, such as Twitter, called Valence Aware Dictionary for sEntiment Reasoning (VADER). The way this classifier works is by examining the polarity and intensity of each word. The results from this analysis were separated into three categories – positive, negative and neutral sentiment. In the sense of this research, a positive result would imply a positive attitude towards regular application of sunscreen products and other skin cancer preventing measures. A neutral would imply no opinion towards the topic, and a negative score would give us the people that are resistant to use such protection. The analysis gives scores for each of the three categories, as well as a fourth category called "compound" that calculates an average of all of the scores. For the purpose of this paper, the compound result was taken into consideration with positive sentiment being classified with a compound ≥ 0.5, neutral with a compound between -0.5 and 0.5, and negative sentiment with a compound ≤ -0.5 (Figure 3). The VADER lexicon is believed to perform exceptionally well when analyzing social media content such as tweets. VADER's correlation shows that it outperforms human raters (F1=0.96) at correctly classifying tweets into positive, neutral and negative classes (Hutto & Gilbert, 2014).

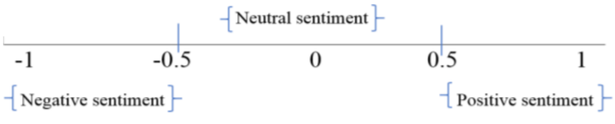

Figure 3. Sentiment scale (document level)

3.3.2 Content analysis

Once the sentiment analysis was finished, the data was analyzed to a greater extent to further identify the topics of the tweets. The investigation of the topics was done through content analysis. The content analysis was first conducted through manual annotation by using the Health Belief Model as a guideline. To apply the theory into practice, each of the concepts was supported by examples, which can be seen in Table 1 that was presented when the theory was conceptualized (section 2.2). This was created to guide the coder through the process and help identify which tweets belong to which category. The HBM has not been used before to explore individuals' tweets and therefore some crucial adjustments had to be made to fit the context. Given the nature of the model and the fact that it has not been previously used for Twitter skin cancer communities, it was expected that the conversations observed will not perfectly fit the categories and some of the opinions expressed might not at all be related to the five concepts. That is why it was important to make some adjustment to the model by loosening the definitions of the categories, so they can work in an online environment.

Furthermore, an automatic tool was used next to the manual coding, so results can then be drawn based on more than one analysis. This tool is used to validate the sentiment analysis and the manual annotation and allows for more insight as different categories are being presented with it. In order to provide an effective method for studying the various emotional, cognitive and structural components present in individuals' written speech samples, a text analysis application called Linguistic Inquiry and Word Count, or LIWC (pronounced Luke) was developed (Pennebaker et al., 2015). The LIWC software works by counting the frequency of the words and separating them into four main processes: linguistic processes (e.g. articles, prepositions), psychological processes (e.g., emotional, cognitive, perceptual), words denoting relativity (e.g., time, space) and personal concerns (e.g., religion, work). A previous research project that employed LIWC to analyze suicide-related tweets showed clear evidence that this content analysis gives an insight on people's behaviors, attitudes and their emotions (Dea et al., 2016). For example, a "greater use of first-person singular pronouns has been shown to suggest feelings of depression" while "second-person or plural pronouns indicate reaching out to others and a sense of community or group identity".

From the linguistic processes, the categories that were of importance were the use of first vs. second person. Examples of first person is the use of words such as "I", "me", "mine", whereas second person means frequently using the words "you" "you're". Thus, this category was used to see if messages created by individuals have the goal to share their own experiences and feelings or if they are trying to reach others and engage in a conversation. The adverbs score was also considered because low score in adverbs is normally seen in scientific papers and informative texts. Such results can show if the content consists mostly of opinions or of facts. This links back to the research question by showing if the content created mainly consists of people's opinions or if the nature of the tweets is strictly informative. Furthermore, some psychological process categories were also evaluated. Positive and negative emotion, anger, sadness and feel were the categories of interest. Positive and negative emotion were used to further confirm the results previously seen by VADER. Anger (e.g. hate, annoyed), sadness (e.g. crying, grief, sad) and feel were added to evaluate the prevailing emotion in the messages. This will show in the results if the opinions expressed show anger and frustration. Those categories can further confirm the manual annotation using the HBM as anger and frustration are feelings associated with negativity and therefore can be linked to the categories 'barriers' and 'severity'. The positive emotion on the other hand is linked to the 'benefits' category. Finally, from the present concerns, health was extracted to assess to what extent the data used to assess skin cancer conversation is actually connected to concerns about health. The LIWC model on its own will not give definite results but used with the manual annotation gives clear representation of key opinions by strengthening the reliability of the results previously presented.

4. RESULTS

The presented results will be supported by examples from the data. It must be noted that when providing examples, some tweets were slightly adjusted from the originals, while keeping the actual content untouched so the readability is easier. Therefore, multiple unneeded punctuation symbols and other irrelevant symbols were extracted.

