

Publications scientifiques



Kap^oCode
FROM DATA TO HEALTH



L'équipe Kap Code



Stéphane Schück

PRÉSIDENT



Nathalie Texier

VICE-PRÉSIDENT



Adel Mebarki

DIRECTEUR GÉNÉRAL



Paméla Voillot

RESPONSABLE ÉTUDES &
OPÉRATIONS



Paul Loussikian

RESPONSABLE INNOVATION &
PARTENARIAT STRATÉGIQUE



Tom Marty

RESPONSABLE MÉDICAL



Joelle Malaab

CHEF DE PROJET



Imad Bendimerad

CHEF DE PROJET MÉDICAL



Anaïs Gedik

RESPONSABLE DATA SCIENCE



Mickaïl Khadhar

DATA SCIENTIST



Manissa Talmatkadi

DATA SCIENTIST



Quels sont les principaux symptômes du Covid Long ? Données médicalisées de vie réelle croisés avec les contenus des réseaux sociaux



Claire Paris, Tom Marty, Medhi Dahoumane, Caroline Eteve-Pitsaer, Adel Mebarki, Charlotte Renaudat

Kap•Code



QUELS SONT LES PRINCIPAUX SYMPTÔMES DU COVID LONG ? DONNÉES MÉDICALISÉES DE VIE RÉELLE CROISÉS AVEC LES CONTENUS DES RÉSEAUX SOCIAUX



Relecture du poster et discussion des résultats

Claire Paris⁽¹⁾, Tom Marty⁽²⁾, Medhi Dahoumane⁽³⁾, Caroline Eteve-Pitsaer⁽¹⁾, Adel Mebarki⁽³⁾, Charlotte Renaudat⁽²⁾

⁽¹⁾GERS SAS-Cegedim Health Data, Boulogne Billancourt, France

⁽²⁾Kap Code, 146 rue Montmartre, 75002 Paris, France

⁽³⁾Kynapse, 28 rue Jacques Ibert, 92300 Levallois-Perret, France



INTRODUCTION

Contexte

- Mars 2020 : émergence pandémie mondiale COVID-19
- Emergence infection COVID long : persistance des symptômes
- Maladie décrites par les patients qui cherchent compréhension et reconnaissance
- La majorité des études ont été réalisées chez des patients ayant été hospitalisés [1] pour une infection COVID sévère
- Le médecin généraliste reste le premier contact médical des patients, quelle qu'ait été la gravité de leur infection initiale [2]

Objectif principal

- Examiner, à grande échelle, les principaux symptômes rapportés par les patients souffrant de COVID long
- Croiser 2 différentes sources de données de vie réelle, recueillies au cours du suivi médical des patients ou issues des témoignages publics des patients sur les réseaux sociaux

MATÉRIEL ET MÉTHODES

Réseaux sociaux

Messages réseaux sociaux francophones
Entre Janvier 2020 - Août 2021
Via outil Brandwatch® et dictionnaire MedDRA
5 364 internautes (6 484 messages)

Type d'étude

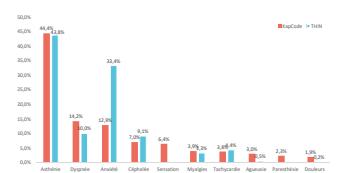
observationnelle rétrospective

Consultations médicales

Symptômes et diagnostics COVID long en absence autre pathologie [3] dans la base de données THIN France, 6 mois post infection aigüe
Entre Mars 2020 – Mars 2022
73 338 patients lors consultation médecine générale

RÉSULTATS

Les symptômes associés au COVID long: réseaux sociaux vs consultation



Type d'étude

observationnelle rétrospective

Consultations médicales

Symptômes et diagnostics COVID long en absence autre pathologie [3] dans la base de données THIN France, 6 mois post infection aigüe

Entre Mars 2020 – Mars 2022

73 338 patients lors consultation médecine générale

Les symptômes associés au COVID long: réseaux sociaux vs consultation



Type d'étude

observationnelle rétrospective

Consultations

- 20,1% des patients associent douleurs abdominales & diarrhées
- 17,1% des patients associent asthénie & anxiété
- 9,9% des patients associent asthénie & dyspnée

Symptômes et diagnostics COVID long en absence autre pathologie [3] dans la base de données THIN France, 6 mois post infection aigüe

Entre Mars 2020 – Mars 2022

73 338 patients lors consultation médecine générale

Résumé

Les 3 symptômes dans les deux populations identifiées COVID long sont les mêmes: asthénie, dyspnée et anxiété.

Co-occurrence des symptômes concomitants selon les sources



Résumé

Les 3 symptômes dans les deux populations identifiées COVID long sont les mêmes: asthénie, dyspnée et anxiété.

Co-occurrence des symptômes concomitants selon les sources



Résumé

- 35,3% des patients associent asthénie & dyspnée
- 22,5% des patients associent asthénie & anxiété
- 17,3% des patients associent asthénie & céphalée

RÉSULTATS

Les symptômes associés au COVID long: réseaux sociaux vs consultation



Type d'étude

observationnelle rétrospective

Consultations

- 20,1% des patients associent douleurs abdominales & diarrhées
- 17,1% des patients associent asthénie & anxiété
- 9,9% des patients associent asthénie & dyspnée

Symptômes et diagnostics COVID long en absence autre pathologie [3] dans la base de données THIN France, 6 mois post infection aigüe

Entre Mars 2020 – Mars 2022

73 338 patients lors consultation médecine générale

RÉSULTATS

Les symptômes associés au COVID long: réseaux sociaux vs consultation



Type d'étude

observationnelle rétrospective

Consultations

- 20,1% des patients associent douleurs abdominales & diarrhées
- 17,1% des patients associent asthénie & anxiété
- 9,9% des patients associent asthénie & dyspnée

Symptômes et diagnostics COVID long en absence autre pathologie [3] dans la base de données THIN France, 6 mois post infection aigüe

Entre Mars 2020 – Mars 2022

73 338 patients lors consultation médecine générale

RÉSULTATS

Les symptômes associés au COVID long: réseaux sociaux vs consultation



Type d'étude

observationnelle rétrospective

Consultations

- 20,1% des patients associent douleurs abdominales & diarrhées
- 17,1% des patients associent asthénie & anxiété
- 9,9% des patients associent asthénie & dyspnée

Symptômes et diagnostics COVID long en absence autre pathologie [3] dans la base de données THIN France, 6 mois post infection aigüe

Entre Mars 2020 – Mars 2022

73 338 patients lors consultation médecine générale

RÉSULTATS

Les symptômes associés au COVID long: réseaux sociaux vs consultation

Type d'étude

observationnelle rétrospective

Consultations

- 20,1% des patients associent douleurs abdominales & diarrhées
- 17,1% des patients associent asthénie & anxiété
- 9,9% des patients associent asthénie & dyspnée

Symptômes et diagnostics COVID long en absence autre pathologie [3] dans la base de données THIN France, 6 mois post infection aigüe

Entre Mars 2020 – Mars 2022

73 338 patients lors consultation médecine générale

RÉSULTATS

Les symptômes associés au COVID long: réseaux sociaux vs consultation

Type d'étude

observationnelle rétrospective

Consultations

- 20,1% des patients associent douleurs abdominales & diarrhées
- 17,1% des patients associent asthénie & anxiété
- 9,9% des patients associent asthénie & dyspnée

Symptômes et diagnostics COVID long en absence autre pathologie [3] dans la base de données THIN France, 6 mois post infection aigüe

Entre Mars 2020 – Mars 2022

73 338 patients lors consultation médecine générale

RÉSULTATS

Les symptômes associés au COVID long: réseaux sociaux vs consultation

Type d'étude

observationnelle rétrospective

Consultations

- 20,1% des patients associent douleurs abdominales & diarrhées
- 17,1% des patients associent asthénie & anxiété
- 9,9% des patients associent asthénie & dyspnée

Symptômes et diagnostics COVID long en absence autre pathologie [3] dans la base de données THIN France, 6 mois post infection aigüe

Entre Mars 2020 – Mars 2022

73 338 patients lors consultation médecine générale

RÉSULTATS

Les symptômes associés au COVID long: réseaux sociaux vs consultation

Type d'étude

observationnelle rétrospective

Consultations

- 20,1% des patients associent douleurs abdominales & diarrhées
- 17,1% des patients associent asthénie & anxiété
- 9,9% des patients associent asthénie & dyspnée

Symptômes et diagnostics COVID long en absence autre pathologie [3] dans la base de données THIN France, 6 mois post infection aigüe

Entre Mars 2020 – Mars 2022

73 338 patients lors consultation médecine générale

RÉSULTATS

Les symptômes associés au COVID long: réseaux sociaux vs consultation

Type d'étude

observationnelle rétrospective

Consultations

- 20,1% des patients associent douleurs abdominales & diarrhées
- 17,1% des patients associent asthénie & anxiété
- 9,9% des patients associent asthénie & dyspnée

Symptômes et diagnostics COVID long en absence autre pathologie [3] dans la base de données THIN France, 6 mois post infection aigüe

Entre Mars 2020 – Mars 2022

73 338 patients lors consultation médecine générale

RÉSULTATS

Les symptômes associés au COVID long: réseaux sociaux vs consultation

Type d'étude

observationnelle rétrospective

Consultations

- 20,1% des patients associent douleurs abdominales & diarrhées
- 17,1% des patients associent asthénie & anxiété
- 9,9% des patients associent asthénie & dyspnée

Symptômes et diagnostics COVID long en absence autre pathologie [3] dans la base de données THIN France, 6 mois post infection aigüe

Entre Mars 2020 – Mars 2022

73 338 patients lors consultation médecine générale

RÉSULTATS

Les symptômes associés au COVID long: réseaux sociaux vs consultation

Type d'étude

observationnelle rétrospective

Consultations

- 20,1% des patients associent douleurs abdominales & diarrhées
- 17,1% des patients associent asthénie & anxiété
- 9,9% des patients associent asthénie & dyspnée

Symptômes et diagnostics COVID long en absence autre pathologie [3] dans la base de données THIN France, 6 mois post infection aigüe

Entre Mars 2020 – Mars 2022

73 338 patients lors consultation médecine générale

RÉSULTATS

Les symptômes associés au COVID long: réseaux sociaux vs consultation

Type d'étude

observationnelle rétrospective

Consultations

- 20,1% des patients associent douleurs abdominales & diarrhées
- 17,1% des patients associent asthénie & anxiété
- 9,9% des patients associent asthénie & dyspnée

Symptômes et diagnostics COVID long en absence autre pathologie [3] dans la base de données THIN France, 6 mois post infection aigüe

Entre Mars 2020 – Mars 2022

73 338 patients lors consultation médecine générale

RÉSULTATS

Les symptômes associés au COVID long: réseaux sociaux vs consultation

Type d'étude

observationnelle rétrospective

Consultations

- 20,1% des patients associent douleurs abdominales & diarrhées
- 17,1% des patients associent asthénie & anxiété
- 9,9% des patients associent asthénie & dyspnée

Symptômes et diagnostics COVID long en absence autre pathologie [3] dans la base de données THIN France, 6 mois post infection aigüe

Entre Mars 2020 – Mars 2022

73 338 patients lors consultation médecine générale

RÉSULTATS

Les symptômes associés au COVID long: réseaux sociaux vs consultation

Type d'étude

observationnelle rétrospective

Consultations

- 20,1% des patients associent douleurs abdominales & diarrhées
- 17,1% des patients associent asthénie & anxiété
- 9,9% des patients associent asthénie & dyspnée

Symptômes et diagnostics COVID long en absence autre pathologie [3] dans la base de données THIN France, 6 mois post infection aigüe

Entre Mars 2020 – Mars 2022

73 338 patients lors consultation médecine générale

RÉSULTATS

Les symptômes associés au COVID long: réseaux sociaux vs consultation

Type d'étude

observationnelle rétrospective

Consultations

- 20,1% des patients associent douleurs abdominales & diarrhées
- 17,1% des patients associent asthénie & anxiété
- 9,9% des patients associent asthénie & dyspnée

Symptômes et diagnostics COVID long en absence autre pathologie [3] dans la base de données THIN France, 6 mois post infection



Détection des parcours de soin décrits par les patients sur les réseaux sociaux, évaluation de plusieurs modèles de machine learning combinés



Imad Bendimerad, Manissa Talmatkadi, Anaïs Gedik, Tom Marty, Adel Mebarki, Nathalie Texier, Stéphane Schück

Kap•Code

DÉTECTION DES PARCOURS DE SOIN DÉCRITS PAR LES PATIENTS SUR LES RÉSEAUX SOCIAUX, ÉVALUATION DE PLUSIEURS MODÈLES DE MACHINE LEARNING COMBINÉS

Imad Bendimerad^{1*}, Manissa Talmatkadi^{1†}, Anaïs Gedik¹, Tom Marty¹, Adel Mebarki¹, Nathalie Texier¹, Stéphane Schück¹

¹Kap Code, 146 rue Montmartre 75002 Paris, France

Co-premiers auteurs

Detect' t

INTRODUCTION

Le parcours de soins, sa standardisation et sa coordination sont des enjeux de santé individuelle et publique, censés répondre aux problématiques d'évolution du système de santé, dans un contexte d'augmentation des maladies chroniques. Sur les réseaux sociaux, les communautés de patients échangent et partagent leur prise en charge. Analyser ces données de vie réelle provenant des échanges publics de patients permettrait d'augmenter la compréhension des difficultés associées à certaines étapes, afin d'orienter des efforts de restructuration, d'accompagnement et d'adaptation des parcours de soins, en fonction des besoins des patients.

MATÉRIEL ET MÉTHODES

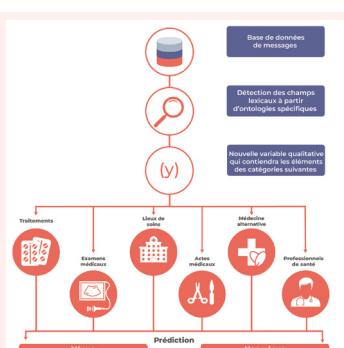
A partir de sources internet, entre forums et réseaux sociaux (Twitter, Facebook), une extraction de messages contenant des mots clés faisant référence à différentes pathologies a été réalisée. 1887 messages incluant au moins un élément de parcours de soins ont été sélectionnés. Les catégories du parcours de soin à détecter sont :

Les professionnels de santé consultés, les examens médicaux passés, les actes médicaux réalisés, les médecines alternatives recourues, les traitements pris ainsi que les lieux de santé visités.

Chaque message a été manuellement annoté en deux catégories : les expériences vécues et celles non vécues.

Sur la base de ce *gold standard*, un modèle binaire *xgboost* a été créé et entraîné par expérience afin de prédire si cette dernière a été vécue ou non par le patient (Figure 1).

Figure 1 - Algorithme parcours de soin



L'algorithme se base sur une sélection de variables parmi les catégories de variables suivantes :

Des variables de pronoms :

- Un comptage des pronoms au singulier (1^{ère}, 2^{ème} et 3^{ème} personnes)
- Une proportion des pronoms au singulier (1^{ère}, 2^{ème} et 3^{ème} personnes)

Des variables de sentiments :

- Un comptage de mots liés à différents sentiments

Des variables de temps des verbes :

- Un comptage des verbes sous différentes conjugaisons dans le message et autour du traitement

Des variables de négation :

- Un comptage des mots « ne » et « pas » dans le message et autour du traitement

Des variables de Regex :

- Champs lexicaux (CL) des différentes expériences couplées aux déterminants possessifs « mon » ou « ma »
- Fréquence de l'élément de parcours de soins dans le message
- Longueur de l'élément de parcours de soins dans le message

RÉSULTATS

Afin de valider ces modèles, ces derniers ont été appliqués sur un corpus de messages traitant un sujet d'addictologie et de volonté de sevrage. 100 messages sélectionnés aléatoirement par catégorie ont été analysés et revus manuellement par deux annotateurs afin d'évaluer les performances (figure 2).

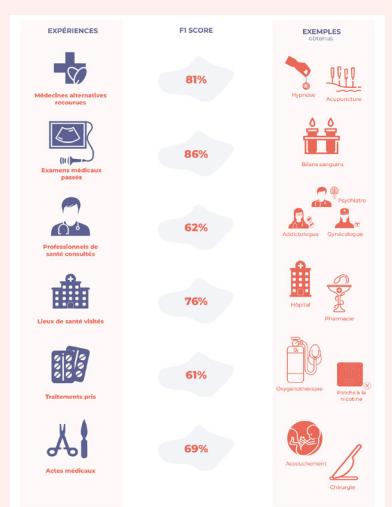


Figure 2
Performances des 6 modèles appliqués sur des messages traitant l'addictologie avec des exemples pour chaque catégorie d'expériences

CONCLUSION

L'algorithme constitué de six modèles d'apprentissage automatique permet de détecter les différentes catégories d'éléments des parcours de soins décrites par les patients ou leurs aidants, qu'elles soient vécues ou non vécues. Les performances de l'algorithme permettent une reconstruction rétrospective du parcours de soins général emprunté par une population face à une affection donnée, comme l'addiction tabagique pour le test de validation utilisé, dans un objectif de santé publique.



