

Evaluation of Internet Social Networks using Net scoring Tool: A Case Study in Adverse Drug Reaction Mining

Sandrine Katsahian^{a,c,d,h}, Erica Simond Moreau^h, Damien Leprovost^{b,c,f}, Jeremy Lardon^{b,e}, Cedric Bousquet^{b,c}, Gaétan Kerdelhué^g, Redhouane Abdellaouiⁱ, Nathalie Texierⁱ, Anita Burgun^{a,c,d,h}, Abdelali Boussadi^{a,c,d,h} and Carole Faviezⁱ

^aINSERM, UMR_S 1138, équipe 22, Centre de Recherche des Cordeliers, F-75006, Paris, France

^bINSERM, U1142, LIMICS, F-75006, Paris, France

^cSorbonne Universités, UPMC Univ Paris 06, UMR_S 1138, Centre de Recherche des Cordeliers, F-75006, Paris, France

^dUniversité Paris Descartes, Sorbonne Paris Cité, UMR_S 1138, Centre de Recherche des Cordeliers, F-75006, Paris, France

^eUniversity of Saint Etienne, Department of Public Health and Medical Informatics, Saint-Etienne, France.

^fUniversité Paris 13, Sorbonne Paris Cité, F-93430, Villetaneuse, France.

^gCISMef, Rouen University Hospital, Cour Leschevin, Rouen, Cedex, France.

^hAP-HP, HEGP, Département d'Informatique Hospitalière, Paris, France.

ⁱKappa Santé, Paris, France.

Abstract. Background and objectives: Suspected adverse drug reactions (ADR) reported by patients through social media can be a complementary tool to already existing ADRs signal detection processes. However, several studies have shown that the quality of medical information published online varies drastically whatever the health topic addressed. The aim of this study is to use an existing rating tool on a set of social network web sites in order to assess the capabilities of these tools to guide experts for selecting the most adapted social network web site to mine ADRs. **Methods:** First, we reviewed and rated 132 Internet forums and social networks according to three major criteria: the number of visits, the notoriety of the forum and the number of messages posted in relation with health and drug therapy. Second, the pharmacist reviewed the topic-oriented message boards with a small number of drug names to ensure that they were not off topic. Six experts have been chosen to assess the selected internet forums using a French scoring tool: Net scoring. Three different scores and the agreement between experts according to each set of scores using weighted kappa pooled using mean have been computed. **Results:** Three internet forums were chosen at the end of the selection step. Some criteria get high score (scores 3-4) no matter the website evaluated like accessibility (45-46) or design (34-36), at the opposite some criteria always have bad scores like quantitative (40-42) and ethical aspect (43-44), hyperlinks actualization (30-33). Kappa were positives but very small which corresponds to a weak agreement between experts. **Conclusion:** The personal opinion of the expert seems to have a major impact, undermining the relevance of the criterion. Our future work is to collect results given by this evaluation grid and proposes a new scoring tool for Internet social networks assessment.

Keywords. Adverse drug reaction, Internet, quality measurement, evaluation

1. Introduction

Although some adverse drug reaction (ADR) are identified during clinical trials, many other (new) ADR escape to these pre marketing processes, monitoring of ADR is then delegated to post marketing surveillance systems mostly based on spontaneous reporting databases [1]. However, in recent years, these systems have shown several limitations mainly due to underreporting [1].

Online social networking is increasingly used for communication among patients with the same health concerns and also among patients and their physicians [2]. Patients share a lot of information in these web sites in relation with their health condition, illnesses, feelings, medications taken and as well as ADR ...and they receive emotional support from each other and from their physicians. For psychological reasons, patients are often more comfortable sharing personal experiences in support groups, with other participants who are going through similar issues [3].

Suspected ADR reported by patients through social media can be a complementary tool to already existing ADR signal detection processes [4]. Different studies have already and successfully mine different kind of social media with different strategies for ADR signal detection [5][6][7]. In France, a study by Abou Taam et al. [8] strongly suggests that analysis of health forums can inform on drug ADR.

However, several studies have shown that the quality of medical information published online varies drastically whatever the health topic addressed [9][10][11]. Seeking useful and valid information on the Internet can be difficult because of the lack of rigorous control and upgrade of the material conveyed in these web sites. Filtering through information on the Internet may also be very time consuming. These issues led main key opinion leaders to argue for web- sites evaluation and rating tools for patients and also for domain experts [12].

Many of these rating tools already exist [13]: the Silberg scores, the HogneSandvik scale, the Jim Kapoun's criteria, the Health Information Technology Institute (HITI) criteria, the HON (health on the net) Code of Conduct for medical and health Web sites and for the French web sites, the Net scoring tool [14]. However, none of these tools have been originally designed to evaluate and rate social network web sites which are one of the five major aspects that emerged from web 2.0 in health care with participation, apomediation, collaboration, and openness [15].