4.1 Sentiment analysis

The output of the VADER analysis showed the following results: out of the 567 tweets that were analyzed, only 28 (4.9%) of them scored negatively on the analysis with only 1 (0.17%) tweet scoring highly negatively (compound = -0.9). On the other hand, the positive sentiment was evaluated with a total of 117 (20.6%) tweets. This left the other 422 (74.5%) tweets with a neutral result. This striking neutrality of the tweets was further confirmed by having 206 (36.3%) of the tweets absolutely neutral with a score of 0.0.

Further manual analysis was done to test the reliability of the classifier. As the negative sentiment sample was small with n=28, it allowed to check each of the tweets and evaluate what made them score negatively. What was found was a flaw of the sentiment tool. An example of a negative tweet seen in the data is "@jojokempson: This is bullshit. I'm being punished for being sunsmart? Will attempt to dull the pain with champers". As it can be seen, although this results in a negative sentiment, it does not represent a negative opinion towards the use of skin protection. Other examples include "When your head blisters because you forgot to take your hat to warped tour stay sunsmart guys. this shit hurts" and "Skin cancers caught early kill more! Prevention, Prevention, Prevention #SunSmart." Some tweets also included words such as "neglect" and "hate" that explain the result, however those words were used negatively towards the sickness and neutrally/positively towards the prevention of it. These results showed that out of the 28 negatively classified posts, none of them were directly against the campaign or its values and advices to put on sun protection. Although the tweets did indeed possess negative sentiment, this was due to parts of sentences and specific words that are associated with unpleasant experiences.

After observing that the VADER results classified some of the tweets into the negative category when in fact they didn't express a negative opinion, a sample of the positive and neutral tweets was extracted to further examine their polarity. For the positively scored posts, a small sample was extracted by using only "highly" positive scores such as tweets that scored 0.8 or 0.9 on the classifier. Examples of the results are "Happy Friday and don't forget to check the UV as you enjoy this beautiful weekend! #VIC #SunSmart" and "Great to hear Peppa Pig is encouraging kids to be sunsmart @SusannaFreymark". The results align with the original expectations of what positive sentiment towards the use of sun protection would look like. As it can be seen from the examples, the positive tweets communicate messages about using protection and being cautious about the sun levels and possible damage. Finally, the neutral tweets showed messages such as "Dysplastic moles look different to ordinary moles and may evolve to melanoma..." and "Today's predicted max UV is 4. Sun protection is required from 10:10 am to 1:50 pm - Get SunSmart for iPhone". What can be clearly observed is the informative nature of those Tweets. They scored neutral as the tweets strictly provide information and guidance on the matter and do not involve any personal feelings about skin cancer.

Manually checking the sentiment classifier output gave more clarity of the situation. The negative results come from the words used in the sentences, whereas positive results further confirm the positive attitude expressed by people towards the use of protection. The immense size of neutral tweets was explained by the informative nature of most of the messages communicated on Twitter which could have been expected given the informative nature of the social media itself. Overall, these results showed that little to no negative opinion is seen on Twitter skin cancer campaigns with the majority of the posts giving mostly factual information that doesn't possess personal opinions.

4.2 Content analysis

By using the HBM, 134 (23.6%) tweets fell under one of the five categories presented in the model. While some tweets could be classified in more than one of the categories, the most dominant theme was the one chosen and coded for the final analysis.

The 'benefits' category, defined by the benefits a person believes come from engaging in sun protective behavior, had the biggest number of tweets with 74 (13.05% out of the 567 posts) of them falling under this category. People tweeting about their own Sun Smart behavior without giving further insights on how they were sun protective fell under this category as providing further explanation was rarely seen (e.g. "Stay sunsmart folks"). In the times when there was more information provided, few key opinions were expressed. People showed positive attitude towards the shadow-friendly nature of Australia. Information about Australia putting 10 million dollars on improving the conditions outside and giving more places with shade was expressed as an argument that being protective becomes increasingly easier, and there is no reason not to be. Other perks communicated were the various jobs offering free sunscreen, all the events that give away free hats ("Thanks #sunsmartWA for the awesome sombreros! Saved my skin") and an app for the phone that has been created with the specific purpose of making the use of protection easy and effortless. This app shows everything that needs to be known about UV levels and protection. All those factors were considered benefits as they make sun protection behavior easy and accessible.

The category that had the second highest number of tweets was the barriers. Barriers was previously defined as the belief of potential inconveniences and risks associated with engaging in preventive behavior. 28 Tweets or 4.9% fell under this category. Tweets were categorized as barriers not only in the sense of lack of desirability to engage in the behavior, but also when they were identifying the unprotective behavior that people from the surroundings have. Examples of messages that fell under this category are people admitting to not always being sun protective, not because of lack of motivation, but because it is not embedded in their brain that they need to always bring their hat and sunscreen with them when they leave the house. Other barriers that were expressed by people were the costs coming from continuously buying sunscreen "Really wish that @TheWarehouse can hurry up and do a buy 1 get 1 free sale on sunscreens!! Going get burnt #soexpensive".