Comparaison des effets indésirables liés à la prise d'antidépresseurs mentionnés sur les réseaux sociaux et dans la littérature scientifique



Paul Loussikian, Anaïs Gedik, Manissa Talmakadi, Blandine Derycke, Emma Le Priol, Tom Marty, Candice Tonlorenzi, Mohamed Khemakhem, Adrien Moufid, Romain Clément, Nathalie Texier, Stéphane Schück, Adel Mebarki

Kap Code



COMPARAISON DES EFFETS INDÉSIRABLES LIÉS À LA PRISE D'ANTIDÉPRESSEURS MENTIONNÉS SUR LES RÉSEAUX SOCIAUX ET DANS LA LITTÉRATURE SCIENTIFIQUE

P.Loussikian¹, A. Gedik¹, M.Talmakadi¹, B. Derycke¹, E. Le Priol¹, T. Marty¹, C. Tonlorenzi², M. Khemakhem², A. Moufid², R. Clément², N. Texier¹, S. Schück¹, A. Mebarki¹

¹Kap Code, 146 rue Montmartre 75002 Paris, France

²ArcaScience, 130 rue de Lourmel, 75015 Paris, France

INTRODUCTION

Les antidépresseurs (ADs) sont largement utilisés pour le traitement de la dépression et d'autres troubles de l'humeur dans le monde entier et leur consommation a plus que doublée dans les pays de l'OCDE entre 2000 et 2019 [1]. Alors que l'efficacité de ces médicaments a été bien établie dans la littérature scientifique [2], la question des effets indésirables (Els) associés à leur utilisation reste un sujet d'intérêt et de préoccupation. En effet, ces Els peuvent avoir un impact significatif sur la tolérance et l'observance du traitement, ainsi que sur la qualité de vie des patients [3]. Ils sont notamment la cause, dans 66 % des cas, de l'arrêt de ces traitements [4].

Les réseaux sociaux en tant que plateformes de libre échanges d'expériences personnelles comme la prise de traitements médicamenteux notamment d'ADs [5]. Ces témoignages en ligne sont une source d'information unique sur les Els potentiels des ADs, complémentaire à la littérature scientifique traditionnelle.

Dans cette étude, les Els liés à une prise d'ADs rapportés par les patients en ligne ont été comparés à ceux mentionnés dans la littérature scientifique. L'objectif était d'évaluer la concordance et la fiabilité de ces données et d'analyser les éventuelles différences entre ces deux sources d'information afin de mieux comprendre les risques potentiels associés à ces médicaments.

MATÉRIELS & MÉTHODES

Les données, présentes dans la littérature scientifique et relatives à la prise d'ADs ainsi que les liens entre les traitements et les Els ont été extraits à l'aide de l'outil ArcalDF (Intelligence Data Federator). Cet outil repose sur l'analyse de contexte par IA et la corrélation directe entre pathologie, traitement et El. Ce modèle d'harmonisation ArcalDF a permis d'extraire ces données à partir de plus de 35 millions de références en libre accès. Parmi ces références, 50 000 articles étaient relatifs aux ADs.

Par le biais de modèles de reconnaissance sémantique nommés ArcaPV et ArcaPatho, des algorithmes de Machine Learning basés sur des ontologies telles que MedDRA, une douzaine de pathologies associées et 14 000 Els relatifs aux ADs ont été identifiés.

21 596 messages postés sur les réseaux sociaux et forums médicaux en France entre Janvier 2020 et Janvier 2022 et mentionnant des ADs ou des termes évocateurs d'ADs ont été extraits via l'outil Brandwatch®.

Après avoir passé un pipeline d'algorithme de traitement automatique du langage (TAL) parmi ce premier corpus, 8 665 messages écrits par 6 877 internautes mentionnant une expérience-patient [6] et évoquant une prise d'ADs ont été retenus. Une sélection a ensuite été effectuée par un algorithme de Machine Learning, basé et entraîné à partir du dictionnaire MedDRA, afin de détecter les messages comprenant un ou plusieurs Els liés à un ou plusieurs ADs.

Par cette méthode, 304 Els associés à l'utilisation d'ADs ont pu être identifiés en ligne.

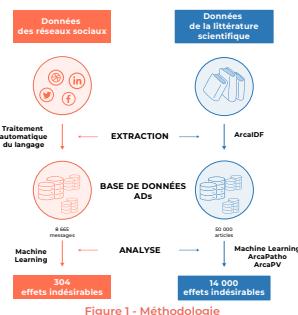


Figure 1 - Méthodologie

RÉSULTATS

4 grandes classes thérapeutiques d'antidépresseurs identifiées en ligne et dans la littérature scientifique

L'analyse des messages en ligne et de la littérature scientifique permet de distinguer 4 grandes classes thérapeutiques d'antidépresseurs avec des proportions comparables entre chaque classe (Figure 1). En effet, les inhibiteurs sélectifs de la recapture de la sérotonine (ISRS) sont les ADs les plus mentionnés (67,1% et 78,0%) ce qui semble cohérent compte tenu de leur volume de prescription sur la période d'analyse [7]. Les inhibiteurs de la recapture de la sérotonine et de la noradrénaline (ISRN), généralement prescrits en première intention comme les ISRS, sont présents, quant à eux, dans 17,1% et 20,0% respectivement. Les Imipraminiques ou tricycliques (10,9% et 1,5%) ainsi que les inhibiteurs de monoamine oxydase ou IMAO (0% et 0,5%), moins couramment utilisés, complètent le classement. Le type d'AD n'étant pas précisé dans le reste des messages en ligne.



Figure 2
Proportions de classes thérapeutiques d'antidépresseurs retrouvées en ligne et dans la littérature scientifique

La quasi-totalité des effets indésirables liés aux antidépresseurs rencontrés en ligne sont présents dans la littérature scientifique

Les principaux Els liés à des ADs identifiés dans les messages issus des réseaux sociaux et des forums de santé sont comparés avec les Els décrits dans la littérature scientifique (Figure 2).

L'état de fatigue (9,9%), la modification du poids (8,9%) et les troubles du sommeil (6,9%) sont les Els les plus fréquemment rapportés dans les messages en ligne. Concernant l'état de fatigue et les troubles du sommeil, ces Els sont également retrouvés dans la littérature scientifique, dans des proportions importantes (11,9% et 13,4% respectivement).

Toutefois, la modification du poids n'est retrouvée que dans les messages en ligne et peut suggérer que les patients en ligne accordent plus d'importance à cet effet indésirable que les études scientifiques traditionnelles.

L'anxiété, les troubles neurologiques, les troubles de l'humeur, les vertiges et les troubles de la sexualité sont des Els qui sont mentionnés à des proportions similaires en ligne et dans la littérature scientifique.

D'autres catégories d'Els comme les douleurs et les nausées/vomissements sont présents dans les deux types de sources de données mais semblent sous-représentés dans les messages en ligne (4,6% et 4,3%) par rapport aux données de la littérature scientifique (8,2% et 9,0%).

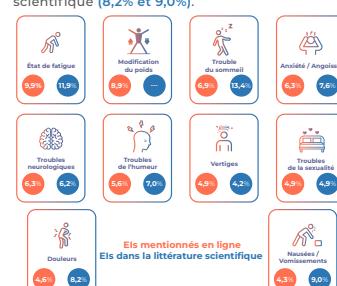


Figure 3
Proportions des Els liés à des antidépresseurs retrouvés en ligne comparés à ceux retrouvés dans la littérature scientifique

CONCLUSION

Cette étude permet de mettre en lumière l'importance d'examiner à la fois les sources en ligne et l'ensemble de la littérature scientifique pour obtenir une vision plus complète des Els associés à l'utilisation d'ADs. En effet, l'utilisation d'une méthodologie innovante combinant l'étude de ces deux types de données a mis en évidence des signaux similaires mais aussi des différences concernant les principaux Els liés à des ADs.

De plus, l'étude des réseaux sociaux et des forums de santé peut également fournir des informations uniques et spécifiques, qui peuvent ne pas être rapportées dans les études scientifiques traditionnelles. Ainsi, l'intégration de ces sources complémentaires peut enrichir la compréhension des risques potentiels associés à l'utilisation des ADs.

Des études futures pourraient continuer à développer et à optimiser ce type de méthodologie sur d'autres aires thérapeutiques afin de tester la reproductibilité et la fiabilité de cette approche.

En ce qui concerne les ADs, l'ajout de techniques de fouille de données et de visualisation de réseaux pourraient permettre d'explorer les relations entre les ADs, les Els et les caractéristiques des patients.

Ainsi cet accès à la vie réelle des patients, pourrait mettre en lumière la tolérance et l'importance accordée à certains Els, pouvant expliquer un défaut d'adhésion thérapeutique ou des switches médicamenteux.

RÉFÉRENCES

- [1] Executive summary | Health at a Glance 2021 | OECD Indicators | OECD Library [Internet]. [cité 14 mars 2023]. Disponible sur: <https://www.oecd-ilibrary.org/health-data/2019-edition-an-index.html#content/publication/da301629-en>
- [2] Cipriani A, et al. Comparative efficacy and tolerability of antidepressants for major depressive disorder in children and adolescents: a network meta-analysis. *Lancet*. 2019 Jun 16;393(10174):881–890. doi: 10.1016/j.lane.2019.05.025.
- [3] Mogabgab R, et al. Impact of adverse effects of antidepressants on medication adherence: a systematic review and meta-analysis. *J Clin Psychiatry*. 2014 Jul;75(7):e47–e47. doi: 10.4088/JCP.13b02824.
- [4] Gurevitz M, et al. Online social media and healthcare: a systematic review of quantitative and qualitative evidence. *J Med Internet Res*. 2014 Aug 7;16(8):e202. doi: 10.2196/jmir.3304.
- [5] De Bruyn A, et al. Online social media and healthcare: a systematic review of quantitative and qualitative evidence. *J Med Internet Res*. 2014 Aug 7;16(8):e202. doi: 10.2196/jmir.3304.
- [6] Lou Y, Kataoka Y, Ostlere E, Cipriani A, & Furukawa T.A. (2020). National Prescription Patterns of Antidepressants in the Treatment of Adults With Major Depression in the US Between 1996 and 2015: A Population Representative Survey Based Analysis. *Frontiers in Psychiatry*. 11. <https://doi.org/10.3389/fpsyg.2020.00035>
- [7] Lou Y, Kataoka Y, Ostlere E, Cipriani A, & Furukawa T.A. (2020). National Prescription Patterns of Antidepressants in the Treatment of Adults With Major Depression in the US Between 1996 and 2015: A Population Representative Survey Based Analysis. *Frontiers in Psychiatry*. 11. <https://doi.org/10.3389/fpsyg.2020.00035>





2023

Patient perceptions of electronic cigarettes for smoking cessation and related healthcare pathways, infodemiology study based on social media data in France



Imad Bendimerad, Tom Marty, Manissa Talmatkadi, Anaïs Gedik, Adel Mebarki, Nathalie Texier, Stéphane Schück, Ammar Amrouche, Amine Benyamina

Kap•Code



PATIENT PERCEPTIONS OF ELECTRONIC CIGARETTES FOR SMOKING CESSION AND RELATED HEALTHCARE PATHWAYS, INFODEMIOLOGY STUDY BASED ON SOCIAL MEDIA DATA IN FRANCE

Imad Bendimerad¹, Tom Marty¹, Manissa Talmatkadi¹, Anaïs Gedik¹, Adel Mebarki¹, Nathalie Texier¹, Stéphane Schück¹, Ammar Amrouche², Amine Benyamina²

¹Kap Code, 146 rue Montmartre 75002 Paris, France

²Unité de recherche psychiatrie-comorbidités-addictions - PSYCOMADD, Université Paris Saclay, France

Detect'

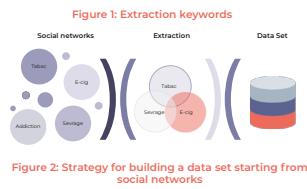
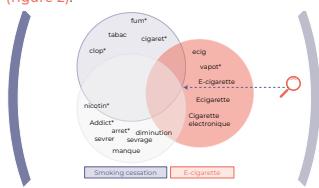
INTRODUCTION

Amid a lack of scientific evidence on the effectiveness and the risk-benefit ratio of the electronic cigarette in smoking cessation (1,2), healthcare professionals are generally supportive of patients using them, while still favorable of medical solutions as a priority treatment (3,4).

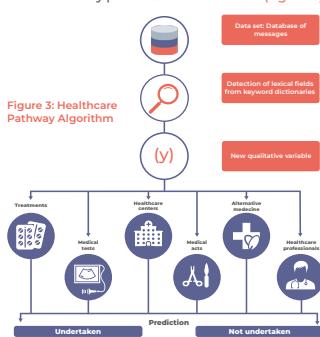
Social networks and forums are places of free expression for many patients. They describe their experiences with smoking cessation and the role of electronic cigarette in achieving this goal. These sources of real-life data constitute a valuable basis for tracing and better understanding the different healthcare pathways reported by patients in their quest to quit their addiction.

MATERIALS AND METHODS

Based on keywords referring to addiction, smoking cessation and e-cigarette use (figure 1), messages were extracted from public forums and social networks to constitute a corpus (figure 2).



The extraction included messages posted between January 2018 and February 2023, geolocated in France. Using an algorithm composed of several artificial intelligence models involving natural language processing -NLP-, different stages of healthcare pathways undertaken by patients were detected (figure 3).



RESULTS

416,530 messages mentioning smoking cessation and addiction were identified (figure 4) on several public platforms (figure 5).

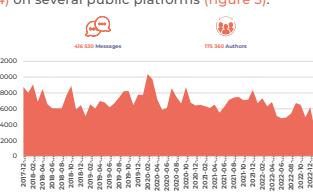


Figure 4: Evolution of monthly message volume

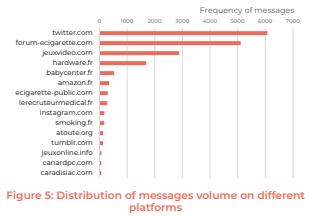


Figure 5: Distribution of messages volume on different platforms

4,865 messages from patients or their caregivers, mentioning e-cigarette use and at least one healthcare pathway category, were identified (figure 6-8).



Figure 6: Distribution of authors by category of healthcare pathway undertaken



Figure 7: Proportions of healthcare professionals consulted

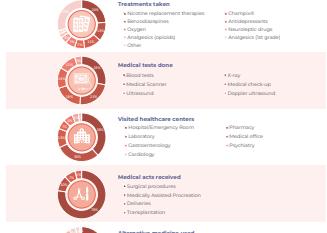


Figure 8: Proportions of other categories' items

CONCLUSION

Based on the results of these healthcare pathways, we have identified 5 contexts where healthcare professionals recommend the electronic cigarette as a means of smoking cessation: spontaneous desire to quit smoking, pregnancy, preparation for a medical procedure (e.g., surgery and dental care), the exacerbation of the addiction and psychiatric complications, and lastly medical conditions requiring cessation (e.g., cancers and cardiovascular pathologies) (figure 9).



Figure 9: Contexts where healthcare professionals recommend electronic cigarettes as a means of smoking cessation

Patients describe the electronic cigarette as their main means for smoking cessation. Other messages mention the use of medical cessation methods as well as alternative medicines (hypnosis, acupuncture) when describing the provided support, but these remain a minority compared to the electronic cigarette. Whether it is considered as a weaning tool or simply a different form of addiction, further analysis is necessary to assess its risk as a public health issue.

REFERENCES

- Berlin I. La cigarette électronique, outil thérapeutique, phénomène social ou business ? La Revue de Médecine Interne. Juin 2015;36(6):405-10.
- Gómez-García A, Martínez-Benito J, Bermejo-Bonilla M, et al. A propósito de la eficacia del uso de la tabagisme en cours de grossesse—tête court. Gynécologie Obstétrique Fertilité & Sénologie. Juill 2020;48(7-8):539-45.
- Hauszenberg B, Adler M, Carelle D, Loupries JF, Mathem G, Peiffer C, et al. Practical guidelines on e-cigarettes for practitioners and others health professionals. A French 2016 expert's statement. Revue des Maladies Respiratoires. Févr 2017;34(2):155-64.
- Kollath-Cattano C, Thrasher JF, Osman A, Andrews JO, Strayer SM. Physician Advice for e-Cigarette Use. The Journal of the American Board of Family Medicine. 1 nov 2016;29(6):741-7.