The aim of this study is to use the existing rating tools on a set of social network web sites in order to assess the capabilities of these tools to guide experts for selecting the most adapted social network web site to mine ADR.

2. Methods

2.1. Forums and Internet social networks selection

First, we identified, reviewed and rated 132 French Internet forums and social networks according to three major criteria: the number of visits, the notoriety of the forum and the number of messages posted in relation with health and drug therapy. To estimate those three parameters, we used: i) the Cismef web site (www.cismef.org), a catalog and index of French language health resources on the Internet to estimate the total number of messages ii) the 1001 forums website (<http://www.1001forums.fr/>) which is one of the larger French forum's indexer on the Internet to estimate the

number of messages per day and iii) Google, Alexa (<http://www.alexa.com/>) and Yoovi (<http://www.yoovi.com/>) to estimate the notoriety, traffic data and global ranking of the forum. Second, private and non-active Internet forums were discarded. Third, each Internet forum was checked by a pharmacist with a small number of drug names to ensure that they were not off topic.

2.2. Data collection

Six experts with different backgrounds: four computer science researchers in the field of social networks, one physician and one statistician with medical informatics backgrounds were chosen in the context of the ADR-Prism project (<http://adr-prism.com/>). The experts were asked to use Net scoring tool [14] composed of 46 criterion on a 4 point-Likert scale, where 1 indicates very bad agreement, 2 bad agreement, 3 good agreement and 4 indicates total agreement. Experts were also asked to assess the interest and the importance of each criterion as: “to keep”, “to revise” or “to remove”. In addition, experts were allowed to propose new, non-mentioned criterion and to give a reason of their propositions. Before the assessment step, experts were briefed on the definition of each criterion and how to use them. Each expert received an Excel file with the different criterion and was asked to rate the Internet forums selected.

2.3. Data analysis

At the end of the assessment step by the experts, three different scores were computed for each Internet forum rated using the following formulas:

Table 1: Formulas used to compute the different scores according to the different way of handling the “not found” values

$ScoreA = \frac{\sum \text{criteriaLikert note}}{\sum \text{max Likert note}}$	Where « Not found » values are omitted
$ScoreB = \frac{\sum \text{criteriaLikert}}{\sum \text{max score}}$	Where « Not found » values correspond the worst score of 1
$ScoreC = \frac{\sum(\text{criteria scores} \neq \text{"notfound"}) - \text{penalty} \times \text{numberofcriteria "Notfound"}}{4 \times \text{numberofcriteriafound}}$	Where « Not found » values correspond to a “penalty” of “-2”

We estimated the agreement between experts according to each set of scores using weighted kappa pooled using mean in order to take into consideration the distance between the scores. “Not found” values were considered in two different ways: i) as missing data, ii) 0 was assigned to the not found values. The value of 0 was chosen in a consistence way as distance to avoid biasing weighted kappa computation.

3. Results

3.1. The expert's scores

Three forums were chosen at the end of the selection step: Asperansa (<http://www.asperansa.org>) which is an Internet forum dedicated to adolescents, young adults with autism and high level of Asperger syndrome and their parents, Atoute (<http://www.atoute.org/>) and Doctissimo (<http://www.doctissimo.fr/>), which are a generalist web sites and forums dedicated to health topics. Table 2 describes the three scores computed for each Internet forum rated using the 46 criterion-based Internet scoring tool. Table 2 shows that despite the score i.e. despite the way to deal with missing data, forums rated keep the same ranking. Doctissimo has a better quality score than atoute and asperansa. Scores are logically higher when not found values are omitted than when we replaced them with the worst score or a penalty.

Table 2: The three scores computed for each Internet social network

		Doctissimo	Atoute	Asperansa
score A	mean	80%	78%	73%
	median	80%	79%	75%
score B	mean	72%	71%	66%
	median	73%	70%	62%
score C	mean	71%	69%	62%
	median	73%	69%	58%

Figure 1 shows that some criteria get high score (scores 3-4) no matter the website evaluated like accessibility (45-46) or design (34-36) ; at the opposite some criteria always have bad scores like quantitative (40-42) and ethical aspect (43-44), hyperlinks actualization (30-33).

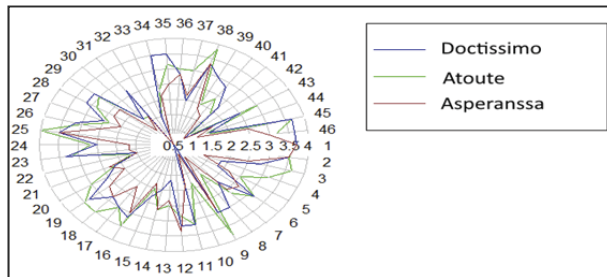


Figure 1: Radar plot of mean of scores computed for the 6 experts for each 46 criterion of the Net scoring tool

3.2. The expert's agreement

Table 3 shows that Kappa are positives but very small which corresponds to a weak agreement between experts.