The category that had the third highest number of tweets, with 27 tweets or 4.7%, was 'severity'. It is defined by the belief about the seriousness and consequences of skin cancer. It presented opinions about the various symptoms that individuals should look out for and gave different website articles about skin checks and how to recognize dangerous beauty marks (e.g. "Are you getting a good skin cancer check? Know what to look for during an exam: *link to an article*"). It also gave insights on how to be sun protective and the negative effects sun has on people.

Furthermore, the susceptibility resulted as fourth in terms of frequency. Susceptibility is the belief about the risk of getting the condition. As this category is closely related to severity, it was often difficult to separate the message content. This resulted in only 5 tweets falling under the category of susceptibility. These five tweets were coming from people that are under high risk (e.g. individuals with pale skin) explaining their issues associated with their high risk. The other messages observed were information on high risk behaviors such as “sun bathing/bedding, prolonged exposure, sitting behind glass all day, high veg oil intake”.

Finally, efficiency gave 0 results. No tweets talking about the risk of being diagnosed with skin cancer even when being sun protective were identified. To further understand the results, a connection between all the existing categories from the content analysis and the output from the sentiment analysis is illustrated in Table 2.

Table 2. Content and Sentiment results

Content categories	Sentiment count
barriers	**28**
Negative	6
Neutral	18
Positive	4
benefits	**74**
Negative	3
Neutral	43
Positive	28
severity	**27**
Negative	3
Neutral	21
Positive	3
susceptibility	**5**
Negative	1
Neutral	3
Positive	1
Grand Total	**134**

The benefits category is linked to the highest amount of positive sentiment. About 38% of the tweets in this category gave a positive outcome. This outcome was expected as benefits was the category that presented positive attitude towards the use of sun protection and also included the people that expressed online how important it is to be sun protective and that they themselves are. Even with only 6 negative tweets, the barriers category presented the biggest number of negative sentiment. This aligns with the assumption that barriers would show negativity, because by definition this category talks about the monetary and personal inconveniences coming from engaging in a behavior. It must however be noted that the VADER classifier, once manually checked, showed some inaccuracy in predicting negative sentiment, so the negative results presented in this table only show posts using words associated with negativity but do not necessarily show negative opinion towards the use of protection.

4.3 Validation of results using LIWC analysis

The categories that were analyzed by LIWC were used to further confirm the results seen by the sentiment analysis and the content manual annotation. The output of the LIWC analysis provides a mean number for each of the categories. The number for the variables is expressed as a percentage of the total words in the file. The numbers seen for each of the categories will then be judged on the basis of the means provided for each type of text in the LIWC2007 Language Manual (e.g. emotional writing, blogs and scientific writing) (Pennebaker, Booth and Francis, 2011). For the linguistic processes, the first person “I” resulted in 1.14, compared to “We” with 0.47 and “You” with 1.57. The low means of “We” and “You” for social media shows that the individuals do not see themselves as part of the group and try to start conversations. This can also be seen when analyzing the content by the fact that there is little conversation going on between the people, and most of the tweets are not intended to create a conversation. Adverbs scored low compared to the means presented in the LIWC manual with a score of 2.09. For a comparison, the means for emotional writing, blogs and talking are all higher than 5 as expressing emotions results high on this category. 2.09 is a mean comparable to the one of scientific articles (1.35) which shows that most of the content of the tweets is factual and informative. The positive and negative emotion categories were used to further confirm the VADER results. The psychological processes results lined up with the results observed from the sentiment analysis. The positive emotion category scored 3.37, compared to the negative which scored 0.76. A result of 3.37 aligns with the mean usually seen in blogs (3.72), whereas a negative outcome of 0.76 is a low mean for any type of text that has been previously analyzed by LIWC. This shows that the overall positive sentiment is bigger than the negative sentiment and that negative sentiment can rarely be seen in the analyzed case. Furthermore, anxiety scored lower than the mean usually seen in emotional writing but higher than the mean observed in blogs and scientific writings. Anger and sadness, however, both showed low results. The final psychological category, feel, scored relatively high, showing that sentiment might be mostly positive to neutral, but there is still anxiety and different types of feelings expressed by people in their tweets. The final category is the present concerns. Health scored exceptionally high (1.4), whereas the means of any type of text evaluated by the creators of LIWC showed a mean below 1. Nonetheless, this result was no surprise given the tweets analyzed come from a health campaign. It is, however, a proof that the dataset used in the present study is a reliable source of information.