Patients' and caregivers perceptions of bariatric surgery: A France and United States comparative infodemiology study using social media data mining



Sébastien Czernichow, Nathalie Rassy, Joelle Malaab, Paul Loussikian, Adel Mebarki, Mickail Khadhar, Tigran Poghosyan, Guy Fagherrazi, Claire Carette, Stéphane Schück, Claire Rives-Lange



Frontiers in Digital Health

TYPE Original Research
PUBLISHED 18 April 2023
DOI 10.3389/fdgth.2023.1136326



OPEN ACCESS

EDITED BY

Alex Mariakakis,
University of Toronto, Canada

REVIEWED BY

Daniel Gero,
University Hospital Zurich, Switzerland
Artur Strzelecki,
University of Economics of Katowice, Poland

*CORRESPONDENCE

Sébastien Czernichow
✉ [sebastien.czernichow@aphp.fr](mailto:sbastien.czernichow@aphp.fr)

[†]These authors share first authorship

SPECIALTY SECTION

This article was submitted to Digital Health Communication, a section of the journal Frontiers in Digital Health

RECEIVED 02 January 2023

ACCEPTED 21 March 2023

PUBLISHED 18 April 2023

CITATION

Czernichow S, Rassy N, Malaab J, Loussikian P, Mebarki A, Khadhar M, Poghosyan T, Fagherrazi G, Carette C, Schück S and Rives-Lange C (2023) Patients' and caregivers' perceptions of bariatric surgery: A France and United States comparative infodemiology study using social media data mining. *Front. Digit. Health* 5:1136326. doi: 10.3389/fdgth.2023.1136326

COPYRIGHT

© 2023 Czernichow, Rassy, Malaab, Loussikian, Mebarki, Khadhar, Poghosyan, Fagherrazi, Carette, Schück and Rives-Lange. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Patients' and caregivers' perceptions of bariatric surgery: A France and United States comparative infodemiology study using social media data mining

Sébastien Czernichow^{1,2,3*†}, Nathalie Rassy^{1†}, Joelle Malaab⁴,
Paul Loussikian⁴, Adel Mebarki⁴, Mickail Khadhar⁴,
Tigran Poghosyan⁵, Guy Fagherrazi⁶, Claire Carette^{1,2,7},
Stéphane Schück⁴ and Claire Rives-Lange^{1,2,3}

¹Assistance Publique-Hôpitaux de Paris (AP-HP), Service de Nutrition, Centre Spécialisé Obésité, Hôpital Européen Georges Pompidou, Paris, France, ²Université Paris Cité, Paris, France, ³INSERM, UMR1153, Epidemiology and Biostatistics Sorbonne Paris Cité Center (CRESS), METHODS Team, Paris, France, ⁴Kap Code, Paris, France, ⁵Assistance Publique-Hôpitaux de Paris (AP-HP), Service de chirurgie digestive, Hôpital Bichat, Paris, France, ⁶Deep Digital Phenotyping Research Unit, Department of Precision Health, Luxembourg Institute of Health, Strassen, Luxembourg, ⁷Assistance Publique-hôpitaux de Paris (AP-HP), Centre d'investigation clinique, Inserm 1418, Hôpital Européen Georges Pompidou, Paris, France

Background: People are conversing about bariatric surgery on social media, but little is known about the main themes being discussed.

Objective: To analyze discussions regarding bariatric surgery on social media platforms and to establish a cross-cultural comparison of posts geolocated in France and the United States.

Methods: Posts were retrieved between January 2015 and April 2021 from general, publicly accessed sites and health-related forums geolocated in both countries. After processing and cleaning the data, posts of patients and caregivers about bariatric surgery were identified using a supervised machine learning algorithm.

Results: The analysis dataset contained a total of 10,800 posts from 4,947 web users in France and 51,804 posts from 40,278 web users in the United States. In France, post-operative follow-up ($n = 3,251$, 30.1% of posts), healthcare pathways ($n = 2,171$, 20.1% of the posts), and complementary and alternative weight loss therapies ($n = 1,652$, 15.3% of the posts) were among the most discussed topics. In the United States, the experience with bariatric surgery ($n = 11,138$, 21.5% of the posts) and the role of physical activity and diet in weight-loss programs before surgery ($n = 9,325$, 18% of the posts) were among the most discussed topics.

Conclusion: Social media analysis provides a valuable toolset for clinicians to help them increase patient-centered care by integrating the patients' and caregivers' needs and concerns into the management of bariatric surgery.

KEYWORDS

bariatric (weight loss) surgery, social media, patients, caregiver, perception, infodemiology



Patient and Caregiver Perceptions of Advanced Bladder Cancer Systemic Treatments: Infodemiology Study Based on Social Media Data



Simon Renner, Paul Loussikian, Pierre Foulquié, Alexia Marrel, Valentin Barbier, Adel Mebarki, Stéphane Schück, Murtuza Bharmal

JMIR CANCER

Renner et al

Original Paper

Patient and Caregiver Perceptions of Advanced Bladder Cancer Systemic Treatments: Infodemiology Study Based on Social Media Data

Simon Renner^{1*}, PharmD; Paul Loussikian^{1*}, MS; Pierre Foulquié¹, MSc; Alexia Marrel², MA; Valentin Barbier², MSc; Adel Mebarki¹, MSc; Stéphane Schück¹, PhD; Murtuza Bharmal³, BPharm, MS, PhD

¹Kap Code, Paris, France

²ICON, Lyon, France

³EMD Serono, Rockland, MA, United States

*these authors contributed equally

Corresponding Author:

Murtuza Bharmal, BPharm, MS, PhD

EMD Serono

1 Technology Place

Rockland, MA, 02370

United States

Phone: 1 781 681 2051

Email: murtuza.bharmal@emdserono.com

Abstract

Background: In 2022, it was estimated that more than 80,000 new cases of bladder cancer (BC) were diagnosed in the United States, 12% of which were locally advanced or metastatic BC (advanced BC). These forms of cancer are aggressive and have a poor prognosis, with a 5-year survival rate of 7.7% for metastatic BC. Despite recent therapeutic advances for advanced BC, little is known about patient and caregiver perceptions of different systemic treatments. To further explore this topic, social media can be used to collect the perceptions of patients and caregivers when they discuss their experiences on forums and online communities.

Objective: The aim of this study was to assess patient and caregiver perceptions of chemotherapy and immunotherapy for treating advanced BC from social media–posted data.

Methods: Public posts on social media in the United States between January 2015 and April 2021 from patients with advanced BC and their caregivers were collected. The posts included in this analysis were geolocalized to the United States; collected from publicly available domains and sites, including social media sites such as Twitter and forums such as patient association forums; and were written in English. Posts mentioning any line of chemotherapy or immunotherapy were qualitatively analyzed by two researchers to classify perceptions of treatments (positive, negative, mixed, or without perception).

Results: A total of 80 posts by 69 patients and 142 posts by 127 caregivers mentioning chemotherapy, and 42 posts by 31 patients and 35 posts by 32 caregivers mentioning immunotherapy were included for analysis. These posts were retrieved from 39 public social media sites. Among patients with advanced BC and their caregivers, treatment perceptions of chemotherapy were more negative (36%) than positive (7%). Most of the patients' posts (71%) mentioned chemotherapy factually without expressing a perception of the treatment. The caregivers' perceptions of treatment were negative in 44%, mixed in 8%, and positive in 7% of posts. In combined patient and caregiver posts, immunotherapy was perceived positively in 47% of posts and negatively in 22% of posts. Caregivers also posted more negative perceptions (37%) of immunotherapy than patients (9%). Negative perceptions of both chemotherapy and immunotherapy were mainly due to side effects and perceived lack of effectiveness.

Conclusions: Despite chemotherapy being standard first-line therapy for advanced BC, negative perceptions were identified on social media, particularly among caregivers. Addressing these negative perceptions of treatment may improve treatment adoption. Strengthening support for patients receiving chemotherapy and their caregivers to help them manage side effects and understand the role of chemotherapy in the treatment of advanced BC would potentially enable a more positive experience.

(*JMIR Cancer* 2023;9:e45011) doi: [10.2196/45011](https://doi.org/10.2196/45011)



Perceived Unmet Needs in Patients Living With Advanced Bladder Cancer and Their Caregivers: Infodemiology Study Using Data From Social Media in the United States



Simon Renner, Paul Loussikian, Pierre Foulquié, Benoit Arnoult, Alexia Marrel, Valentin Barbier, Adel Mebarki, Stéphane Schück, Murtuza Bharmal

JMIR CANCER

Renner et al

Original Paper

Perceived Unmet Needs in Patients Living With Advanced Bladder Cancer and Their Caregivers: Infodemiology Study Using Data From Social Media in the United States

Simon Renner^{1*}, PharmD; Paul Loussikian^{1*}, MS; Pierre Foulquié¹, MSc; Benoit Arnould², PhD; Alexia Marrel², MA; Valentin Barbier², MSc; Adel Mebarki¹, MSc; Stéphane Schück¹, PhD; Murtuza Bharmal³, PhD

¹Kap Code, Paris, France

²Icon, Lyon, France

³EMD Serono, Billerica, MA, United States

*these authors contributed equally

Corresponding Author:

Murtuza Bharmal, PhD

EMD Serono

45 Middlesex Turnpike

Billerica, MA, 01821

United States

Phone: 1 978 294 1100

Email: murtuza.bharmal@emdserono.com

Abstract

Background: Locally advanced or metastatic bladder cancer (BC), which is generally termed advanced BC (aBC), has a very poor prognosis, and in addition to its physical symptoms, it is associated with emotional and social challenges. However, few studies have assessed the unmet needs and burden of aBC from patient and caregiver perspectives. Infodemiology, that is, epidemiology based on internet health-related content, can help obtain more insights on patients' and caregivers' experiences with aBC.

Objective: The study aimed to identify the main discussion themes and the unmet needs of patients with aBC and their caregivers through a mixed methods analysis of social media posts.

Methods: Social media posts were collected between January 2015 and April 2021 from US geolocalized sites using specific keywords for aBC. Automatic natural language processing (regular expressions and machine learning) methods were used to filter out irrelevant content and identify verbatim posts from patients and caregivers. The verbatim posts were analyzed to identify main discussion themes using bitemporal topic modeling. Difficulties or unmet needs were further explored using qualitative research methods by 2 independent annotators until saturation of concepts.

Results: A total of 688 posts from 262 patients and 1214 posts from 679 caregivers discussing aBC were identified. Analysis of 340 randomly selected patient posts and 423 randomly selected caregiver posts uncovered 33 unique unmet need categories among patients and 36 among caregivers. The main unmet patient needs were related to challenges regarding adverse events (AEs; 28/95, 29%) and the psychological impact of aBC (20/95, 21%). Other patient unmet needs identified were prognosis or diagnosis errors (9/95, 9%) and the need for better management of aBC symptoms (9/95, 9%). The main unmet caregiver needs were related to the psychological impacts of aBC (46/177, 26.0%), the need for support groups and to share experiences between peers (28/177, 15.8%), and the fear and management of patient AEs (22/177, 12.4%).

Conclusions: The combination of manual and automatic methods allowed the extraction and analysis of several hundreds of social media posts from patients with aBC and their caregivers. The results highlighted the emotional burden of cancer for both patients and caregivers. Additional studies on patients with aBC and their caregivers are required to quantitatively explore the impact of this disease on quality of life.

(*JMIR Cancer* 2022;8(3):e37518) doi: [10.2196/37518](https://doi.org/10.2196/37518)

KEYWORDS

real-world evidence; unmet needs; quality of life; social media; bladder cancer; caregivers



Psoriasis Et Altérations De La Qualité De Vie Au Travail Une Étude À Partir De Deux Sources De Données De Vie Réelle Consultations Médicales Croisées Avec Les Réseaux Sociaux



Caroline Eteve-Pitsaer, Tom Marty, Angie Nguyen, Emma Le Priol, Adel Mebarki, Nathalie Texier, Stephane Schück

Kap Code

cegedim
Health Data

THIN
THE HEALTH IMPROVEMENT NETWORK

PSORIASIS ET ALTÉRATIONS DE LA QUALITÉ DE VIE AU TRAVAIL : UNE ÉTUDE DE DEUX SOURCES DE DONNÉES DE VIE RÉELLE, CONSULTATIONS MÉDICALES CROISÉES AVEC LES RÉSEAUX SOCIAUX

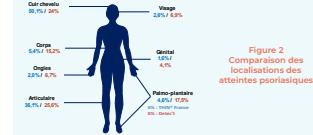
Caroline Eteve-Pitsaer¹, Tom Marty², Angie Nguyen¹, Emma Le Priol², Claire Paris¹, Adel Mebarki², Nathalie Texier², Stéphane Schück²

¹Gér SAS-Cegedim Health Data, Boulogne-Billancourt, France

²Kap Code, 146 rue Montmartre, 75002 Paris, France

INTRODUCTION

Le psoriasis est une maladie chronique inflammatoire de la peau touchant 2,5 millions de personnes en France [1]. Si la littérature montre des liens de plus en plus perceptibles entre psoriasis et troubles psychologiques[3], l'analyse des conséquences de ces comorbidités sur la vie quotidienne des patients a pu difficilement être conduite à grande échelle, puisqu'elle requiert la plupart du temps le recrutement de patients répondant à un questionnaire structure[2,3]. Ainsi, l'objectif de cette étude était d'examiner, à grande échelle, les altérations de la qualité de vie au travail de patients atteints de psoriasis, en croisant des données de vie réelle dont certaines issues des réseaux sociaux.



La répartition des psoriasis est relativement concordante entre les 2 sources de données, où le cuir chevelu et les articulations sont les atteintes les plus importantes. Suivies des atteintes du « corps » et du psoriasis palmo-plantaire. Les 3 dernières suivent le même classement, avec l'atteinte du visage, puis des ongles et enfin des psoriasis génitaux.

MATÉRIEL ET MÉTHODES

Une cohorte de 52 492 patients a été extraite de la base de données médicalisée, ambulatoire, en vie réelle THIN® France, conforme au RCPD, intégrant l'historique de remboursement de la CNAM. Tous les patients associés à un psoriasis diagnostiquée (CIM 10 : L40*) entre Janvier 2019 et Juin 2021 ont été sélectionnés. De plus, une cohorte contrôlée, appariée sur l'âge, le sexe, et la région, a été constituée pour mener une étude comparative.



Les messages relatifs au psoriasis, postés sur les réseaux sociaux francophones accessibles de Janvier 2019 à Juin 2021, ont été collectés via l'outil Brandwatch®, conforme au RCPD. Un algorithme de traitement automatisé du langage a permis de sélectionner les messages développant une expérience médicale vécue, écrits par les patients et les proches, composant un corpus d'analyse de 2 383 messages (2 036 internautes).

RÉSULTATS

LES DEUX BASES DE DONNÉES SONT COMPARABLES, MALGRÉ LES DIFFÉRENCES DE POPULATION

Place des antidépresseurs & anxiolytiques

Les antidépresseurs et anxiolytiques sont présents dans le top 10 des traitements détectés dans les deux sources de données, ce qui corrobore les conclusions d'études précédentes mettant en exergue le lien entre psoriasis et troubles psychologiques [3]. Sur les réseaux sociaux (Detect), ils sont en 5ème position des traitements les plus évoqués. Dans les données médicales, ils sont en 8ème position des traitements prescrits aux patients ayant un arrêt de travail (THIN).

Comparaison des atteintes psoriasiques selon leur fréquence

La figure 2 montre la répartition des localisations des atteintes psoriasiques lorsqu'elles sont renseignées (27,7%; 21%), disponibles dans le tableau 1.



L'ANALYSE DES ARRÊTS DE TRAVAIL (THIN®) ET DES RÉSEAUX SOCIAUX (KAP CODE) MONTRÉE L'IMPACT DU PSORIASIS SUR LA SANTÉ MENTALE ET L'ACTIVITÉ PROFESSIONNELLE

La comparaison entre la cohorte de patients atteints de psoriasis et la cohorte contrôle dans les données médicales (THIN®) montre une probabilité plus forte d'avoir un arrêt de travail avec une raison liée à la santé mentale (ICD10 F30-F39 et F40-F48) chez les patients psoriasiques. La différence est particulièrement forte pour les troubles de l'humeur (F30-F39), avec un odds ratio égal à 1,43. De plus, on observe une durée moyenne d'arrêt de travail plus longue dans la cohorte d'étude, 17,6 jours, contre 14,9 jours dans la cohorte contrôle (figure 3).



Plus de 50% des internautes ayant écrit un message lié aux arrêts de travail, évoquent un rhumatisme psoriasique. Ils témoignent notamment des plaintes concernant la gêne et la douleur ainsi qu'au rhumatisme/arthrite en lui-même.

La figure 5 montre les 194 mots les plus fréquents dans les concepts médicaux détectés sur les réseaux sociaux, ainsi que dans les diagnostics associés aux arrêts de travail. La taille des mots est représentative du nombre d'occurrences total des mots tandis que l'échelle de couleur reflète leur fréquence relative dans chaque source de données (bleu : THIN FR, Rouge : Kap Code réseaux sociaux, échelle du rouge au bleu selon la fréquence relative des mots communs aux deux bases de données).

Les douleurs (« douleur », « douleurs ») sont particulièrement rapportées sur les réseaux sociaux. La souffrance psychologique et le thème du travail sont également mis en lumière avec les mots-clés « travail », « anxiété », et « malaise » en violet, indiquant leur présence dans les deux bases de données. Tandis que les motifs d'arrêt maladie soulignent l'importance des troubles psychologiques et de la santé mentale (« dépression », « burn-out » en bleu), la fusion de ces données avec les messages publiés par des patients permet de qualifier plus finement le ressenti de ces derniers (« aidez-moi », « horre », « complexé », « peur » en rouge).



Figure 5: Motifs d'arrêts de travail (source THIN®) associés à des messages qualitatifs - Source Kap Code

CONCLUSION

Cette étude a apporté, à travers la synergie de 2 sources de données de vie réelle, quantitatives structurées de la base de données THIN® et qualitatives non structurées des réseaux sociaux, un nouvel éclairage concernant la nature des souffrances des patients atteints de psoriasis, qui se retrouvent principalement dans la dépression et les douleurs, particulièrement en lien avec leur vie professionnelle.

