

# An Analysis of Personal Medical Information Disclosed in YouTube Videos Created by Patients with Multiple Sclerosis

Luis FERNANDEZ-LUQUE<sup>a,1</sup>, Najeed ELAHI<sup>b</sup>, Francisco J. GRAJALES III<sup>c</sup>

<sup>a</sup> *Northern Research Institute (Norut), Tromsø, Norway*

<sup>b</sup> *Norwegian Center of Telemedicine, University Hospital of North Norway, Tromsø, Norway*

<sup>c</sup> *eHealth Strategy Office, Faculty of Medicine, The University of British Columbia, Vancouver, Canada*

**Abstract.** The Internet has become one of the main sources of health information. Today, content generation is no longer limited to the healthcare professionals of the late nineties; Web 2.0 services and platforms have empowered patients to create and interact with various forms of Patient-Generated Content (PGC); these include: videos, blogs, and social networking pages, among others. This investigation evaluated the characteristics of PGC found within YouTube video comments. We selected a random sample of 25 out of 769 Multiple Sclerosis patient-generated videos and analyzed their corresponding 557 comments for health information. 320 comments met the inclusion criteria and 70 contained personal health information (PHI). Comments with PHI were sub-characterized for the type of medical information (i.e., diagnosis, date of diagnosis, medication, among others). In this descriptive study, we present the strata within this content and postulate some of the corresponding patient risks and ethical challenges associated with PGC found in YouTube video comments.

**Keywords.** Web 2.0, privacy, YouTube, health information, patient-generated content

## 1. Introduction

The Internet is one of the main sources of health information. Recent studies have found that most Europeans are using the Internet to search for health information [1]. Web 2.0 platforms like YouTube, Facebook, and Blogger are among the most popular. The common characteristic of all these platforms is the availability of User-Generated Content. This has lead to a new scenario where patients are not only accessing content, but also creating new content (e.g., blogs, videos); predominantly, this content is also used as a form of primary communication between the users, especially when commenting or “evaluating” the content created by other patients. For example, there are users who publish videos relating to their everyday life; including being blind, deaf or coping with chronic debilitating diseases. Many of these videos also contain comments from other users, mainly patients and relatives, which contain personal

---

<sup>1</sup> Corresponding Author: Northern Research Institute, Postboks 6434 Forskningsparken, 9294 Tromsø, Norway; E-mail: luis.luque@norut.no.

health information or medical advice. Although we have published research on the usage of the web services by health consumers elsewhere [2], there is paucity in the literature with regards to the characteristics of user comments posted within Patient-Generated Content (PGC). Other researchers have explored the content quality of YouTube videos and the disclosure of personal health information within social network public profiles [3–5]. The analysis of comments is important due the prevalence and incidence of Web 2.0 platforms and the PGC within; with issues related to privacy and ethics. Additionally, automatic comment analysis, through the use of crawlers and algorithms, may be used to harvest more information about the users for personalization purposes (i.e., user-targeted advertising) [6]. In our project MyHealthService [7] we are investigating the use of PGC for the enhancement and personalization of health education.

The overarching objective of our study is to increase the understanding of the Patient 2.0 phenomenon. In particular, it is focused on PGC found within YouTube video comments created by patients with Multiple Sclerosis and the disclosure of personal health information within.

## 2. Method

During the first week in December of 2008, we queried the YouTube database for users who had published at least three videos in English about their everyday life with Multiple Sclerosis. Twenty-seven patients were found and using the YouTube application-programming interface, user, video and comment data were extracted. A total of 769 videos, 7,047 comments and 2,426 user profiles were collected (video creators and comments). Using random selection, we selected 25 of these videos and analyzed their corresponding 557 comments. Comments which were not posted on health-related videos or were not written in English were excluded. A final sample of 320 comments was analyzed and classified into the following strata:

- Personal health information: related to personal health experiences. Subcategories: diagnosis, date of diagnosis, symptoms and medication.
- Video discussion: discussions about the videos (e.g., adding information or discussing the video topic).
- Appreciation: comments expressing gratitude for content or comments.
- Criticism: Video content criticism based on the content or quality of the information contained within the video.
- Unrelated comments: comments not covered in the other categories.

No ethical approval was necessary for the data used and gathered for this study; we used publicly accessible data that was voluntarily released by YouTube's users.

## 3. Results

The majority of comments met more than one category definition (Figure 1) and most of them discussed the video (77%, n=247) or expressed gratitude for the author(s) (55%, n=177). A total of 22 comments met the unrelated comment criteria (e.g., *\*\*\*~KABOOOOM~\*\*\**) and eight criticized video authors (e.g., *I'm skeptical you*

REALLY have MS). Overall, 70 comments (22%) contained personal health information concerning their creators or a third party (e.g., relatives).