5. DISCUSSION

The findings provided the opportunity for some conclusions to be drawn in order to answer the research question: “What are the opinions, observed on Twitter, that people are expressing in regard to health campaigns about skin cancer and its prevention?”. It was observed that in online skin cancer health campaigns, people share little about their own feelings and opinions. The little involvement by the people was already seen in the methodology section – filtering only the individuals’ tweets from the data resulted in a decrease from 1,570 tweets to 567. This showed that only 36% of the conversation on Twitter comes from individual people. The first step of the analysis was to evaluate the sentiment. The results showed that a great amount of tweets were neutral and of a very informative nature. Moreover, the final set of 567 posts was further decreased by number when the tweets were used for the content analysis and were trying to be categorized based on a model that was created to show individuals’ health beliefs. When undertaking the manual annotation, only 23% of the 567 tweets were able to fit into the Health Belief Model categories. This revealed that the HBM is difficult to apply in this setting. However, when answering the research question based on our categorized tweets, the results showed an overall positive attitude and expressed opinions towards the use of protection. Based on the categories presented by the Health Belief Model, the majority of expressed opinions is about the benefits coming from engaging in sun protective behavior. 55% (73 Tweets out of 134) of the categorized expressed opinions were in relation to the need to use sun protection, the ease to use it and all the perks coming from engaging in the behavior. However, the research allowed for some broader conclusions. The outcomes showed clear unwillingness by people to actively participate in the conversation and share their opinions. Organizations and campaigns fail to encourage individuals to share their experiences and opinions. In order for organizations to understand their customers and optimize their campaigns, they need to stimulate their target group to communicate their needs and concerns. Currently, it is difficult to see what's missing in the campaign's strategy as little people talk about it. The little amount of expressed opinions leads to a new question: what are the rest of the people talking about?

Given only 23.6% of the 567 tweets were able to fit into the model, a need for evaluating the other 76.4% of the tweets and the nature of their content was seen. The big gap observed between the model introduced by Rosenstock and the actual content seen on Twitter gave an opportunity for additional categories to be suggested so all of the data can be categorized. In the next section, supplementary categories will be suggested so further interpretation of the data can be done, and conclusions can be drawn.

5.1 Proposed categories

Three new categories were added- promotional, informative and advisory. Furthermore, a fourth category called "uncategorized" was also added when tweets were seen to not have any content or when they were not related to the topic of the research.

Analyzing manually the rest of the tweets lead to the finding that many of the uncategorized tweets were retweets. Retweeting is reposting information provided by another user and gives no new insight and instead reinforces something previously said. What was further observed was that most of the retweets were informative and promotional posts originally coming from campaigns. Further evidence of the influence organizations have on the individuals' content of tweets was seen by the considerable number of tweets posted by individuals but having their content about organizational events and information. It was seen that people tweet from their personal account about events on behalf of the organizations they work for or post about radio shows and talks they participate in. Furthermore, teachers and school employees also post on behalf of their workplace about sun protection rules. Example "A reminder that from Sept - April, all students are required to wear their SunSmart hats each day, so please send them along from Monday!". The fact that almost 40% of the individuals were retweeting organizations and campaigns showed the need to reintroduce the two main subcategories suggested in the data processing section for the organizational tweets: "**Promotional**" and "**Informative**". The promotional category was previously defined by tweets that advertise an event, a product or a service – "Need funding to build shaded areas? Apply for a SunSmart Grant today! Go to URL or call 13 11 20". The informative category only states facts "Check sun protection times as you head to the Royal Adelaide Show- be SunSmart from 10.20am-2.10pm (UV 4)" and often answers the question "How?" How to apply it, how to recognize the illness etc. Frequently those questions were referring to a link where more information was provided.

The third additional proposed category is called "**Advisory**". It was seen that a big percentage of the tweets are informative and give only facts about the conditions outside and symptoms of the illness. However, a more personal approach was also observed. The advisory category presents the personal advices that individuals gave to others to push them to be Sun Smart. Some examples were simply stating that one should always engage in protective behaviors ("So important for our babies and kids. Get in the SunSmart habit every day!") while others gave more specific examples of how to be Sun Smart – "Remember, caps only shade your nose and forehead, so a broad-brimmed hat is best". As it can be seen the biggest difference between informative and advisory is that there is a more personal touch to the advisory category.

Finally, some tweets were seen to contain little to no information. The messages included tagging other people without fostering a conversation. Another issue when interpreting the content was because of all the tweets that used the hashtag SunSmart without talking directly about the topic. For example, "Clarke, Doolan, Johnson and Siddle have all opted for the floppy white instead of the Baggy Green #bigissues #sunsmart" or "#sunsmart #flipflops #summer #sun #beach". Other examples of tweets that add no knowledge about the conversation and opinions observed are „@abijones_xx @bumpshow #sunsmart" and "Our cow looked the snazziest. @adelaideshow #sunsmart". Such content gave no insight on the topic and therefore wasn't beneficial when answering the research question. Because of the aforementioned reasons, an "**Uncategorized**" classification had to be added. This category is defined by tweets that do not consist of content related to the topic and do not give information about skin cancer opinions.

Introducing the additional categories allowed for all 567 to be categorized. The 433 tweets that were previously unable to fit the HBM were able to fit the additional categories, and the posts that did not fit any of the proposed criteria were put into Uncategorized. The final outcome for the previously unclassified 433 tweets gave 126 Uncategorized, 111 Promotional, 111 Informative and 85 Advisory tweets. Being able to put the data into categories (Figure 2) allowed for the results to be interpreted and conclusions about the nature of online skin cancer conversation to be drawn. Furthermore, a connection between the old and new categories with the sentiment analysis can be seen in Appendix 1. The prevailing sentiment for each category is neutral with the Informative category scoring highest on neutrality and lowest on positive and negative sentiment, once again showing the little opinions expressed in relation to this campaign and the stress on information-based tweets.