Les principales limites de cette étude sont les suivantes : (i) les publications sur les réseaux sociaux concernent une partie réduite de la population atteinte de psoriasis (dans cette étude, les femmes sont majoritaires à 70%, âge moyen de 32,9 ans) et sont soumises à certains biais ; (ii) la période d'étude coïncide en grande partie avec celle de la pandémie de COVID-19, qui s'est accompagnée d'une forte augmentation générale des troubles psychologiques[4]. Ainsi, les travaux futurs porteront sur l'extension de cette analyse sur une période plus large ; et l'analyse des relations temporelles entre troubles psychologiques et crises psoriasiques. Au-delà des résultats, ce travail illustre la valeur ajoutée que le croisement de deux sources de données de vie réelle - structurée et non structurée - peut représenter pour les décideurs en santé, dans un objectif de compréhension fine de la réalité patient.

RÉFÉRENCES

- [1] https://francescp.scp.org/uploads/2016/08/DOSSIER_PRESSE_JOURNÉE_MONDIALE_PSORIASIS_RHUMATISME_PSORIASIQUE_2014.pdf
- [2] Francesca Sampogna, Francesco Sera, Damiano Abeni. Measures of Clinical Severity, Quality of Life, and Psychological Distress in Patients with Psoriasis: A Cluster Analysis. Journal of Investigative Dermatology. <https://doi.org/10.1016/j.jid.2022.03.010>.
- [3] Raishan S. Bakar, Shafiqah Z.S. Jaspal, Afra F. Asmi, Yeoh C. Aun, Depression and anxiety among patients with psoriasis: A correlation with quality of life and associated factors. Journal of Taibah University Medical Sciences, Volume 16, Issue 4, 2022. <https://doi.org/10.1016/j.jtumed.2021.02.008>
- [4] Résultats de l'étude CovPrev : une enquête pour suivre l'évolution des comportements et de la santé mentale pendant l'épidémie de COVID-19 en France

Les Difficultés Exprimées Par Les Patients Atteints De Dermatoses Visibles Sur Les Réseaux Sociaux



Tom Marty, Mickail Khadar, Lina Autelino, Catherine Baissac, Simon Renner, Adel Mebarki, Nathalie Texier, Stephane Schück

Kap Code

LES DIFFICULTÉS EXPRIMÉES PAR LES PATIENTS ATTEINTS DE DERMATOSES VISIBLES SUR LES RÉSEAUX SOCIAUX

Tom Marty¹, Mickail Khadar¹, Lina Autelino², Catherine Baissac², Simon Renner¹, Adel Mebarki¹, Nathalie Texier¹, Stéphane Schück¹
¹Kap Code, 146 rue Montmartre, 75002 Paris, France
²Pierre Fabre, 3 avenue Hubert Curien, 31100 Toulouse, France

INTRODUCTION

Les dermatoses visibles comme l'eczème, l'acné, le psoriasis, le vitiligo ou la rosacée peuvent considérablement impacter la qualité de vie des patients, notamment sur le plan psycho-social. Les patients concernés et leurs proches peuvent se tourner vers les réseaux sociaux et forums pour échanger entre eux. Ces échanges constituent des données de vie réelle qui peuvent être analysées pour mieux comprendre l'expérience vécue par les patients, l'impact de leur maladie et les difficultés rencontrées au quotidien.

MATÉRIEL ET MÉTHODES

Les messages écrits en ligne sur les forums et réseaux sociaux publics entre septembre 2018 à septembre 2021 ont été récupérés avec l'outil Brandwatch®. L'identification des messages pertinents s'est basée sur la présence de mots clés en lien avec les cinq dermatoses étudiées : le psoriasis, l'eczème, l'acné, le vitiligo et la rosacée. Les messages témoignant d'une expérience vécue de patient ou de proche ont été sélectionnés grâce à un filtre de recherche spécifique. Ensuite, les messages, issus de 2082 messages par dermatose, aléatoirement sélectionnés, ont ensuite été analysés manuellement selon les difficultés rencontrées par les patients et leurs proches. L'analyse s'est basée sur une grille d'annotation standardisée présentant 51 possibles difficultés différentes. Pour chaque dermatose, les difficultés les plus souvent associées au sein des messages ont ensuite été identifiées.

RÉSULTATS

Sur 2082 messages récoltés, les maladies les plus citées sont l'acné (63,1%), l'eczème (22,3%), le psoriasis (9,6%), le vitiligo (3,8%) et la rosacée (1,2%). Les messages proviennent de différents forums et réseaux sociaux (fig. 1). Twitter est la source principale. Les messages y sont courts, les patients et leurs proches se répondent en mentionnant leurs malades et leurs symptômes. Ensuite, les patients et leurs proches évoquent leurs problèmes dermatologiques d'une célébrité. YouTube est la 2ème source, les messages sont retrouvés au sein des commentaires de vidéos de tutoriels et tests de produits dermatologiques. YouTube est particulièrement important chez les patients atteints de rosacée car les internautes sont à la recherche de retours d'expérience sur certains produits (cosmétiques et dermatologiques). La communauté issue de Docitissimo détaille le fardeau de la maladie, les patients s'y retrouvent pour avoir des réponses à leurs questions et trouver des « groupes de parole et de soutien ». BabyCenter est une source importante des messages en lien avec l'eczème à cause de l'eczème du nourrisson qui y est abordé par les mères. Le psoriasis est une maladie qui bénéficie de canaux spécifiques sur Carenity, ce qui explique sa présence importante par rapport aux autres maladies.

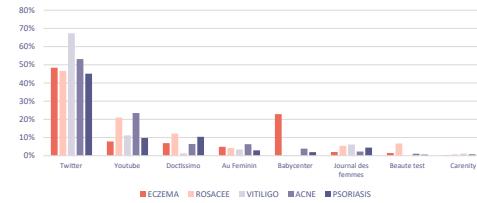


Figure 1 - Répartition des messages dans le classement des tops sources

Les verbatims utilisés pour témoigner des impacts sont : allergie (eczème), rougeur (rosacée), peur (vitiligo), fatigue et peur (acné), fatigue et douleur (psoriasis). Les impacts de la maladie sont majoritairement physiques (56,6%) suivis de l'impact psychique (26,6%). Toutes dermatoses confondues, les difficultés les plus rencontrées sont « crainte et gestion des symptômes de la maladie » (n=94), suivie de « l'impact psychique au sens général » (n=76) puis « l'atteinte de l'image de soi » (n=67). L'analyse des co-occurrences (fig. 2) révèle l'existence de liens entre les difficultés rencontrées qui sont associées. Ainsi, dans le psoriasis, l'impact psychologique est lié à la crainte des poussées et de la progression de la maladie. Dans l'acné, comme dans le vitiligo, l'impact psychologique est lié à la dégradation de l'image de soi. Pour l'eczème, l'impact s'étend aux proches.

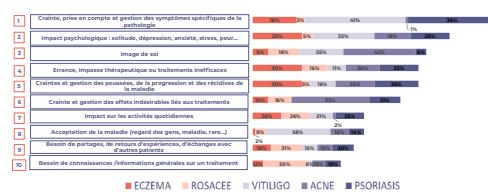
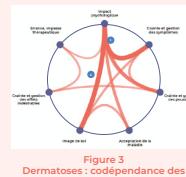


Figure 2 - Top 10 des difficultés les plus mentionnées. Répartition des difficultés par maladies.

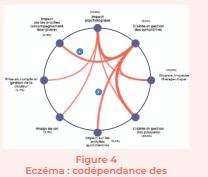
Les patients et leurs proches peuvent mentionner plusieurs difficultés de manière conjointe dans les messages. L'analyse des liens et détecter les co-occurrences des difficultés peuvent permettre de mieux appréhender les interactions couvertes des patients et de leur aidants. Toutes dermatoses confondues (fig. 3), l'impact psychologique est particulièrement lié aux symptômes des maladies affichantes et à la difficulté de leur gestion (lien 1), notamment lorsqu'elles évoluent par poussées. L'impact physique a un retentissement psychique qui s'ajoute à l'image de soi perturbée (lien 2). En trame de fond, l'acceptation de la maladie, l'errance thérapeutique et la crainte des effets indésirables des traitements contribuent à la situation de mal-être exprimée par les patients.

L'eczème (fig. 4) est la seule maladie pour laquelle l'impact s'étend aux proches.

Energivore, source d'inquiétude et de questionnements sur la gestion des symptômes et des habitudes de vie à adapter (lien 1). De même, les impacts sur les activités quotidiennes concernent la façon dont les symptômes perturbent les nourrissons au niveau du sommeil ou au niveau de l'alimentation qu'il faut modifier.



Dermatoses : codépendance des difficultés rencontrées

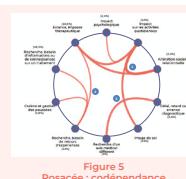


Eczème : codépendance des difficultés rencontrées

L'anxiété sociale (fig. 5) due aux rougessements pousse des patients à moduler certaines activités afin d'éviter de se retrouver dans des situations générantes (lien 1). Le besoin de retours d'expériences est causé par l'errance thérapeutique et le fait que chaque patient réagisse différemment aux mêmes produits (lien 2). Des messages évoquent le besoin de reconsulter en quête d'un autre avis médical, évoquant l'errance diagnostique associée à la rosacée. Les messages évoquent l'absence de reconnaissance de la rosacée par le professionnel de santé.

Dans le vitiligo (fig. 6), l'impact psychologique est directement lié aux symptômes qui perturbent l'image de soi, sur une trame de fond d'une difficulté d'acceptation de la maladie, qui s'ajoute au poids du regard des autres (lien 1).

La dépigmentation est le principal symptôme évoqué par les patients. Ils décrivent une crainte de la progression des zones dépigmentées et de l'aggravation du caractère « affichant » de la maladie (lien 2).



Rosacée : codépendance des difficultés rencontrées



Vitiligo : codépendance des difficultés rencontrées

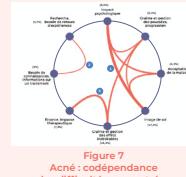
L'image de soi (fig. 7) est surtout liée à l'impact psychologique dans l'acné, en cause les boutons présents en permanence et s'aggravant lors des poussées (lien 1). Le besoin de retours d'expériences est lié à la recherche d'informations sur les traitements possibles. Les patients évoquent le besoin de reconnaître ce besoin, où la balance bénéfices/risques est au cœur des interrogations (lien 2).

La difficulté de l'errance thérapeutique est fréquemment mentionnée avec la crainte des effets indésirables. Après avoir testé divers produits, les patients en errance envisagent des produits « plus forts » mais recherchent de retours d'expériences (lien 3).

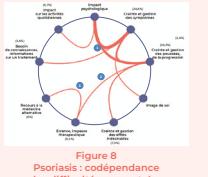
Les patients experimentent le fardeau du cercle vicieux stress & poussées de psoriasis (fig. 8) : le stress source des crises qui alimentent un état de mal-être. La prise en compte de la dimension psychique étant un sujet en soi pour ces patients qui souhaitent mettre un terme à ce cercle vicieux (lien 1).

Concernant l'errance thérapeutique, l'inefficacité des traitements est liée à la crainte et à la gestion des poussées. Les biothérapies sont décrites ici comme une solution est une liberté retrouvée (lien 2).

Le besoin de recourir aux médecines alternatives, notamment aux magnétiseurs, est à la fois envisagé et décrié par les patients face à l'errance thérapeutique (lien 3).



Acné : codépendance des difficultés rencontrées



Psoriasis : codépendance des difficultés rencontrées

CONCLUSION

Sur 2082 messages récoltés, les maladies les plus citées sont l'acné (63,1%), l'eczème (22,3%), le psoriasis (9,6%), le vitiligo (3,8%) et la rosacée (1,2%). Les verbatims utilisés pour témoigner des impacts sont : allergie (eczème), rougeur (rosacée), peur (vitiligo), fatigue et peur (acné), fatigue et douleur (psoriasis). Les impacts de la maladie sont majoritairement physiques (56,6%) suivis de l'impact psychique (26,6%). Toutes dermatoses confondues, les difficultés les plus rencontrées sont « crainte et gestion des symptômes de la maladie » (n=94), suivie de « l'impact psychique au sens général » (n=76) puis « l'atteinte de l'image de soi » (n=67). L'analyse des co-occurrences (fig. 1) révèle l'existence de liens entre les difficultés rencontrées qui sont associées. Ainsi, dans le psoriasis, l'impact psychologique est lié à la crainte des poussées et de la progression de la maladie. Dans l'acné, comme dans le vitiligo, l'impact psychologique est lié à la dégradation de l'image de soi. Pour l'eczème, l'impact s'étend aux proches.

Création D'un Algorithme D'identification D'expériences Vécues Par Des Patients ou Leurs Proches À Partir De Messages Issus Des Réseaux Sociaux : Un Cas D'usage Sur Le Covid Long



Manissa Tamaltkadi, Pierre Foulquié, Amélia Déguilhem, Simon Renner, Léa Chateauneuf, Adel Mebarki, Nathalie Texier, Stéphane Schück

Kap•Code

CONCEPTION D'UN ALGORITHME D'IDENTIFICATION D'EXPÉRIENCES VÉCUES PAR DES PATIENTS OU LEURS PROCHES À PARTIR DE MESSAGES ISSUS DES RÉSEAUX SOCIAUX : UN CAS D'USAGE SUR LE COVID LONG

Manissa Tamaltkadi¹, Pierre Foulquié¹, Amélia Déguilhem¹, Simon Renner¹, Léa Chateauneuf¹, Adel Mebarki¹, Nathalie Texier¹, Stéphane Schück¹

¹Kap Code, 146 rue Montmartre 75002 Paris, France

Detect'

INTRODUCTION

La pandémie de COVID19 a entraîné une masse d'informations sur les réseaux sociaux et forums provenant de multiples acteurs : gouvernement, journalistes ou encore citoyens.

Parmi ces messages, de nombreux patients se sont saisis de ces outils pour se regrouper en communauté et partager leurs symptômes, comme c'est le cas des patients atteints de COVID long, avec le hashtag #aprèsJ20.

Un algorithme de traitement automatique du langage naturel capable d'identifier les messages constituant des expériences de santé vécues par des patients ou des proches permettrait d'exploiter ces données de vie réelle dans un objectif de santé.

MATÉRIEL ET MÉTHODES

Un corpus de 12 430 messages issus de différents forums et réseaux sociaux a été extrait sur diverses pathologies : Cancer, Psoriasis, Maladie de Fabry, Sevrage tabagique, Diabète et Vaccination. Une annotations en 3 catégories a été réalisée :

- **Patient**: une personne malade ou prenant un traitement.
- **Proche** : un proche s'occupant d'un patient ou ayant un lien avec lui (aussi appelé Aidant).
- **Répondant** : un utilisateur qui répond sans préciser son lien avec une maladie un traitement, un professionnel de santé ou autre.

Ces messages annotés ont ainsi constitué un gold standard (GS) comportant 6 725 expériences médicales vécues (4 730 Patients et 1 995 Proches) et 5 605 expériences non vécues (Répondants).

Le GS a par la suite été réparti en deux jeux de données pour validation croisée. Le premier, contenant 70% des messages, est utilisé pour entraîner le modèle, et le second, avec 30% des messages, sert à la validation de la méthode.

Par la suite, deux modèles binaires Xgboost ont été créées et entraînés. Le premier détermine dans un premier temps, si le message a été posté par un proche du patient ou non, puis dans un second temps si c'est un patient ou non. Ces deux catégories (Patients et Proches) sont ensuite regroupées pour constituer les expériences médicales vécues (figure 1).

Le calcul au niveau du message de la probabilité d'appartenir à la classe des expériences vécues est fait en se basant sur :

- Des formes régulières (Regex) (« je suis atteint de [TERME d'INTERET] », « mon père a été diagnostiquée [TERME d'INTERET] »)
- Couplées à des variables sur les pronoms

Pour finir, afin de valider ces modèles, ces derniers ont été appliqués à un échantillon de messages spécifiques au COVID long, sélectionnés aléatoirement et annotés par 3 annotateurs, ce qui a permis d'évaluer les performances de cet algorithme.

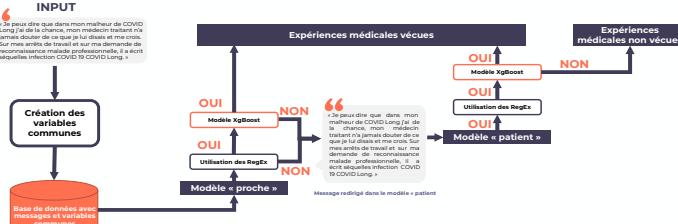


Figure 1 - Algorithme Patient-Aidant

RÉSULTATS

Après l'application de l'algorithme d'identification d'expériences vécues par des patients ou des proches sur des messages relatifs au COVID long, une revue manuelle de 700 messages a été réalisée en premier lieu, puis en second lieu, un regroupement des catégories Patient et Aidant a été fait, cela permet de rassembler les nombreux cas de formes familiales du COVID long retrouvés sur les réseaux sociaux, qui réunissent à la fois un message de Proche et de Patient dans une seule et même classe, celle des expériences médicales vécues (figure 2).