Table 3: Kappa expert's agreement for the three selected Internet social networks

	Doctissimo	Atoute	Asperansa
Where « Not found » values are omitted	0.03870397	0.06770932	0.05367282
Where « Not found » values correspond to 0	0.08845810	0.09893061	0.07135463

4. Discussion

The results show that if some criteria are quite consensual, others are randomizing the results. The personal opinion of the expert seems to have a major impact, undermining the relevance of the criterion. In order to maximize the agreement between experts, it appears we need to make a rigorous selection of criteria. Our future work is to collect results given by this evaluation grid and proposes a new scoring tool adapted to Internet social networks assessment.

Acknowledgement: We are thankful to: Hadyl Asfari, Neyla Kerras, Adel Mebarki and Nassim Nafti for their contribution. This article is part of ADR prism project financed by the DGE and territorial collectivities (Ile de France and Haute Normandie) under the 16th FUI's (Fond Unique Interministériel) call for project.

References

- [1] M. Molokhia, S. Tanna, et D. Bell, « Improving reporting of adverse drug reactions: Systematic review », *Clin. Epidemiol.*, vol. 1, p. 75-92, août 2009.
- [2] S. Fox, « Health Topics », *Pew Research Center's Internet & American Life Project*.
- [3] K. P. Davison, J. W. Pennebaker, et S. S. Dickerson, « Who talks? The social psychology of illness support groups », *Am. Psychol.*, vol. 55, n° 2, p. 205-217, févr. 2000.
- [4] A. Blenkinsopp, P. Wilkie, M. Wang, et P. A. Routledge, « Patient reporting of suspected adverse drug reactions: a review of published literature and international experience », *Br. J. Clin. Pharmacol.*, vol. 63, n° 2, p. 148-156, févr. 2007.
- [5] Y. A. Li, « Medical data mining : improving information accessibility using online patient drug reviews », Thesis, Massachusetts Institute of Technology, 2011.
- [6] A. Yates et N. Goharian, « ADRTrace: Detecting Expected and Unexpected Adverse Drug Reactions from User Reviews on Social Media Sites », in *Advances in Information Retrieval*, P. Serdyukov, P. Braslavski, S. O. Kuznetsov, J. Kamps, S. Rüger, E. Agichtein, I. Segalovich, et E. Yilmaz, Éd. Springer Berlin Heidelberg, 2013, p. 816-819.
- [7] B. W. Chee, R. Berlin, et B. Schatz, « Predicting Adverse Drug Events from Personal Health Messages », *AMIA Annu. Symp. Proc.*, vol. 2011, p. 217-226, 2011.
- [8] M. Abou Taam, C. Rossard, L. Cantaloube, N. Bouscaren, G. Roche, L. Pochard, F. Montastruc, A. Herxheimer, J. L. Montastruc, et H. Bagheri, « Analysis of patients' narratives posted on social media websites on benfluorex's (Mediator®) withdrawal in France », *J. Clin. Pharm. Ther.*, vol. 39, n° 1, p. 53-55, févr. 2014.
- [9] S. Lee, J. J. Shin, M. S. Haro, S. H. Song, et S. J. Nho, « Evaluating the Quality of Internet Information for Femoroacetabular Impingement », *Arthrosc. J. Arthrosc. Relat. Surg. Off. Publ. Arthrosc. Assoc. N. Am. Int. Arthrosc. Assoc.*, vol. 30, n° 10, p. 1372-1379, oct. 2014.
- [10] S. Morr, N. Shanti, A. Carrer, J. Kubeck, et M. C. Gerling, « Quality of information concerning cervical disc herniation on the Internet », *Spine J. Off. J. North Am. Spine Soc.*, vol. 10, n° 4, p. 350-354, avr. 2010.
- [11] R. M. Garcia, P. J. Messerschmitt, et N. U. Ahn, « An evaluation of information on the Internet of a new device: the lumbar artificial disc replacement », *J. Spinal Disord. Tech.*, vol. 22, n° 1, p. 52-57, févr. 2009.
- [12] A. R. Jadad et A. Gagliardi, « Rating health information on the Internet: navigating to knowledge or to Babel? », *JAMA*, vol. 279, n° 8, p. 611-614, févr. 1998.
- [13] A. Aslani, O. Pournik, A. Abu-Hanna, et S. Eslami, « Web-site evaluation tools: a case study in reproductive health information », *Stud. Health Technol. Inform.*, vol. 205, p. 895-899, 2014.
- [14] SJ. Darmoni, V. Leroux, M. Daigne, B. Thirion, P. Santamaria, C. Duvaux, M. Gea. Net Scoring : critères de qualité de l'information de santé sur l'Internet. *Technologie Santé* , Volume 36, Pages 128-142, 1999.
- [15] G. Eysenbach, « Medicine 2.0: Social Networking, Collaboration, Participation, Apomediation, and Openness », *J. Med. Internet Res.*, vol. 10, n° 3, août 2008.