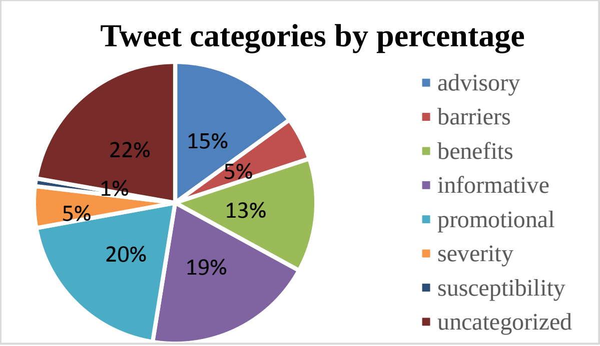

Figure 2. Tweet categories presented by percentage of the total

Overall, the categories from the Model did not fit the data. Some of the categories were very loosely fitting and a lot of adjustments from the original definitions the model gives had to be made. Even after the adjustments it can be seen from Figure 2 that the biggest categories do not come from the HBM. Uncategorized is the classification with highest number of posts.

The second biggest category goes to the strictly informative tweets that do not express any personal opinion. This makes it visible that a great number of the tweets are not containing opinions and beliefs. The little activity seen by individuals revealed no interest in discussing the sickness, but rather concentrated on promotional and informative messages. Results showed very different findings than what was originally expected. The tweets showed little individuality and different opinions. Unlike the original assumption that individuals' tweets will differ significantly from organizations' tweets – they did not. Most of the conversation is driven by what campaigns are saying and opinions that differ from the campaign's messages were rarely seen.

5.2 Social desirability

Based on previous research and on the Health Belief Model, as well as the Social Cognitive Theory that was previously introduced, it was believed that positive, as well as negative sentiment will be observed. Contrary to those initial assumptions, the opinions observed were mostly neutral to positive. An explanation of this occurrence might be due to the fact that only people who have positive attitude towards the use of sun protection engage in conversations about it, as people with negative opinions might not follow the topic or feel the need to share their negativity. As those results could be accurate and everyone posting on Twitter might be engaging in safety behavior, there are reasons to doubt online behavior is the same as offline behavior. One of the reasons those results should be taken with precaution is the social desirability concept. This theory explores potential reasons for observing only positive sentiment towards the use of skin protection. Social desirability is generally defined as the tendency for subjects to act in a manner that consistently presents the self in a favorable light (Edwards, 1957; Jackson, 1984; Wiggins, 1973). A problem facing many health organizations is the tendency of participants to over-report health promotion and disease-prevention activities and this have been observed in a wide range of health activities, including behaviors such as obtaining procedures for early detection for cancer (Warnecke et al., 1997). The accuracy of self-reported information on health behaviors has been brought into question as a result of research conducted by Newell et al (1999) that found that there is often a difference between what people report and what is found in their medical records.

A relevant issue for social researchers and social media marketers is to understand to what extent social desirability can play a role in people's attitudes showed on social media. In other words, do opinions given within a social network reflect the genuine emotions towards a topic? There is evidence from the social interaction literature that states virtual worlds contexts increase social desirability. This is supported by the fact that the interactions going on in online communities are publicly shared, which creates the opportunity for the communication to be seen by thousands of people, thus increasing the need to be socially accepted (Etzioni & Etzioni, 1999). The issue with social desirability extends to the business field. Marketers are provided with an exceptional opportunity to study behaviors and opinions online with almost no costs associated with this research. However, they also should not draw definite conclusions based only on those studies as this might lead them to coming up with a new product/service that was believed to be supported by the online communities only so it later becomes a failure due to the gaps that exist between online and offline behaviors.

Nevertheless, the study of opinions expressed online can shed light on the effectiveness of a campaign in terms of promoting online conversations on the topic. However, it cannot and should not be considered a representation of the actual behaviors and, therefore it cannot be linked to a possible reduction of skin cancer or increase of sunscreen sales.

6. LIMITATIONS AND FUTURE RESEARCH

As with all research, there are some limitations. First, the tweets that were analyzed in this study only examined English tweets and were constrained to the time period between April 2014 and February 2015. They were also limited to Australia and New Zealand leaving the rest of the world unaccounted for. This led to a very small and specific dataset. Tweets that were published in other languages (e.g. Italian, German, Spanish etc.) could have showed significantly different results than the ones in this dataset. The topics seen by people from Australia and New Zealand might have been affected by the way organizations in those locations structure their messages. As it was seen by the original dataset, a great amount of the tweets were reporting events organized by the cancer campaign. If a research is conducted in other regions, it could be seen that because of their differences in events and messages sent by the campaigns, different tweets are being generated in response by the people. While this paper provides an initial insight into the use of microblogging about skin cancer awareness, data collected across multiple points of time and locations would be beneficial to demonstrate a better representation of the main conversations generated. The study must also be extended to social media platforms like Facebook, Tumblr, Pinterest, etc. This would show the applicability and replicability of the results observed in this paper and display if the key topics are generalizable and if there is a baseline framework for the opinions observed.