Accuracy	87%
F1 score	90%
Sensibilité	96%
Spécificité	76%
Précision	84%

Figure 2
Performances du modèle sur les expériences médicales vécues du corpus COVID long

CONCLUSION

La création de cet algorithme capable d'identifier les expériences vécues par des Patients ou leurs Proches à partir des messages issus des réseaux sociaux facilite leur utilisation comme données de vie réelle, notamment grâce aux bonnes performances de ces modèles. Le suivi de ces populations de Patients et de leurs Proches dans un objectif de santé publique est possible.

Il convient tout de même de souligner que la présente étude comporte une limite non négligeable, en effet, d'une part, l'identification des messages comprenant à la fois une expérience d'un patient et celle d'un de ses proches et d'autre part, la catégorisation des proches (cercle familial ou externe) devront être possible dans une prochaine version de l'algorithme.



Concerns Discussed on Chinese and French Social Media During the COVID-19 Lockdown: Comparative Infodemiology Study Based on Topic Modeling



Stéphane Schück, Pierre Foulquié, Adel Mebarki, Carole Faviez, Mickaïl Khadhar, Nathalie Texier, Sandrine Katsahian, Anita Burgun, Xiaoyi Che

JMIR FORMATIVE RESEARCH

Schück et al

Original Paper

Concerns Discussed on Chinese and French Social Media During the COVID-19 Lockdown: Comparative Infodemiology Study Based on Topic Modeling

Stéphane Schück¹, MD; Pierre Foulquié¹, MSc; Adel Mebarki¹, MSc; Carole Faviez², MSc; Mickaïl Khadhar¹, MSc; Nathalie Texier¹, PharmD; Sandrine Katsahian^{2,3}, PhD, MD; Anita Burgun^{2,4,5,6}, PhD, MD; Xiaoyi Chen², PhD

¹Kap Code, Paris, France

²Centre de Recherche des Cordeliers, INSERM, Sorbonne Université, Université de Paris, Paris, France

³Unité d'Épidémiologie et de Recherche Clinique, Hôpital européen Georges Pompidou, Assistance Publique - Hôpitaux de Paris, Paris, France

⁴Département d'informatique médicale, Hôpital européen Georges Pompidou, Assistance Publique - Hôpitaux de Paris, Paris, France

⁵Département d'informatique médicale, Hôpital Necker-Enfants Malades, Assistance Publique - Hôpitaux de Paris, Paris, France

⁶Paris Artificial Intelligence Research Institute, Paris, France

Corresponding Author:

Xiaoyi Chen, PhD

Centre de Recherche des Cordeliers

INSERM

Sorbonne Université, Université de Paris

15 Rue de l'école de médecine

Paris, F-75006

France

Phone: 33 171196369

Email: xiaoyi.chen@inserm.fr

Abstract

Background: During the COVID-19 pandemic, numerous countries, including China and France, have implemented lockdown measures that have been effective in controlling the epidemic. However, little is known about the impact of these measures on the population as expressed on social media from different cultural contexts.

Objective: This study aims to assess and compare the evolution of the topics discussed on Chinese and French social media during the COVID-19 lockdown.

Methods: We extracted posts containing COVID-19-related or lockdown-related keywords in the most commonly used microblogging social media platforms (ie, Weibo in China and Twitter in France) from 1 week before lockdown to the lifting of the lockdown. A topic model was applied independently for three periods (prelockdown, early lockdown, and mid to late lockdown) to assess the evolution of the topics discussed on Chinese and French social media.

Results: A total of 6395; 23,422; and 141,643 Chinese Weibo messages, and 34,327; 119,919; and 282,965 French tweets were extracted in the prelockdown, early lockdown, and mid to late lockdown periods, respectively, in China and France. Four categories of topics were discussed in a continuously evolving way in all three periods: *epidemic news and everyday life*, *scientific information*, *public measures*, and *solidarity and encouragement*. The most represented category over all periods in both countries was *epidemic news and everyday life*. *Scientific information* was far more discussed on Weibo than in French tweets. Misinformation circulated through social media in both countries; however, it was more concerned with the virus and epidemic in China, whereas it was more concerned with the lockdown measures in France. Regarding *public measures*, more criticisms were identified in French tweets than on Weibo. Advantages and data privacy concerns regarding tracing apps were also addressed in French tweets. All these differences were explained by the different uses of social media, the different timelines of the epidemic, and the different cultural contexts in these two countries.

Conclusions: This study is the first to compare the social media content in eastern and western countries during the unprecedented COVID-19 lockdown. Using general COVID-19-related social media data, our results describe common and different public reactions, behaviors, and concerns in China and France, even covering the topics identified in prior studies focusing on specific



Social Media Platforms Listening Study On Atopic Dermatitis: Quantitative And Qualitative Findings



P.Voillot, B.Riche, M.Portafax, P.Foulquié, A.Gedik, S.Barbarot, L.Misery, S.Héas, A.Mebarki, N.Texier, S.Schück

JOURNAL OF MEDICAL INTERNET RESEARCH

Voillot et al

Original Paper

Social Media Platforms Listening Study on Atopic Dermatitis: Quantitative and Qualitative Findings

Paméla Voillot¹, MSc; Brigitte Riche², MS; Michel Portafax², MSc, MS; Pierre Foulquié¹, MSc; Anaïs Gedik¹, MSc; Sébastien Barbarot³, MD; Laurent Misery⁴, PhD; Stéphane Héas⁵, MS; Adel Mebarki¹, MS; Nathalie Texier¹, PharmD; Stéphane Schück¹, MD

¹Kap Code, Paris, France

²Sanofi Genzyme, Gentilly, France

³Centre Hospitalier Universitaire de Nantes, Nantes, France

⁴Centre Hospitalier Universitaire de Brest, Brest, France

⁵Université de Rennes 2, Rennes, France

Corresponding Author:

Paméla Voillot, MSc

Kap Code

28 rue d'Enghien

Paris, 75010

France

Phone: 33 972605764

Email: pamela.voillot@kapcode.fr

Abstract

Background: Atopic dermatitis (AD) is a chronic, pruritic, inflammatory disease that occurs most frequently in children but also affects many adults. Social media have become key tools for finding and disseminating medical information.

Objective: The aims of this study were to identify the main themes of discussion, the difficulties encountered by patients with respect to AD, the impact of the pathology on quality of life (QoL; physical, psychological, social, or financial), and to study the perception of patients regarding their treatment.

Methods: A retrospective study was carried out by collecting social media posts in French language written by internet users mentioning their experience with AD, their QoL, and their treatments. Messages related to AD discomfort posted between July 1, 2010, and October 23, 2020, were extracted from French-speaking publicly available online forums. Automatic and manual extractions were implemented to create a general corpus and 2 subcorpora depending on the level of control of the disease.

Results: A total of 33,115 messages associated with AD were included in the analysis corpus after extraction and cleaning. These messages were posted by 15,857 separate web users, most of them being women younger than 40 years. Tips to manage AD and everyday hygiene/treatments were among the most discussed topics for controlled AD subcorpus, while baby-related topics and therapeutic failure were among the most discussed topics for insufficiently controlled AD subcorpus. QoL was discussed in both subcorpora with a higher proportion in the controlled AD subcorpus. Treatments and their perception were also discussed by web users.

Conclusions: More than just emotional or peer support, patients with AD turn to online forums to discuss their health. Our findings show the need for an intersection between social media and health care and the importance of developing new approaches such as the Atopic Dermatitis Control Tool, which is a patient-related disease severity assessment tool focused on patients with AD.

(*J Med Internet Res* 2022;24(1):e31140) doi: [10.2196/31140](https://doi.org/10.2196/31140)

KEYWORDS

atopic dermatitis; Atopic Dermatitis Control Tool; health-related quality of life; social media use; real world; dermatology; skin disease; social media; online health information; online health; health care

Mental Health Condition Of College Students Compared To Non-Students During COVID-19 Lockdown: The Confins Study



J.Arsandaux, I.Montagni, M.Macalli, N.Texier, M.Pouriel, R.Germain, A.Mebarki, S.Kinouani, M.Tournier, S.Schück, C.Tzourio

Open access

Original research

BMJ Open Mental health condition of college students compared to non-students during COVID-19 lockdown: the CONFINS study

Julie Arsandaux ,¹ Ilaria Montagni ,¹ Mélissa Macalli,¹ Nathalie Texier,² Mathilde Pouriel,² Raphaël Germain,² Adel Mebarki,² Shérazade Kinouani ,¹ Marie Tournier,^{1,3} Stéphane Schuck,² Christophe Tzourio

To cite: Arsandaux J, Montagni I, Macalli M, et al. Mental health condition of college students compared to non-students during COVID-19 lockdown: the CONFINS study. *BMJ Open* 2021;11:e053231. doi:10.1136/bmjopen-2021-053231

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online. (<http://dx.doi.org/10.1136/bmjopen-2021-053231>).

Received 10 May 2021
Accepted 09 July 2021

ABSTRACT

Objectives To estimate the effect of student status on mental health condition during COVID-19 general lockdown in France.

Design Cross-sectional analysis comparing students and non-students recruited in the same study.

Setting Participants of the web-based CONFINS study implemented during the general lockdown in France in spring 2020.

Participants 2260 participants (78% women) including 1335 students (59%).

Primary and secondary outcome measures Through an online questionnaire, participants declared if they have experienced suicidal thoughts, coded their perceived stress on a 10-points scale and completed validated mental health scales (Patient Health Questionnaire-9 for depressive symptoms, Generalised Anxiety Disorder-7 for anxiety symptoms) during the last 7 days. The effect of college student status on each mental health condition was estimated using multivariate logistic regression analyses. Stratified models for students and non-students were performed to identify population-specific factors.

Results Student status was associated with a higher frequency of depressive symptoms (adjusted OR (aOR)=1.58; 95% CI 1.17 to 2.14), anxiety symptoms (aOR=1.51; 95% CI 1.10 to 2.07), perceived stress (n=1919, aOR=1.70, 95% CI 1.26 to 2.29) and suicidal thoughts (n=1919, aOR=1.57, 95% CI 0.97 to 2.53). Lockdown conditions that could be potentially aggravating on mental health like isolation had a higher impact on students than on non-students.

Conclusions College students were at higher risk of mental health disturbances during lockdown than non-students, even after taking into account several potential confounding factors. A close follow-up and monitoring of students' mental health status is warranted during lockdown periods in this vulnerable population.

Strengths and limitations of this study

- Direct comparison between a large sample of students and non-students recruited during the same period in a unique study.
- Large collection of variables, including lockdown conditions, allowing to investigate factors potentially associated to mental conditions for both students and non-students.
- Since students and non-students were volunteers, generalisation of the results could be limited.
- The cross-sectional design did not allow studying the temporal relationship between lockdown and mental health condition.

symptoms (eg, depressive symptoms, anxiety symptoms and suicidal risk) than other adults.^{1–4} In this population, and particularly in the first years at university, a high level of stress-related to academic achievement, a low self-esteem and depressive symptoms are frequently reported.^{5–7}

The unprecedented context of 2020 with the SARS-CoV-2 pandemic and its consequences have raised fears of its psychological impact in the population and more particularly in the fragile population of college students.⁸ In February 2020, the SARS-CoV-2 epidemic, causing the COVID-19 disease, hit France which is one of the most affected countries in the world in number of cases and deaths.⁹ To contain the spread of the epidemic, France established a general lockdown from 17 March to 11 May 2020. During previous lockdowns (eg, SARS epidemic in 2013), it has been shown that quarantine had by itself an impact on mental health (depression, post-traumatic stress symptoms, confusion, anger, suicide).^{10,11} These findings have also been found in the general population during the first COVID-19 lockdown period



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Bordeaux Population Health, Université de Bordeaux, Talence, France

²Kappa Santé, Paris, France

³Charles Perrens Hospital, Bordeaux, France

Correspondence to:
Pr Christophe Tzourio;
christophe.tzourio@u-bordeaux.fr



Fuzzy Matching For Symptom Detection In Tweets: Application To Covid-19 During The First Wave Of The Pandemic In France



Stéphane Schück, Avesta Roustamal, Anaïs Gedik, Paméla Voillot, Pierre Foulquié, Catherine Penfornis, Bernard Job

896

*Public Health and Informatics**J. Mantas et al. (Eds.)*

© 2021 European Federation for Medical Informatics (EFMI) and IOS Press.

This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/SHTI210308

Fuzzy Matching for Symptom Detection in Tweets: Application to Covid-19 During the First Wave of the Pandemic in France

Carole FAVIEZ^{a,1}, Pierre FOULQUIÉ^b, Xiaoyi CHEN^a, Adel MEBARKI^b, Sophie QUENNELL^{a,c}, Nathalie TEXIER^b, Sandrine KATSAHIAN^{a,d}, Stéphane SCHUCK^b, and Anita BURGUN^{a,e,f}

^a*Centre de Recherche des Cordeliers, Sorbonne Université, INSERM, Université de Paris, F-75006, Paris, France*

^b*Kap Code, Paris, France*

^c*M3C-Necker, Hôpital Necker-Enfants Malades, AP-HP, F-75015, Paris, France*

^d*Hôpital européen Georges Pompidou, Unité d'épidémiologie et de recherche clinique, AP-HP, F-75015, Paris, France*

^e*Hôpital Necker-Enfants Malades, Département d'informatique médicale, AP-HP, F-75015, Paris, France*

^f*PaRis Artificial Intelligence Research InstitutE (PRAIRIE), France*

Abstract. The exhaustive automatic detection of symptoms in social media posts is made difficult by the presence of colloquial expressions, misspellings and inflected forms of words. The detection of self-reported symptoms is of major importance for emergent diseases like the Covid-19. In this study, we aimed to (1) develop an algorithm based on fuzzy matching to detect symptoms in tweets, (2) establish a comprehensive list of Covid-19-related symptoms and (3) evaluate the fuzzy matching for Covid-19-related symptom detection in French tweets. The Covid-19-related symptom list was built based on the aggregation of different data sources. French Covid-19-related tweets were automatically extracted using a dedicated data broker during the first wave of the pandemic in France. The fuzzy matching parameters were finetuned using all symptoms from MedDRA and then evaluated on a subset of 5000 Covid-19-related tweets in French for the detection of symptoms from our Covid-19-related list. The fuzzy matching improved the detection by the addition of 42% more correct matches with an 81% precision.

Keywords. Content analysis, social media, fuzzy matching, symptoms, Covid-19



A New Method To Extract Health-Related Quality Of Life Data From Social Media Testimonies : Algorithm Development And Validation



Simon Renner, Tom Marty, Mickaïl Khadhar, Pierre Foulquié, Paméla Voillot, Adel Mebarki, Ilaria Montagni, Nathalie Texier, Stéphane Schück

JOURNAL OF MEDICAL INTERNET RESEARCH

Renner et al

Original Paper

A New Method to Extract Health-Related Quality of Life Data From Social Media Testimonies: Algorithm Development and Validation

Simon Renner^{1*}, PharmD; Tom Marty^{1*}, PharmD; Mickaïl Khadhar^{1*}, MSc; Pierre Foulquié¹, MSc; Paméla Voillot¹, MSc; Adel Mebarki¹, MSc; Ilaria Montagni², DPhil; Nathalie Texier¹, PharmD; Stéphane Schück¹, MD

¹Kap Code, Paris, France

²Bordeaux Population Health Research Center, UMR 1219, Bordeaux University, Inserm, Bordeaux, France

* these authors contributed equally

Corresponding Author:

Simon Renner, PharmD
Kap Code
4 Rue de Cléry
Paris, 75002
France
Phone: 33 9 72 60 57 63
Email: simon.renner@kapcode.fr

Abstract

Background: Monitoring social media has been shown to be a useful means to capture patients' opinions and feelings about medical issues, ranging from diseases to treatments. Health-related quality of life (HRQoL) is a useful indicator of overall patients' health, which can be captured online.

Objective: This study aimed to describe a social media listening algorithm able to detect the impact of diseases or treatments on specific dimensions of HRQoL based on posts written by patients in social media and forums.

Methods: Using a web crawler, 19 forums in France were harvested, and messages related to patients' experience with disease or treatment were specifically collected. The SF-36 (Short Form Health Survey) and EQ-5D (Euro Quality of Life 5 Dimensions) HRQoL surveys were mixed and adapted for a tailored social media listening system. This was carried out to better capture the variety of expression on social media, resulting in 5 dimensions of the HRQoL, which are physical, psychological, activity-based, social, and financial. Models were trained using cross-validation and hyperparameter optimization. Oversampling was used to increase the infrequent dimension: after annotation, SMOTE (synthetic minority oversampling technique) was used to balance the proportions of the dimensions among messages.

Results: The training set was composed of 1399 messages, randomly taken from a batch of 20,000 health-related messages coming from forums. The algorithm was able to detect a general impact on HRQoL (sensitivity of 0.83 and specificity of 0.74), a physical impact (0.67 and 0.76), a psychic impact (0.82 and 0.60), an activity-related impact (0.73 and 0.78), a relational impact (0.73 and 0.70), and a financial impact (0.79 and 0.74).