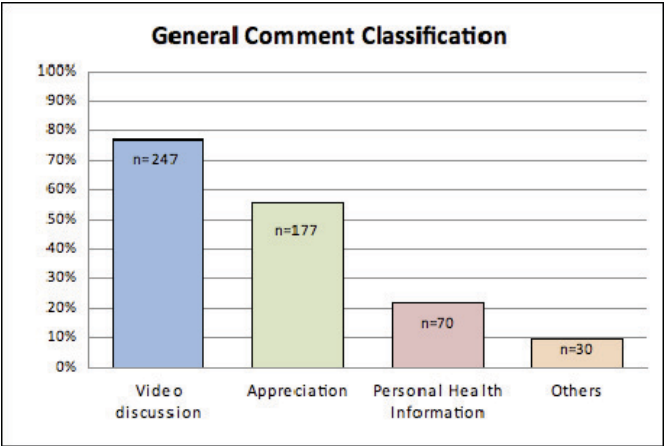


Figure 1. Total number of comments classified into the main categories

The comments with personal health information (PHI) were further stratified. As Figure 2 denotes, almost half of the comments contained information about medications (73%, n=51). Comments about symptoms (50%, n=35) and diagnoses (39%, n=27) were also prevalent. In one case, the information disclosed was the PHI of a third party:

*I have been watching your videos since my daughter was diagnosed with MS on 28-12-07. She was diagnosed with an aggressive form of MS. Betaferon caused liver problems in a very short time. She has now had 4 infusions of tysabri and now feels she is well enough to try and go back to work. Tysabri only became available in Aust on July 1/08. This drug has given her hope that she still has a future to look forward to as she is 26 yrs old. Your improvement since your first dose gave us all hope.*

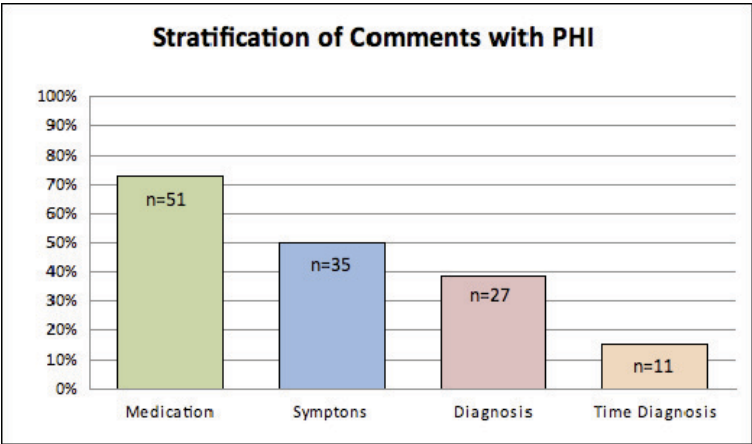


Figure 2. Total number of comments with personal health information by category

Comments also encouraged other patients to take or change their medications; for example: *“im in a wheelchair is it worthwhile taking tysabri?”* The response by the video author was: *“In a word: ABSOLUTELY! It is amazing how much freedom I recovered when I started walking!”*

#### 4. Discussion

Our results show that although the comments with personal health information had a low prevalence, the importance of the information disclosed within is extremely complex. Comments contained detailed user information. The public availability of these comments could constitute a risk to the user's privacy even in if their comments were written under an “anonymous” alias. For example, a YouTube's nickname is often found in other Web 2.0 sites like Flickr or Blogger. This could lead a “curious” investigator to find other profiles in social networking sites, like Facebook and consequently access family member information (e.g., the Facebook profile of the daughter diagnosed with MS). This not as unlikely as it appears when, for example, 25% of the Norwegian population have a profile on Facebook [8]. We also found that some of the commenter's YouTube profiles had been deleted, but their comments remained in the video database. This would make it nearly impossible to delete this public information. The public availability of this information could even have further implications that reach beyond our current ethical paradigms. For example, if the comment's author is identified, could it be used to deny health insurance coverage? Or, could a patient legally sue another for providing medical advice through YouTube?