The use only of the keyword SunSmart and the hashtag with the same name is a second limitation. This limitation is applicable to all studies that gather tweets on a single topic. Users might tweet about a topic without mentioning the keyword and/or hashtag used in the study, making it difficult to gather all the data on a certain topic. Hence, this may result in a sample that is not representable. To better understand skin cancer prevention organizations, a study should also be done using other hashtags. A greater volume of tweets would also allow for a more detailed analysis which could then unveil the existence of less frequently occurring topics.

An additional limitation is the absence of gender consideration from the analysis as it difficult to predict gender on Twitter. A further research can make a separation based on gender to see if there is any difference in sentiment by men compared to women The burden from skin cancer is particularly high in fair skinned men throughout the world. Furthermore, males have two times higher risk of melanoma compared to women (Janda et al., 2010). Examining the content based on gender might reveal significant differences in the content of the posts. If no significant difference is observed, this could give further reason to believe that social desirability plays a role in the conversations observed online. A research can then show to what extent this phenomenon is observed in online communities and what tactics can minimize the difference between online and offline expressed beliefs.

Furthermore, the current research did not look at the network as a whole. Relationships between the campaigns and their followers were not evaluated and thus, key insights of the campaigns strategy were left unexplored. It is recommended that more research is done in the future to analyze not only the receiver (people's) perspective, but also the sender's (organizations). A potential study could manually analyze also the messages posted by the campaigns. This could allow for a connection to be drawn between the two. An answer to the question "How do people react to the different types of messages sent by campaigns?" can be found. Such a research can give more clarity on why people post what they post and how organizations' message content and purposes might affect their behavior. This will also shed light on the potential ways organizations can improve their online activity to affect their target market. The ideal situation when more research has been done is to be able to provide marketers with a set of tools that they can use to better their strategy. This paper, unfortunately, had a limited time framework and coders, but with more participants and time, a lot can be achieved.

Finally, the sentiment analysis must be done with precaution as social media is being used for spamming, and the results from this analysis could therefore be inaccurate. This was attempted to be prevented by conducting a secondary analysis using a smaller sample which allowed to also see if the messages being analyzed are actually about the topic. However, the overall validity and reliability of the sentiment and content results was not evaluated and could have therefore led to some inaccurate conclusions.

7. CONTRIBUTIONS AND CONCLUSIONS

Social media is increasingly being used by marketers to convey their messages and promote their services. With the rise of this new media and the enormous amount of users that join it every day, it makes it a worthwhile marketing tool.

Since the rise of Twitter as a platform, academic research has been conducted analyzing the content of Twitter posts. Various companies have analyzed people's opinions in regard to their products and have made estimations thanks to all the metadata accessible to them. However, much less research has been done in the health-related area. This is, nonetheless, a crucial research topic as social media is constantly being used for health campaigns and it has been proven that those campaigns have an impact on society (George, Rovniak & Kraschnewski, 2013). Analyzing the expressed opinion by people gives a strong baseline and excellent opportunity for further research. This research was the first one to analyze the content of skin cancer campaigns. It gave initial results and interpretation of what is to be seen by people in skin cancer online communities. It helped facilitate the understanding of real time reactions of people about skin cancer related issues. The study offers important information for scholars and practitioners in understanding the use of the Health Belief model framework in social media platforms, in particular Twitter, in regard to promoting health communication about skin cancer. Understanding of the topic generated from this research is also of significant relevance to health organizations and related businesses. It gives them valuable insights into the missed opportunities of effective message creation that stimulates responses. Moreover, health organizations will obtain valuable summary of the opinions expressed. This will thus allow them to design more strategical messages and interventions in the future and use their social media networks in a more effective way in order to share their values successfully. Sunsmart would also benefit from the research findings by further adjusting and improving their value creation.

This thesis offers four main contributions. First, it contributes to research on health campaigns by addressing the gap in knowledge about the content generated about skin cancer campaigns. Second, it contributes to the Health Belief Model by visualizing the strength and pitfalls of the model when applying it to the skin cancer communication domain. Third, in addition to using the HBM to analyze the content, additional categories were suggested as the model did not cover all the topics identified online. Consequentially, it allowed for an elaborate framework resulting in contribution to health behavior theory by building on the existing model. These categories haven't been proposed by previous studies on similar topics and can be used by researchers in the future for content analysis of various health campaigns. Finally, to answer the research question, multi-method approach combining automated tools and manual approach was used. Concrete percentage of tweets per category were additionally presented to help in assessing which topics are seen the most in the campaign communication.

The opinions observed in this research showed positivity towards the use of protection and the campaign's goals. The individuals mostly expressed satisfaction with the availability of sun protection and all the new opportunities presented to help them be Sun Smart. However, some opinions also pointed out the problems with this health behavior. Cost of sunscreens and conditions outside were some of the factors that prevented them from being more protective. One of the biggest conclusions, however, was not related to any expressed opinions. It was related to the lack of them. Only a small amount of the data showed people expressing their beliefs. Most of the activity observed on Twitter came from organizations rather than individuals, which shows that organizations are actively present online, but fail to encourage individuals to share their concerns and remarks. Therefore, marketers are missing a big opportunity to engage with their customers and understand their needs.