Conclusions: The development of an innovative method to extract health data from social media as real time assessment of patients' HRQoL is useful to a patient-centered medical care. As a source of real-world data, social media provide a complementary point of view to understand patients' concerns and unmet needs, as well as shedding light on how diseases and treatments can be a burden in their daily lives.

(*J Med Internet Res* 2022;24(1):e31528) doi: [10.2196/31528](https://doi.org/10.2196/31528)

KEYWORDS

health-related quality of life; social media use; measures; real world; natural language processing; social media; NLP; infoveillance; quality of life; digital health; social listening

Introduction

Most people use the internet regularly to research and discuss health-related topics. Patients give and receive advice on their

diseases and treatments in online forums and social media platforms [1]. These messages are massive, continuously generated, and easy to access [2]. This type of information is direct, genuine, and authentic, offering access to new real-world



Acceptance of a Covid-19 vaccine is associated with ability to detect fake news and health literacy



I Montagni, K Ouazzani-Touhami, A Mebarki, N Texier, S Schück, C Tzourio,
CONFINS group

Acceptance of a Covid-19 vaccine is associated with ability to detect fake news and health literacy

**I Montagni¹, K Ouazzani-Touhami^{1,2}, A Mebarki³, N Texier^{3,4}, S Schück^{3,4},
C Tzourio¹, the CONFINS group**

¹Univ.Bordeaux, Inserm, Bordeaux Population Health Research Center, U1219, F-33000, Bordeaux, France

²Polyclinique Inkermann, groupe Elsan, F-79000 Niort, France

³Kap Code, F-75010 Paris, France

⁴Kappa Santé, F-75002 Paris, France

Address correspondence to Ilaria Montagni, E-mail: ilaria.montagni@u-bordeaux.fr

ABSTRACT

Background During the Covid-19 pandemic fake news has been circulating impacting on the general population's opinion about a vaccine against the SARS-CoV-2. Health literacy measures the capacity of navigating health information.

Methods We used data from a prospective national online cohort of 1647 participants. Descriptive statistics, Chi² and ANOVA independence tests and two multivariable multinomial regression models were performed. Interactions between each variable were tested.

Results Detection of fake news and health literacy scores were associated with intention to get vaccinated against SARS-CoV-2 ($p < 0.01$). The risk of being "anti-vaccination" or "hesitant", rather than "pro-vaccination", was higher among individuals reporting bad detection of fake news, respectively OR = 1.93 (95%CI = [1.30;2.87]) and OR = 1.80 (95%CI = [1.29;2.52]). The risk of being in "hesitant", rather than "pro-vaccination" was higher among individuals having a bad health literacy score (OR = 1.44; 95%CI = [1.04;2.00]). No interaction was found between detection of fake news and health literacy.

Conclusions To promote acceptance of a vaccine against SARS-CoV-2, it is recommended to increase individuals' ability to detect fake news and health literacy through education and communication programs.

Keywords misinformation, fake news, health literacy, vaccination, Covid-19

Introduction

Misinformation about health-related subjects represents a public health threat. Around Covid-19, fake news has been circulating extensively since the beginning of the pandemic thus making the general audience doubt the veracity of health and political authorities concerning information around the SARS-CoV-2.¹ In February, the World Health Organisation's Director-General declared the global 'over-abundance' of Covid-19 information an 'infodemic'.

One of the challenges for health communication introduced by Covid-19 has been to tackle this increasing amount of false content on several communication channels, including social media platforms.² Common fake news about Covid-19 include the fact that only older adults can be infected, that swallowing bleach can kill the virus or that the virus is the product of a laboratory.³ For instance, in Iran, hundreds of people died after ingesting alcohol in a bid to treat Covid-19 as a result of misinformation circulating on social media.⁴

The large spread of misinformation about Covid-19 might be explained by the initial scarce knowledge about the virus among the scientific community and politicians. Confusion generated by the plethora of news across media could have nourished misinformation and lack of trust in scientific evidence, especially in a situation where people have been looking for immediate and reassuring answers regarding the SARS-CoV-2.⁵ Spread of fake news has been fueled also by public figures and politicians who have been giving, in some

I Montagni, Researcher and Lecturer

K Ouazzani-Touhami, Public Health Medical Doctor

A Mebarki, General Co-Director

N Texier, Director

S Schück, Chief Executive Officer

C Tzourio, Professor and Head



Physicians' Perceptions of the Use of a Chatbot for Information Seeking: Qualitative Study



Jason Koman, Khristina Fauvette, Stéphane Schuck, Nathalie Texier, Adel Mebarki

JOURNAL OF MEDICAL INTERNET RESEARCH

Koman et al

Original Paper

Physicians' Perceptions of the Use of a Chatbot for Information Seeking: Qualitative Study

Jason Koman^{1,2,3*}, PhD; Khristina Fauvette⁴, MD; Stéphane Schuck^{1*}, MD; Nathalie Texier¹, MD; Adel Mebarki¹, MD

¹Kap Code, Paris, France

²CNRS, PASSAGES, Bordeaux, France

³Bordeaux Population Health Research Center, University of Bordeaux, Inserm, Bordeaux, France

⁴Sanofi Aventis, Gentilly Cedex, France

*these authors contributed equally

Corresponding Author:

Khristina Fauvette, MD

Sanofi Aventis

82, avenue Raspail

Gentilly Cedex, 94255

France

Phone: 33 772219558

Email: kchristina.fauvette@sanofi.com

Abstract

Background: Seeking medical information can be an issue for physicians. In the specific context of medical practice, chatbots are hypothesized to present additional value for providing information quickly, particularly as far as drug risk minimization measures are concerned.

Objective: This qualitative study aimed to elicit physicians' perceptions of a pilot version of a chatbot used in the context of drug information and risk minimization measures.

Methods: General practitioners and specialists were recruited across France to participate in individual semistructured interviews. Interviews were recorded, transcribed, and analyzed using a horizontal thematic analysis approach.

Results: Eight general practitioners and 2 specialists participated. The tone and ergonomics of the pilot version were appreciated by physicians. However, all participants emphasized the importance of getting exhaustive, trustworthy answers when interacting with a chatbot.

Conclusions: The chatbot was perceived as a useful and innovative tool that could easily be integrated into routine medical practice and could help health professionals when seeking information on drug and risk minimization measures.

(*J Med Internet Res* 2020;22(11):e15185) doi: [10.2196/15185](https://doi.org/10.2196/15185)

KEYWORDS

health; digital health; innovation; conversational agent; decision support system; qualitative research; chatbot; bot; medical drugs; prescription; risk minimization measures



Exploring the Health-Related Quality of Life of Patients Treated With Immune Checkpoint Inhibitors: Social Media Study



François-Emery Cotté, Paméla Voillot, Bryan Bennett, Bruno Falissard, Christophe Tzourio, Pierre Foulquié, Anne-Françoise Gaudin, Hervé Lemasson¹, Valentine Grumberg, Laura McDonald, Carole Faviez, Stéphane Schück

JOURNAL OF MEDICAL INTERNET RESEARCH

Cotté et al

Original Paper

Exploring the Health-Related Quality of Life of Patients Treated With Immune Checkpoint Inhibitors: Social Media Study

François-Emery Cotté¹, PharmD, PhD, MPH; Paméla Voillot², MSc; Bryan Bennett³, PhD; Bruno Falissard^{4,5,6,7}, MD, PhD; Christophe Tzourio^{8,9}, MD, PhD; Pierre Foulquié², MSc; Anne-Françoise Gaudin¹, PharmD; Hervé Lemasson¹, MSc; Valentine Grumberg¹⁰, MSc; Laura McDonald¹¹, PhD; Carole Faviez², MSc; Stéphane Schück², MD

¹Bristol-Myers Squibb, Rueil-Malmaison, France

²Kap Code, Paris, France

³Bristol-Myers Squibb, Stockport, United Kingdom

⁴Paris-Sud University, Paris, France

⁵Paris-Descartes Université, Paris, France

⁶AP-HP, Paris, France

⁷INSERM U1178, Paris, France

⁸Bordeaux University, Bordeaux, France

⁹Inserm U1219, Bordeaux, France

¹⁰UFR Pharmacy, Grenoble Alpes University, Grenoble, France

¹¹Bristol-Myers Squibb, London, United Kingdom

Corresponding Author:

Paméla Voillot, MSc

Kap Code

28 rue d'Enghien

Paris, 75010

France

Phone: 33 624058742

Email: pamela.voillot@kapcode.fr

Abstract

Background: Immune checkpoint inhibitors (ICIs) are increasingly used to treat several types of tumors. Impact of this emerging therapy on patients' health-related quality of life (HRQoL) is usually collected in clinical trials through standard questionnaires. However, this might not fully reflect HRQoL of patients under real-world conditions. In parallel, users' narratives from social media represent a potential new source of research concerning HRQoL.

Objective: The aim of this study is to assess and compare coverage of ICI-treated patients' HRQoL domains and subdomains in standard questionnaires from clinical trials and in real-world setting from social media posts.

Methods: A retrospective study was carried out by collecting social media posts in French language written by internet users mentioning their experiences with ICIs between January 2011 and August 2018. Automatic and manual extractions were implemented to create a corpus where domains and subdomains of HRQoL were classified. These annotations were compared with domains covered by 2 standard HRQoL questionnaires, the EORTC QLQ-C30 and the FACT-G.

Results: We identified 150 users who described their own experience with ICI (89/150, 59.3%) or that of their relative (61/150, 40.7%), with 137 users (91.3%) reporting at least one HRQoL domain in their social media posts. A total of 8 domains and 42 subdomains of HRQoL were identified: Global health (1 subdomain; 115 patients), Symptoms (13; 76), Emotional state (10; 49), Role (7; 22), Physical activity (4; 13), Professional situation (3; 9), Cognitive state (2; 2), and Social state (2; 2). The QLQ-C30 showed a wider global coverage of social media HRQoL subdomains than the FACT-G, 45% (19/42) and 29% (12/42), respectively. For both QLQ-C30 and FACT-G questionnaires, coverage rates were particularly suboptimal for Symptoms (68/123, 55.3% and 72/123, 58.5%, respectively), Emotional state (7/49, 14% and 24/49, 49%, respectively), and Role (17/22, 77% and 15/22, 68%, respectively).

Conclusions: Many patients with cancer are using social media to share their experiences with immunotherapy. Collecting and analyzing their spontaneous narratives are helpful to capture and understand their HRQoL in real-world setting. New measures



The Use of Web Forums Data to Evaluate Online Conversations Associated with Gastrointestinal Discomfort: A Retrospective 15-Year Study of 200 000 Messages from French-Speaking Platforms



Boris Le Nevé, Carole Faviez, Florent Schafer, Jean-François Jeanne, Paméla Voillot, Pierre Foulquié, Guy Fagherazzi, Stéphane Schück

THE USE OF WEB FORUMS DATA TO EVALUATE ONLINE CONVERSATIONS ASSOCIATED WITH GASTROINTESTINAL DISCOMFORT: A RETROSPECTIVE 15-YEAR STUDY OF 200 000 MESSAGES FROM FRENCH-SPEAKING PLATFORMS

Boris Le Nevé, Carole Faviez, Florent Schafer, Jean-François Jeanne, Paméla Voillot, Pierre Foulquié, Guy Fagherazzi, Stéphane Schück

Introduction: Gastrointestinal (GI) discomfort (ex: bloating, flatulence) is very common and can significantly affect well-being and quality of life. Our aim was to explore online conversations on this topic. **MATERIAL & METHODS.** Messages related to GI discomfort were extracted from French-speaking generalist and specialized forums from January 2003 to August 2018 using the Detec't Extractor [1]. Messages were cleaned, deidentified and relevant medical concepts were identified using MedDRA 15.0 (Medical Dictionary for Regulatory Activities). Topic modeling was performed using a correlated topic model (CTM) based on the Latent Dirichlet Allocation (LDA). Users' age category and gender were identified respectively by linear regression and application of a SVM (Support Vector Machine). **Results:** A total of 198 866 messages were extracted from 14 major French-speaking web forums. Active users with available gender were mostly women (11630 vs 1651 men) under 30 years. Average number of messages per user was 4.57. Six classes of topics were identified through topic modelling: "medical consultations/exams", "diet", "symptoms", "quality of life", "treatments" and "stress and symptoms". Within the topic "diet", the most frequently discussed concept was "nausea and vomiting linked to food intake". **CONCLUSIONS.** GI discomfort is an actively discussed topic on French-speaking web forums. Gender and age of most active users tend to mirror the higher prevalence of functional gastrointestinal disorders like irritable bowel syndrome in young women. Our innovative approach has shown that identifying discussion topics associated to GI discomfort online is feasible and can serve as a complementary source of real-world evidence for caregivers. [1]. Abdellaoui R, Schück S, Texier N, et al. Filtering Entities to Optimize Identification of Adverse Drug Reaction From Social Media: How Can the Number of Words Between Entities in the Messages Help? JMIR public health and surveillance, 2017, vol. 3, no 2.



Position Paper : COVID-19, l'enjeu des données



COVID-19 : L'ENJEU DES DONNÉES

RÉSUMÉ EXÉCUTIF

Le Healthcare Data Institute (HDI) est un think tank dédié à la transformation du système de santé à travers l'usage scientifique et économique des données de santé au bénéfice des acteurs du système et des citoyens.

Au printemps 2020, le HDI s'est mobilisé et a créé une Task Force dédiée à la lutte contre l'épidémie de COVID-19. La Task Force a pour mission d'identifier les domaines dans lesquels les données sont apparues, durant la première vague de l'épidémie, comme des facteurs d'amélioration de la réponse aux crises sanitaires ou parfois plus largement, du système de santé.

Après avoir entendu les porteurs d'initiatives publiques ou privées durant une session d'auditions, la Task Force a identifié trois axes majeurs d'amélioration du système et de sa réponse aux crises sanitaires, grâce aux données :

- 1. La data et le numérique en santé:** des outils puissants de lutte contre la discontinuité des soins et des essais cliniques en période de crise sanitaire
- 2. Des données adaptées, accessibles et transparentes:** un élément essentiel du dispositif de pilotage des crises sanitaires
- 3. Le jeu collectif:** les organisations collaboratives émergentes et le partage des données pour améliorer la qualité des réponses aux crises sanitaires.

Analyse des réseaux sociaux pour identifier les motifs de l'hésitation vaccinale anti-HPV : une étude infodémiologique



Simon Renner, Tom Marty, Pamela Voillot, Pierre Foulquié, Adel Mebarki, et Stéphane Schück

Kap•Code

ÉTUDE DES RÉSEAUX SOCIAUX POUR IDENTIFIER LES MOTIFS DE L'HÉSITATION VACCINALE ANTI-HPV : UNE ÉTUDE INFODÉMIOLOGIQUE

Simon Renner¹, Tom Marty¹, Pamela Voillot¹, Pierre Foulquié¹, Adel Mebarki¹, et Stéphane Schück¹

¹Kap Code, 28 rue d'Enghien 75010 Paris, France

Evanex

INTRODUCTION

La recommandation vaccinale contre l'infection à papillomavirus humain (HPV) est un enjeu mondial de santé publique. En 2019, la couverture vaccinale de 21% place la France en dernière position européenne. Parallèlement, les réseaux sociaux sont un espace privilégié pour plus de 32 millions d'internautes actifs. Ces plateformes occupent une place grandissante comme lieu d'expression ou les internautes échangent sur leur santé, leurs préoccupations et souvent sur leur position vis-à-vis de la vaccination. Analyser et comprendre ces messages permettrait d'identifier les leviers associés à une amélioration de la couverture vaccinale anti-HPV.

MATÉRIEL ET MÉTHODES

L'étude a été menée via l'observatoire de la vaccination EVANEX[®]. Les messages associés à la vaccination anti-HPV et écrits en français (dataset HPV) ont été extraits à partir de 23 forums médicaux et de réseaux sociaux, entre 2006 et 2019. Les termes d'extraction faisaient soit référence à un acte de vaccination contre le HPV « vaccin contre le papillomavirus » soit directement à un nom de produit (Gardasil[®] et Cervarix[®]). Une annotation manuelle de messages (présence ou non d'une hésitation) a permis la création d'un Gold standard visant à développer un algorithme de détection d'une hésitation vaccinale. Cet algorithme a ensuite été appliqué sur l'ensemble du dataset HPV. Un premier corpus d'analyse (pré-corpus d'hésitation vaccinale, Figure 1) a pu être constitué. Une annotation manuelle de l'ensemble de ce corpus d'intérêt a été effectuée afin d'identifier les différents facteurs d'hésitation (corpus d'hésitation vaccinale, Figure 1) et les regrouper par grandes thématiques d'hésitation.

RÉSULTATS

Le dataset HPV contenait 12 609 messages uniques rédigés par 5 117 internautes différents. Un pré-corpus de 3 192 messages propres à l'hésitation vaccinale a été déterminé algorithmiquement. L'analyse manuelle de ces 3 192 messages a permis de caractériser 593 messages d'hésitation vaccinale (18.6% du pré-corpus, 4.7% du dataset HPV) (Figure 1).