The majority of comments with personal health information were related to medications. Several users were seeking advice on whether to start or change medication. We were surprised by the number of comments and videos about medications, especially about Tysabri, one of the newest drugs for Multiple Sclerosis. Tysabri is prescribed when a patient is un-responsive to other treatments or when the patient is too affected by side effects. However, many video authors reported recovery since they started Tysabri treatment. The majority of comments in these videos came from patients who were considering this new treatment. For these reasons, PGC and corresponding patient risks need to be studied in further detail. Further research is necessary to understand the implications of PGC in relation to privacy, security, and ethical paradigms.

At present, we are developing an automated semantic-enhanced modeling system of educational resources (e.g., YouTube videos) as part of a project aimed at delivering personalized recommendations for health education. Semantic-based modeling and text analysis techniques could be used for analyzing the comments, e.g., to identify the topics of the videos. Berendt and Navigli [9] have proposed two methods for semantic-enhanced text analysis that allow analysts to integrate domain-specific and general background knowledge. A similar approach has also been suggested by Sheth and Nagarajan [6] where SWRL [10] rules can be defined on top of ontology and serve for classification and information decision-making.

## 5. Conclusion

To our knowledge, this study is the first of its kind to analyze PGC for the disclosure of personal health information. Our study, although limited to a unique dataset and condition, has shown some of the risks users may be subject to, when disclosing medical information in PGC. Although our results cannot be generalized to all conditions, they highlight the important need for further research in this subject. It is particularly important to understand patient's intents and awareness on risk when disclosing personal health information. This research may also contribute to the formation of health consumer guidelines for safety in Web 2.0 service utilization such as not reusing web aliases in platforms where personal health information is publicly disclosed.

Videos analyzed for this study represent a small sample within YouTube's database of approximately 121,200 total video views. It is essential to standardize and promote the adherence to health information quality standards (i.e., HONcode or Patient Blogger Code of Ethics) by the health consumers who are creating content on the Internet. In addition, we believe that studying patients as providers of health information could improve the access and quality of health information on the Internet.

**Acknowledgements.** The authors would like to express their sincere gratitude to the Norwegian Research Council for co-funding this research at the Tromsø Telemedicine Laboratory. We would also like to express our sincere appreciation to the MyHealthService team and to patients using Web 2.0 services for inspiring our work.

## References

- [1] Andreassen, H.K., Bujnowska-Fedak, M.M., Chronaki, C.E., Dumitru, R.C., Pudule, I., Santana, S., Voss, H., Wynn, R. (2007) European citizens' use of e-health services: A study of seven countries. *BMC Public Health* 7:53.
- [2] Luque, L.F., Basagoiti, I., Johnsen, E. et al. (2008) Study of the ePatient as a generator of health content in the Internet. In Eysenbach, G. (Ed.) *Medicine 2.0 Proceedings*, Toronto, Canada, 28–29.
- [3] Keelan, J., Pavri-Garcia, V., Tomlinson, G., Wilson, K. (2007) YouTube as a source of information on immunization: A content analysis. *Journal of the American Medical Association* 298(21):2482–2484.
- [4] Burton, A. (2008) YouTube-ing your way to neurological knowledge. *Lancet Neurology* 7(12):1086–1087.
- [5] Bonander, J. (2008) Personally tailored health information: A health 2.0 approach. In Eysenbach, G. (Ed.) *Medicine 2.0 Proceedings*, Toronto, Canada, 12.
- [6] Sheth, A., Nagarajan, M. (2009) Semantics-empowered social computing. *IEEE Internet Computing* 13, 1:76–80.
- [7] Burkow, T.M., Vognild, L.K., Krogstad, T., Borch, N., Ostengen, G., Bratvold, A., Risberg, M.J. (2008) An easy to use and affordable home-based personal eHealth system for chronic disease management based on free open source software. In *Proceedings of MIE 2008, Studies in Health Technology and Informatics* 136:83–88.
- [8] insidefacebook.com. (2008) Tracking Facebook's 2008 international growth by country July 29<sup>th</sup>, 2008. <http://www.insidefacebook.com/2008/07/29/tracking-facebook-2008-international-growth-by-country/>. (Archived by WebCite® at <http://www.webcitation.org/5eLLehtpM>).
- [9] Berendt, B., Navigli, R. (2006) Finding your way through blogspace: Using semantics for cross-domain blog analysis. In Nicolov, N., Salvetti, F., Martin, J.H., Liberman, M. (Eds.) *Proceedings of the AAAI 2006 Symposium on Computational Approaches to Analysing Weblogs*, Stanford, Technical Report SS-06-03, AAAI Press, Menlo Park, 1–8.
- [10] W3C Member Submission (2004) *SWRL: A Semantic Web Rule Language Combining OWL and RuleML*, <http://www.w3.org/Submission/SWRL/>.