The findings once again prove the huge potential that social media technologies like Twitter have for health-based conversations, but also for any type of business that is interested in what their customers' opinions are. Marketers of skincare organizations are given an insight of what the customers say about using protection, what their opinions are and what the possible reasons for not buying sun protection products are. This can give useful understanding of the customer's needs and allow the companies to build strategies based on this new information and find novel ways to approach their potential target group. Twitter gave the opportunity to analyze people, regardless of their age and gender, in an environment where everyone can share their opinion publicly and can reach limitless amount of people from around the world. For marketers, Twitter and other social media networks, when used adequately, can play a substantial role in effective information communication.

8. ACKNOWLEDGMENTS

First of all, I would like to express my gratitude to my first supervisor - Ms. Anna Priante. Her continuous feedback and support throughout this process as well as the useful insight enriched this thesis and made it a success. Furthermore, I would like to thank my family for their continuous support throughout my studies. Without them, I would have never been able to attain this degree.

9. REFERENCES

Ahadzadeh, A.S., De, M.P., Sharif, S.D., Ong, F.S., Khong, K.W. (2015). Integrating Health Belief Model and Technology Acceptance Model: an investigation of health-related internet use. *Journal of Medical Internet Research, Vol. 17, Issue 2*

Bandura, A. (1998). Health promotion from the perspective of social cognitive theory. *Psychology and Health, Vol.13, pp. 623-649.*

Becker, M.H., Maiman, L.A., Kirscht, J.P., Haefner, D.P and Drachman, H. (1977). The Health Belief Model and prediction of dietary compliance: A field experiment. *Journal of Health and Social Behaviour, Vol. 18, Issue 4, pp.348-366.*

Cancer Council Australia (2018). Skin cancer. Retrieved April 3rd, 2018 from https://www.cancer.org.au/about-cancer/types-of-cancer/skin-cancer.html

Champion, V. L., and Skinner, C. S. (2008). *The health belief model.*

Charles-Smith, L.E., Reynolds, T.L., Cameron, M.A., Conway, M., Lau, E.H.Y., Olsen, J.M., Pavlin, J.A., Shigematsu, M., Streichert, L.C., Suda, K.J. and Corley, C.D. (2015). Using social media for actionable disease surveillance and outbreak management: A systematic literature review. *Public Library of science, Vol. 10, Issue 10*

Dea, B.O., Larsen, M., Batterham, P., Calear, A. and Christensen, H. (2016). Talking about suicide on Twitter: Linguistic style and language processes of suicide-related posts. *European Psychiatry, Vol.33*

Diddi, P. and Lund, L.K. (2017). Organizational Twitter Use: Content Analysis of Tweets during Breast Cancer Awareness Month. *Journal of Health Communication, Vol. 2, Issue 3, pp. 243-253*

Dobbinson, S.J., Wakefield, M.A., Jamsen, K.M., Herd, N.L., Spittal, M.J., Lipscomb J.E. and Hill, D.J. (2008). Weekend sun protection and sunburn in Australia trends (1987-2002) and association with SunSmart television advertising. *American Journal of preventive medicine, Vol. 34, pp. 94–101.*

Edwards, A. E. (1957). *The social desirability variable in personality assessment and research.* New York: The Dryden Press.

Elson, S.B., Yeung, D., Roshan, P., Bohandy, S.R. and Nader, A. (2012). Using Social Media to Gauge Iranian Public Opinion and Mood After the 2009 Election. *Rand Corporation*

Etzioni, A., and Etzioni, O. (1999). Face-to-Face and Computer-Mediated Communities: A Comparative Analysis. *The Information Society, Vol. 15, pp. 241-248.*

Finfgeld-Connett D. (2015). Twitter and health science research. *Western Journal of Nursing Research, Vol. 37, Issue 10, pp. 1269 - 1283.*

George, D., Rovniak, L. and Kraschnewski, J. (2013). Dangers and opportunities for social media in medicine. *Clinical obstetrics and gynecology*

Glanz, K., Rimer, B. K. and Viswanath K. (Eds.), *Health behavior and health education: Theory, research, and practice, pp. 45-65*.

Hayden, J. (2014). *Introduction to health behavior theory.* Burlington, MA: Jones & Bartlett Learning.

Hays, R. and Daker-White, G. (2015). The care. data consensus? A qualitative analysis of opinions expressed on Twitter. *BMC Public Health*

Hill, D. and Marks, R. (2008). Health promotion programs for melanoma prevention: screw or spring? *Archives of dermatology, Vol. 144, pp. 538–540.*

Hutto, C. and Gilbert, E. (2014). VADER: a parsimonious rule-based model for sentiment analysis of social media text. *Association for the Advancement of Artificial Intelligence*

Jackson, D. N. (1984). Personality Research Form manual (3rd ed.). *Port Huron, MI: Research Psychologists Press*

Janda, M., Baade, P.D., Youl, P.H., Aitken, J.F., Whiteman, D.C., Gordon, L. Neale, R.E. (2010). The skin awareness study: Promoting thorough skin self-examination for skin cancer among men 50 years or older. Contemporary clinical trial, Vol. 31, Issue 1, pp. 119-130.