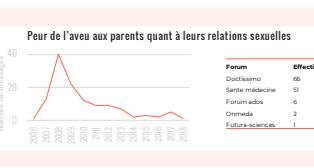
(...) sa fait a peu près l'an et demi que je ne suis plus vierge ma mère m'a fait faire le vaccin HPV mais je ne ve pas lui dire que je ne suis plus vierge eske c'est compte un risque que je me fasse vacciner ?

Figure 2 : Exemples de messages extraits

Les messages (Figure 2) ont été regroupés en 3 principaux groupes, selon les facteurs d'hésitation vaccinale exprimés. Le premier a pour motif l'influence de la vie sexuelle sur la vaccination (339 messages, 7.5% pré-corpus). Au sein de celui-ci, les adolescentes s'interrogent à la fois sur la nécessité de se faire vacciner mais aussi sur une éventuelle dangerosité du vaccin une fois sexuellement actif (n=113; Figure 3) ou si un rapport advient entre deux injections (n=70). La crainte de l'apparition d'effets indésirables en raison de leur activité sexuelle est importante.

Figure 3 : Actif sexuellement - Volumétrie

Parallèlement les partenaires s'interrogent sur les mêmes thématiques (n=30). La peur de l'aveu d'une vie sexuelle aux parents est également source de préoccupations (n=126 ; Figure 4). Les adolescentes sont à la recherche d'un moyen d'informer leur médecin de leur activité sexuelle sans le signaler à leurs parents, présents lors de la consultation médicale.



Le second groupe concerne le manque d'informations (196 messages 5.6% pré-corpus). Ce manque d'informations peut

être global (n=60, Figure 5) sur les modalités vaccinales : obligation vaccinale, efficacité, utilité, remboursement ...

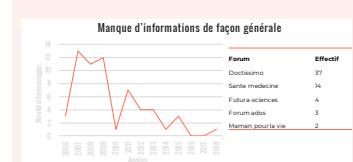


Figure 5 : Manque d'informations générale – Volumétrie

Le manque d'informations porte également sur les recommandations sur les schémas d'injections et les rappels (n=16) ou l'âge limite de vaccination, notamment si une adolescente est toujours vierge (n=15).

Le troisième groupe exprime l'influence qu'ont les différentes sources extérieures sur l'hésitation vaccinale (n=59, 1.8% du pré-corpus). Ces messages, de personnes en cours de vaccination anti-HPV ou non, mettent en avant une peur des effets indésirables (n=30 ; Figure 6) ou des aspects négatifs des injections (n=20) après consultation d'un site internet, de témoignages ou une discussion.

9 messages évoquent une source médicale déconseillant explicitement la vaccination anti-HPV malgré la volonté des parents et de l'adolescente de se faire vacciner.



Figure 6 : Volumétrie impact sources extérieures

Appartenant aux groupes 2 et 3, les mères (n=39) justifient leur hésitation à faire vacciner leur fille par le manque d'informations claires et concordantes.

CONCLUSION

Cette étude a permis de mettre en avant l'existence d'une forte communauté d'internautes en débat et en recherche d'informations concernant la vaccination anti-HPV. L'analyse du contenu des messages a permis d'isoler deux motifs principaux d'hésitation vaccinale : l'impact des relations sexuelles et le manque d'informations. La compréhension fine de ces facteurs d'hésitation permettent d'identifier les leviers associés. A des échelles locales, en fonction des comportements et caractéristiques des populations, ce type d'étude infodémiologique pourrait servir les stratégies et politiques de santé publique d'amélioration de la couverture vaccinale. Ces travaux sur l'hésitation vaccinale anti-HPV ont été effectués avant la recommandation de la HAS d'ouvrir la vaccination aux garçons. Ils pourraient être continués spécifiquement sur l'hésitation vaccinale exprimée sur internet par les adolescents afin de déchiffrer leurs motifs d'hésitation.

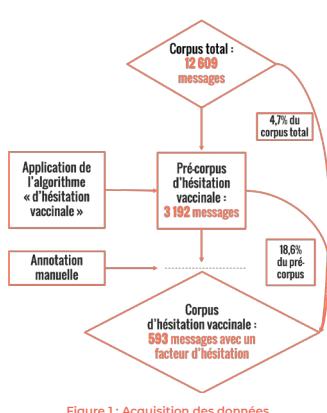


Figure 1 : Acquisition des données

Conception d'un algorithme permettant de détecter l'hésitation vaccinale anti-papillomavirus humain au sein de messages issus des réseaux sociaux



Pierre Foulquié, Anaïs Gedik, Simon Renner, Paméla Voillot, Adel Mebarki et Stéphane Schück

Kap•Code

CONCEPTION D'UN ALGORITHME PERMETTANT DE DÉTECTER L'HÉSITATION VACCINALE ANTI-PAPILLOMAVIRUS HUMAIN AU SEIN DE MESSAGES ISSUS DES RÉSEAUX SOCIAUX

Pierre Foulquié¹, Anaïs Gedik¹, Simon Renner¹, Paméla Voillot¹, Adel Mebarki¹ et Stéphane Schück¹

¹Kap Code, 28 rue d'Enghien 75010 Paris, France

Detect'

INTRODUCTION

La France est un des pays où l'hésitation vaccinale anti-papillomavirus humain (HPV) est la plus forte au monde. Cette hésitation s'observe particulièrement sur les réseaux sociaux, interface où les internautes peuvent s'exprimer librement sur leur santé. L'amélioration de l'acceptabilité vaccinale HPV passe par la compréhension des déterminants de l'hésitation. Un algorithme d'analyse sémantique capable d'identifier les messages exprimés sur les réseaux sociaux contenant une hésitation vaccinale anti-HPV permettrait d'analyser et de comprendre ce phénomène.

MATÉRIEL ET MÉTHODES

Un corpus de messages associés à la vaccination anti-HPV, postés entre 2006 et 2019, a été extrait dans le cadre du projet Detect' [1] à partir de 17 sources francophones. Les 23 mots-clés d'extraction évoquaient plusieurs sujets associés à la sphère du papillomavirus : la vaccination anti-HPV, la sexualité et l'anatomie. Une annotation d'un échantillon du corpus a été effectuée par 3 annotateurs qui disposaient d'une charte d'annotation basée sur la définition de l'hésitation vaccinale de l'OMS [2]. Elle a permis de classer les messages comme exprimant de l'hésitation ou non et d'extraire des expressions des différentes perceptions vaccinales (anti-vaccin, pro-vaccin). Le gold standard (GS) ainsi créé a été réparti en 2 jeux de données. Un premier jeu, dit « d'entraînement » et contenant 85% des données, a été utilisé pour entraîner le modèle. Le deuxième jeu, désigné comme jeu « de validation » et constitué des 15% restants, a servi à la validation de la méthode.

À partir du jeu d'entraînement, plusieurs variables ont été déterminées à l'aide des formes syntaxiques des messages (N-grams), de la présence des mots de champs lexicaux spécifiques (anti-vaccin, pro-vaccin, etc.) et du word embedding (représentation contextuelle des mots via un modèle Glove [3]).

Par la suite et afin de mettre en place un modèle performant en termes de précision de détection d'hésitation vaccinale, une recherche de la meilleure combinaison entre différents classificateurs (support vector

classification, logistic regression, random forest, et extreme gradient boosting) et les différentes variables identifiées précédemment a été effectuée.

RÉSULTATS

1 370 messages contenant une mention de vaccination anti-HPV ont été extraits pour annotation. Les sources les plus présentes dans l'échantillon étaient les forums Doctissimo et Santé Médecine, avec respectivement 615 et 193 messages, suivis de Twitter (n=395). Le terme « Gardasil » est le mot d'extraction ayant permis de recueillir le plus de messages dans cet échantillon (n=967).

L'annotation a permis d'identifier 497 messages (36%) représentant une hésitation vaccinale et 891 (64%) une perception positive, négative ou neutre de la vaccination. Les jeux de données d'entraînement et de validation étaient composés respectivement de 1 164 et 206 messages. La meilleure combinaison de variables identifiées est constituée des 300 premiers N-grams en terme de pouvoir prédictif de l'hésitation vaccinale. Ces formes syntaxiques sont composées d'un mot unique ou d'associations de deux ou trois mots telles que « savoir »; « mère veut »; « rapports sexuels ».

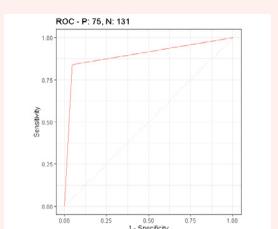
Suite aux tests, le meilleur modèle choisi est un classifieur binaire Random Forest, discriminant correctement l'hésitation des autres perceptions dans 91% des cas. Les performances du modèle, calculées à partir de la matrice de confusion (Tableau 1), sont regroupées dans le Tableau 2. Les messages classés comme évocateur d'hésitation vaccinale exprimaient une réelle hésitation dans 91% des cas (valeur prédictive positive). Parmi tous les messages annotés comme exprimant une hésitation vaccinale, 84 % des messages ont été identifiés par notre modèle (sensibilité). Le Graphique 1 présente la courbe ROC du modèle.

		Référence	
		Oui	Non
Prédiction	Oui	63	6
	non	12	125

Tableau 1 - Matrice de confusion du modèle

Precision	Sensibilité	Spécificité	Value prédite positive	Value prédite négative
91,26%	84,00%	95,42%	91,30%	91,24%

Tableau 2 – Performances du modèle



Graphique 1 – Courbe ROC du modèle

CONCLUSION

Développer un algorithme d'analyse sémantique capable d'identifier une hésitation vaccinale anti-HPV au sein de messages issus des réseaux sociaux pourrait se révéler être un nouvel outil d'aide à l'identification des déterminants de l'hésitation dans le cadre d'une couverture insuffisante.

Les performances de l'algorithme sur des données n'ayant pas servi au développement de ce dernier démontrent que ce type d'outil est efficace pour identifier puis analyser les messages d'internautes exprimant une hésitation vaccinale anti-HPV.

Ce travail ouvre de nombreuses possibilités de travaux futurs. Des méthodes complémentaires, permettant par exemple d'identifier des causalités, pourraient permettre d'identifier les facteurs de cette hésitation. De plus, l'étude des utilisateurs exprimant une hésitation vaccinale permettrait d'établir des profils types et d'étudier l'évolution temporelle de cette hésitation. Ceci pourrait ouvrir la voie à l'instauration d'outils de monitoring de la vaccination sur les réseaux sociaux dans un objectif de santé publique.

RÉFÉRENCES

- [1] Abdellaoui, R., Schück, S., Texier, N., & Burgun, A. (2017). Filtering entities to optimize identification of adverse drug reaction from social media: how can the number of words between entities in the messages help? JMIR public health and surveillance, 3(2), e36.
- [2] Ten threats to global health in 2019, <https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019>
- [3] Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 1532-1543).



Segmentation des internautes issus des réseaux sociaux au cœur de la controverse sur la vaccination



Paméla Voillot, Avesta Roustamal, Anaïs Gedik, Pierre Foulquié, Simon Renner, Adel Mebarki et Stéphane Schück

Kap•Code

SEGMENTATION DES INTERNAUTES ISSUS DES RÉSEAUX SOCIAUX AU CŒUR DE LA CONTROVERSE SUR LA VACCINATION

Paméla Voillot¹, Avesta Roustamal¹, Anaïs Gedik¹, Pierre Foulquié¹, Simon Renner¹, Adel Mebarki¹ et Stéphane Schück¹

¹Kap Code, 29 rue d'Enghien 75010 Paris, France

Evanex

INTRODUCTION

L'élargissement de l'obligation vaccinale dès 2018 a permis d'améliorer le taux de couverture vaccinale en France [1]. En effet, l'éradication de certaines maladies comme la rougeole nécessite un niveau de couverture vaccinale de 95% [2]. L'obtention d'une couverture vaccinale suffisante est un objectif majeur de santé publique. Néanmoins, une partie de la population reste toujours réticente à cette mesure. L'hésitation vaccinale persistante a été classée par l'OMS comme l'une des plus grandes menaces sur la santé. En complémentarité avec les bases de données publiques, les réseaux sociaux, avec plus de 32 millions d'internautes, sont une source d'échanges et de partage d'information où la parole est libérée [3] et permet d'exprimer ses doutes et interrogations sur la vaccination. La caractérisation de ces internautes pourrait aider à mieux comprendre l'attitude de la population envers la vaccination. L'objectif de cette étude était d'identifier la typologie des internautes qui s'expriment sur les réseaux sociaux au sujet de la vaccination.

MATÉRIEL ET MÉTHODES

Dans le cadre de la mise en place d'un observatoire sur la vaccination appelé Evanex®, les messages d'internautes issus des réseaux sociaux ont été extraits à partir d'un webcrawler développé par la société Kap Code. L'analyse des messages a été réalisée au moyen de divers outils de Machine Learning. La méthode de classification hiérarchique ascendante (CAH) a été utilisée pour former des catégories d'internautes [4]. Celle-ci se base sur plusieurs variables telles que l'activité de l'internaute sur les réseaux sociaux (nombre de messages, fréquence de publication, etc.), les sujets abordés dans ses messages (méthode de topic modeling) et les concepts médicaux exprimés (déttection de termes médicaux à partir du dictionnaire médical MedDRA, enrichi avec du vocabulaire d'internaute). Les individus qui présentaient des centres de gravité similaires après le calcul des variables, ont été regroupé ensemble par la méthode de Ward. Les groupes obtenus sont interprétés à partir de leurs centres de gravité.

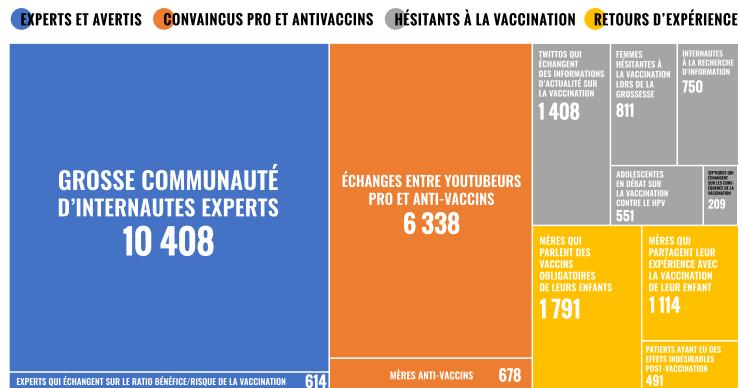


Figure 1 – Segmentation des internautes

RÉSULTATS

L'analyse a été réalisée sur 88 894 messages provenant de 37 forums français (Doctissimo, Au Féminin...) et a mis en évidence une diversité d'internautes actifs sur des discussions autour de la vaccination (25163 internautes). La segmentation a permis d'identifier 12 catégories d'internautes regroupées en 4 grands groupes d'utilisateurs (figure 1). Le groupe rassemblant le plus d'individus (n=11 022; 44%) est composé d'internautes « experts » et avertis qui échangent sur les risques et bénéfices de la vaccination. Ils échangent également des informations scientifiques factuelles. Ces internautes constituent une communauté très active et aguerrie avec une moyenne de rédaction de 5 messages et 14 concepts médicaux par post. Le deuxième groupe est composé des convaincus pro et anti-vaccins (n=7016, 28%) qui soutiennent une certitude dogmatique. On y retrouve des YouTubers pro ou anti-vaccins ainsi que des mères anti-vaccins. Ce groupe est défini par un nombre important d'expression de termes médicaux tels que l'autisme, les effets indésirables et l'aluminium. Le troisième groupe d'internautes est composé d'hésitants (n=3729, 15%). Ils s'expriment sur des forums spécifiques comme Forum Ados ou le Journal des Femmes. Les hésitants sont majoritairement des mères qui se questionnent sur les obligations vaccinales. Il est également composé d'adolescentes qui débâtent sur les risques de la vaccination anti-HPV, avec des termes médicaux exprimés comme « tumeur maligne », « sclérose en plaque » et « grossesse ». Le dernier groupe est représenté par des internautes qui décrivent un retour d'expérience (n=3396; 13%). Ce groupe est caractérisé par des individus qui partagent les effets indésirables post-vaccination vécus personnellement ou par un proche ainsi que des mères témoignant d'expériences de vaccination de leurs enfants.

CONCLUSION

Cette étude démontre que la vaccination est un sujet qui continue à diviser. Les internautes partagent leurs inquiétudes mais sont également à la recherche d'informations. Les réseaux sociaux sont une source à surveiller et à exploiter pour contribuer aux stratégies de Santé publique d'amélioration de la couverture vaccinale.