Janz, N.K., and Becker, M.H. (1984). The Health Belief Model: A decade later. *Health Education Quarterly, Vol. 11, Issue 1, pp. 1-47.*

Java, A., Song, X., Finin, T. and Tseng, B. (2007). Why we Twitter: Understanding microblogging usage and communities. Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis

Ji X., Chun, S.A., Geller, J. (2013). Monitoring public health concerns using twitter sentiment classifications. *Healthcare informatics*

Ji, X., Chun, S.A., Wei, Z. and Geller, J. (2015). Twitter sentiment classification for measuring public health concerns. *Social network analysis and mining*

Lister, C., Royne, M., Payne, H.E., Cannon, B., Hanson, C. and Barnes, M. (2015). The laugh model: Reframing and rebranding public health through social media. *Am. J. Public Health*

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*

Mahler, H. I. M., Fitzpatrick, B., Parker, P. and Lapin, A. (1997). The relative effects of a health-based versus an appearance-based intervention designed to increase sunscreen use. *American Journal of Health Promotion, Vol. 11, pp. 426–429.*

Massey, P.M., Boansi, R.K., Gipson, J.D., Adams, R.M, Riess, H., Dieng, T., Prelip, M.L., Glik, D.C. (2017). Human papillomavirus (HPV) awareness and vaccine receptivity among Senegalese adolescents. *Tropical medicine and international health, Vol.22, Issue 1, pp. 113-121*

Miller, C. C. (2009). Who's driving Twitter's popularity? Not teens. *The New York Times*

Moorhead, A., Hazlett, D.E., Harrison, L., Carroll, J.K., Irwin, A., and Hoving, C. (2013). A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. *Journal of medical Internet research, Vol.15, Issue 4.*

Newell, S.A., Girgis, A., Sanson-Fisher, R.W., Savolainen N.J. (1999). The accuracy of self-reported health behaviors and risk factors relating to cancer and cardiovascular disease in the general population: a critical review. *American Journal of Preventive Medicine, Vol.* 17, Issue 3, pp. 211-29.

Pennebaker, J.W., Booth, R.J., and Francis, M.E. (2011). LIWC2007: Linguistic Inquiry and Word Count (Version 1.12) [Computer software]. Austin, TX: LIWC.net.

Pennebaker, J.W., Boyd, R.L., Jordan, K., and Blackburn, K. (2015). The development and psychometric properties of LIWC2015. Austin, TX: University of Texas at Austin.

Rosenstock, I.M. (1974). Historical origins of the health belief model. *Health education monographs, Vol. 2, Issue 4*

Scanfeld, D., Scanfeld, V., and Larson, E. L. (2010). Dissemination of health information through social networks: Twitter and antibiotics. *American journal of infection control, Vol. 38, Issue 3, pp. 182– 188.*

Sentiment. (2018) In Cambridge Dictionary. Retrieved June 1st, 2018 from https://dictionary.cambridge.org/dictionary/english/sentiment

Skin Cancer Foundation (n.d.). May is Skin Cancer Awareness Month. Retrieved April 1st, 2018 from https://www.skincancer.org/get-involved/skin-cancer-awareness-month

Skin cancer (n.d.). Common skin cancer types. Retrieved April 3rd, 2018 from https://www.cancer.org/cancer/skin-cancer.html

Vance, K., Howe, W., and Dellavalle, R. (2009). Social internet sites as a source of public health information. *Dermatologic clinics, Vol. 27, Issue 2, pp. 133–136.*

Wakefield, M.A., Loken, B., Hornik, R.C. (2010). Use of mass media campaigns to change health behavior. *Vol. 376, Issue 9748, pp. 1261-1271*

Warnecke, R.B., Sudman, S., Johnson, T.P., O'Rourke, D., Davis, A.M. and Jobe, J.B. (1997). Cognitive aspects of recalling and reporting health-related events: Papanicolaou smears, clinical breast examinations, and mammograms. *American Journal of Epidemiology. Vol. 146, pp. 982–992.*

Wiggins, J. S. (1973). *Personality and prediction: Principles of personality assessment.* Reading, MA: Addison-Wesley.

Wong, K.O., Davis, F.G., Zaiane, O.R., Yasui, Y. (2016). Sentiment analysis of breast cancer screening in the United States using Twitter. *Vol. 1, pp. 265-274*

Yarbrough, S., Braden, C.J. (2001). Utility of health belief model as a guide for explaining or predicting breast cancer screening behaviors. *Journal of Advanced Nursing, Vol.33 Issue 5, pp. 677-688*

Yazdanifard, R., Obeidy, W.K., Yusoff, W.F.W, Babaei, H.R. (2011). Social network and microblogging; the emerging marketing tcharends & tools of the twenty-first century. International conference on computer communication and management.

10. APPENDIX

10.1 Sentiment classification for each of the message categories presented in the paper