RÉFÉRENCES

- [1] Qu'est-ce que la couverture vaccinale ?, Dossier thématique, <https://www.santepubliquefrance.fr/determinants-de-sante/vaccination/articles/qu-est-ce-que-la-couverture-vaccinale>, 29 Juin 2019
- [2] Les programmes de vaccination sont de plus en plus confrontés aux hésitations de la population, Communiqué de Presse OMS, 18 Août 2015, <https://www.who.int/fr/news-room/detail/18-08-2015-vaccine-hesitancy-a-growing-challenge-for-immunization-programmes>
- [3] Fox S, Duggan M. PewInternet. 2013 Jan 15. Health Online 2013 URL: <http://www.pewinternet.org/2013/01/15/health-online-2013/> [accessed 2018-02-27] [WebCite Cache ID 6xXiCgjLK]
- [4] S. Schück et al, Que nous apportent les réseaux sociaux quant à la crise sanitaire du Levothyrox® d'août 2017 ?, Revue d'Épidémiologie et de Santé Publique, Volume 66, Supplément 4, 2018, Page S225, ISSN 0398-7620

Analyse temporelle du volume et de la perception relatifs aux masques à partir de tweets français pendant la période de confinement



Pierre Foulquié, Ludovic Figarella, Léa Châteauneuf, Simon Renner, Adel Mebarki, Nathalie Texier et Stéphane Schück

Kap•Code

ANALYSE TEMPORELLE DU VOLUME ET DE LA PERCEPTION RELATIFS AUX MASQUES À PARTIR DE TWEETS FRANÇAIS PENDANT LA PÉRIODE DE CONFINEMENT

Pierre Foulquié¹, Ludovic Figarella¹, Léa Châteauneuf¹, Simon Renner¹, Adel Mebarki¹, Nathalie Texier¹ et Stéphane Schück¹

¹Kap Code, 28 rue d'Enghien 75010 Paris, France

Detect'

INTRODUCTION

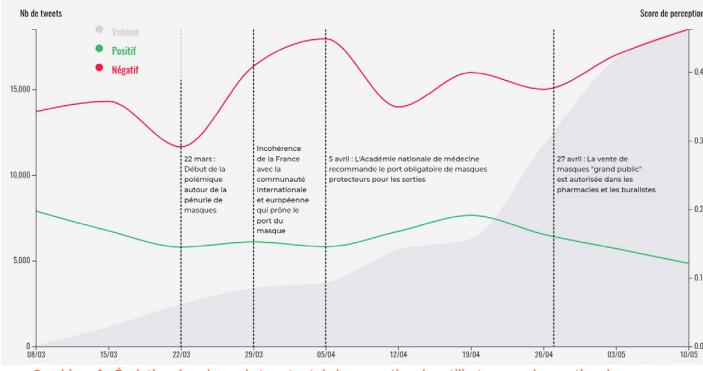
En 2020, en pleine pandémie de Covid-19, la gestion du matériel sanitaire, dont font partie les masques, a été au cœur de plusieurs débats : efficacité du port du masque comme moyen de prévention, communication du gouvernement ou encore gestion du stock d'État. De tels sujets, suscitant des divergences de discours, sont analysables à l'aide de réseaux sociaux comme Twitter. Ces canaux permettent à la population de réagir à l'actualité, d'y exprimer ses opinions et sa perception des sujets de santé publique. Cette étude propose une analyse temporelle de l'apparition du sujet des masques et de l'évolution des opinions associées.

MATÉRIEL ET MÉTHODES

Un corpus de tweets français liés au Covid-19 a été constitué grâce à l'API Twitter. Les tweets extraits contenaient des mots-clés ou des hashtags évocateurs du coronavirus (#coronavirusfrance, #covid19fr, ...) et du confinement (#restezchezvous), postés dans une période allant du 10 mars au 11 mai 2020. Elle correspond à la semaine précédant le confinement jusqu'à la date du déconfinement en France.

Un second filtre textuel a ensuite été appliqué sur les tweets dans le but de conserver uniquement les messages mentionnant les mots "masque" ou "masques". Ceci permet d'obtenir un corpus de tweets liés au coronavirus et aux masques.

Pour chaque tweet, les proportions de sentiments positif ou négatif exprimées ont été estimées par la méthode d'analyse de sentiment de Microsoft Azure Cognitive Services [1]. Trois séries temporelles bimensuelles ont pu être construites. La première série temporelle est constituée du volume cumulé de tweets postés au sujet des masques, toutes les douze heures. Elle correspond à l'évolution de l'importance de ce sujet parmi les discussions des internautes. Les deux autres séries temporelles sont constituées des proportions de messages avec un sentiment positif et négatif, toutes les douze heures. Ces deux courbes visent à créer un indicateur de la perception des utilisateurs concernant les sujets connexes aux masques. Elles ne sont pas des indicateurs de proportions d'utilisateurs pour ou contre le port du masque. En effet, un tweet peut exprimer une indignation (identifiée comme un sentiment négatif par la méthode) concernant la non-obligation de porter le masque dans l'espace public.



Graphique 1 – Evolution du volume de tweets et de la perception des utilisateurs sur la question des masques.

Les pointillés verticaux indiquent des événements susceptibles d'avoir impacté la perception de ce sujet par le public

RÉSULTATS

773 115 messages en français liés au coronavirus ont été extraits de Twitter entre le 10 mars au 11 mai. La détection des termes liés aux masques a permis d'isoler 18 508 messages, soit 2,4% du corpus total. Les dates importantes dans le débat sur le masque sont identifiables à partir des évolutions des trois séries temporelles (Graphique 1). Au début de la crise sanitaire, en même temps que l'instauration du confinement, il est décidé que les masques seront en priorité pour le personnel soignant et le ministre de la Santé, Olivier Véran, annonce que la France dispose d'un stock assez important pour eux. Ceci est rapidement contesté et, à partir du 22 mars, des premiers débats émergent quant à la responsabilité de la pénurie. A compter de cette date, une hausse importante de messages négatifs sur le sujet des masques est observée, passant de 29% à 45% en deux semaines. Ce pic de 45% a lieu la semaine du 5 avril, durant laquelle le port obligatoire d'un masque, même alternatif (c'est-à-dire autre que FFP2), a été recommandé par l'Académie Nationale de Médecine.

Sur les trois dernières semaines de confinement, le volume de messages s'est accru avec 12 245 messages (deux tiers de notre corpus final). La proportion de nouveaux messages négatifs augmente tandis que le nombre de tweets positifs diminue, en passant de 16% à 12%. Ces tendances correspondent aux inquiétudes grandissantes des Français à l'approche du déconfinement, à l'obligation du port du masque est obligatoire dans les transports et commerces.

CONCLUSION

La question des masques a pris une place grandissante au cours de la période de confinement. Bien qu'en pénurie et jugé inutile au début de la crise sanitaire, le masque a progressivement été accepté par la majorité de la population comme essentiel pour lutter contre l'épidémie. L'analyse proposée dans cette étude suggère des conclusions similaires. Le volume de contenus échangés sur le sujet et les fluctuations de perceptions sont directement corrélés aux annonces et événements relatifs à la problématique des masques.

Ces résultats souffrent de limites inhérentes à toutes études réalisées à partir des réseaux sociaux. La représentativité d'abord, puisque seuls 34% des internautes français utilisent activement Twitter [2]. Une deuxième limite est liée à l'imperfection de méthodes de fouilles de texte comme l'analyse de sentiment. Bien que validées scientifiquement, une telle méthode ne peut refléter l'opinion complète d'un individu sur une question complexe comme celle que nous traitons ici.

Ce travail démontre néanmoins si la possibilité de suivre l'impact d'une politique de santé publique et sa perception par une partie de la population en utilisant des données de vie réelle issues de témoignages sur Twitter.

RÉFÉRENCES

[1] <https://azure.microsoft.com/fr-fr/services/cognitive-services/text-analytics/#features>.

[2] Hootsuite & We Are Social. Digital 2020: Global Digital Overview. DataReportal – Global Digital Insights <https://datareportal.com/reports/digital-2020-global-digital-overview>.

Application de tracking StopCovid : identification des thématiques de discussions et des opinions par l'analyse des réseaux sociaux



Paméla Voillot, Pierre Arwidson, Pierre Foulquié, Anne-Juliette Serry, Adel Mebarki, Nathalie Texier, Stéphane Schück

Kap Code

APPLICATION DE TRACKING STOPCOVID : IDENTIFICATION DES THÉMATIQUES DE DISCUSSIONS ET DES OPINIONS PAR L'ANALYSE DES RÉSEAUX SOCIAUX

Paméla Voillot¹, Pierre Arwidson², Pierre Foulquié¹, Anne-Juliette Serry², Adel Mebarki¹, Nathalie Texier¹, Stéphane Schück¹

¹Kap Code, 28 rue d'Enghien 75010 Paris, France
²Santé publique France, 12, rue du Val d'Osne, 94 415 Saint-Maurice, France



INTRODUCTION

L'application StopCovid, mise en ligne le 2 juin 2020, permet de prévenir les individus qui ont été proches d'une personne testée positive au Covid-19 pendant au moins 15 minutes et à moins d'un mètre [1]. Celle-ci est téléchargeable sur la base du volontariat. Le fonctionnement de l'application a été approuvé par la CNIL et le code source est public. Malgré ces précautions, cette application divise la population. L'application a été téléchargée plus de 2,3 millions de fois à la mi-août. À ce jour, 1 169 usagers se sont déclarés positifs sur l'application permettant de repérer 72 contacts à risque.

Les réseaux sociaux sont un espace privilégié pour plus de 32 millions d'internautes. Ces plateformes occupent une place grandissante comme lieu d'expression où les internautes échangent sur leurs santé, leurs préoccupations et où la parole est libérée [2].

L'identification des opinions et des thématiques de discussions des internautes autour de l'application durant le confinement pourrait aider à mieux appréhender le comportement de la population envers cet outil numérique en post-confinement.

MATÉRIEL ET MÉTHODES

L'étude a été menée via l'outil **Detect'it**. Les messages associés à l'application StopCovid (dataset StopCovid) ont été extraits de Twitter entre le 13/04/2020 et le 06/05/2020.

La méthode d'analyse de sentiment de Microsoft Azure Cognitive Services a été utilisée pour identifier la perception générale des messages vis-à-vis de StopCovid (neutre, négative ou positive).

Par la suite et afin de caractériser le contenu de ces messages Twitter, l'algorithme Bitem TopicModel (BTM) a été appliqué sur l'ensemble du corpus afin de permettre une identification automatique des différentes thématiques de discussions abordées [3].

RÉSULTATS

Le dataset StopCovid contenait 7 110 tweets rédigés par 4 248 internautes différents avec une majorité de jeunes postants (moyenne = 27 ans) (figure 1).

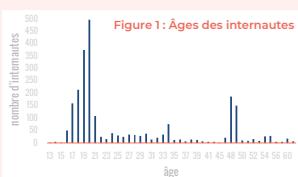


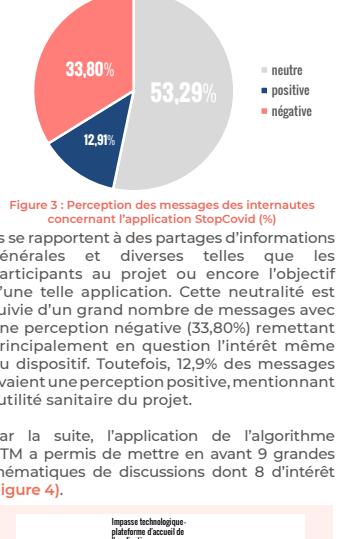
Figure 1: Âges des internautes



Figure 2 : Perception des messages des internautes concernant l'application StopCovid

Ils se rapportent à des partages d'informations générales et diverses telles que les participants au projet ou encore l'objectif d'une telle application. Cette neutralité est suivie d'un grand nombre de messages avec une perception négative (33,80%) remettant principalement en question l'intérêt même du dispositif. Toutefois, 12,9% des messages avaient une perception positive, mentionnant l'utilité sanitaire du projet.

Par la suite, l'application de l'algorithme BTM a permis de mettre en avant 9 grandes thématiques de discussions dont 8 d'intérêt (figure 4).



Réseaux sociaux et coronavirus : caractérisation des contenus échangés en France sur Twitter pendant la crise sanitaire du Covid-19



Pierre Foulquié, Léa Châteauneuf, Paméla Voillot, Simon Renner, Adel Mebarki, Nathalie Texier et Stéphane Schück

Kap•Code

RÉSEAUX SOCIAUX ET CORONAVIRUS : CARACTÉRISATION DES CONTENUS ÉCHANGÉS EN FRANCE SUR TWITTER PENDANT LA CRISE SANITAIRE DU COVID-19

Pierre Foulquié¹, Léa Châteauneuf¹, Paméla Voillot¹, Simon Renner¹, Adel Mebarki¹, Nathalie Texier¹ et Stéphane Schück¹

¹Kap Code, 28 rue d'Engleien 75010 Paris, France

INTRODUCTION

Suite aux premiers cas de coronavirus enregistrés sur son territoire en février 2020, la France a pris différentes mesures sanitaires pour répondre à la crise mondiale. Ces mesures, dont le confinement mis en place le 17 mars, ont impacté le quotidien des français de façon inédite. Les réseaux sociaux, notamment Twitter, source reconnue de données de vie réelle, ont permis aux français de communiquer et d'échanger des informations pendant cette crise. Cette étude infodémiologique se propose d'étudier la nature et l'évolution des contenus échangés sur twitter à partir de plusieurs méthodes de fouille de texte.

MATÉRIEL ET MÉTHODES

Un corpus de tweets français liés au Covid-19 a été constitué grâce à l'API Twitter. Les tweets extraits contenaient des mots-clés ou des hashtags évocateurs du coronavirus (#coronavirusfrance, #covid19fr etc.) et du confinement (#restezchezvous). Trois périodes d'intérêt ont été définies : du 10 au 31 mars, le mois d'avril et le mois de mai.

Le contenu des tweets est caractérisé par trois méthodes et les résultats de chaque période sont comparés.

Un modèle de sujet adapté aux textes courts (*bitem topic model* [1]) a été utilisé pour déterminer les thématiques des discussions. Ce modèle classe les messages selon les thématiques qu'ils abordent. Celles-ci se présentent sous la forme de listes de mots apparaissant ensemble dans les messages. Une classification manuelle des adresses web partagées a été opérée, basée sur la nature du site internet (presse, site gouvernemental etc.). Les proportions de ces regroupements dans le total des contenus web partagés est étudiée, ainsi que leur évolution au cours des trois périodes d'intérêt, dans le but de distinguer les types de contenus échangés.

Une identification des terminologies du dictionnaire MedDRA (*Medical Dictionary for Regulatory Activities*), enrichie de vocabulaire patient [2], a permis d'extraire le vocabulaire médical et les éventuels symptômes du coronavirus et du confinement. Les terminologies détectées sont ensuite regroupées par catégories médicales, grâce à la structure arborescente du dictionnaire MedDRA. Par exemple, « expectorer » et « tousser » sont regrouper dans la catégorie « Toux ».

RÉSULTATS

Le corpus constitué via l'API Twitter contient 2,5 millions de tweets : 933 481 en mars, 833 108 en avril et 774 778 en mai. La figure 1 montre l'évolution temporelle hebdomadaire nombre de messages. Le pic à la gauche du graphique correspond à la première semaine de confinement, qui a généré beaucoup de contenu. Le tableau 1 présente les cinq thèmes les plus abordés, chaque mois. Au mois de mars, ils correspondent à des inquiétudes et difficultés liées au confinement (37% des tweets), ainsi que des commentaires de la gestion de la crise par l'État (27%). Ces thèmes persistent en avril. Ils sont accompagnés par des thématiques plus positives décrivant des solidarités (4,68%) et des idées d'occupations (8,80%). Au mois de mai, les conditions du déconfinement sont largement discutées, avec par exemple la réouverture des écoles (18,09%) et le port du masque (22,55%).



Figure 1 – Évolution hebdomadaire du volume de tweets

Thème	Mars			Avril			Mai		
	%	P	%	P	%	P	%	P	
Difficultés liées au confinement	37,48%	P	29,43%	P	22,55%	P			
Non conformité aux mesures de santé publique	27,00%	P	15,64%	P	10,09%	P			
Intervention de l'État	13,52%	P	10,54%	P	10,44%	P			
Positiver face au confinement	9,16%	P	8,80%	P	9,26%	P			
Le confinement n'est pas si mal	6,37%	P	6,37%	P	5,37%	P			
Hébergement de vidéos	20,00%	P	19,25%	P	23,92%	P			
Presse quotidienne nat.	13,76%	P	15,78%	P	14,49%	P			
Revue sociale	12,59%	P	10,03%	P	10,45%	P			
Presse quotidienne rég.	1,31%	P	6,92%	P	8,47%	P			
Site Service Public	0,50%	P	0,52%	P	0,47%	P			

Tableau 1
Thèmes de discussion abordés (haut)
et type de sites partagés (bas)

Terminologie	Mars			Avril			Mai		
	Concept	P	Concept	P	Concept	P	Concept	P	
Mort	9,33%	P	Mort	11,06%	Mort	10,26%			
Ennui	2,82%	P	Ennui	2,25%	Anxiété	1,73%			
Faiblesse	1,04%	P	Faiblesse	1,20%	Faiblesse	1,09%			
Humeur dépressive	1,98%	P	Humeur dépressive	1,77%	Dépression	1,77%			
Crise	1,29%	P	Usage abusif de substances	1,29%	Étouffement	1,29%			

Les sites de presse et d'informations constituaient un tiers des adresses web échangées en mars et en avril, et plus de la moitié en mai. Youtube, avec des vidéos d'ordre informatif, humoristique ou créatif, est le site le plus partagé en mars (20% des liens). Cette proportion diminue à chaque période. Une tendance similaire est constatée pour les réseaux sociaux (autres que Twitter). Les sites de services publics sont présents au mois de mars, notamment du fait des décisions de confinement, du partage des gestes barrières et d'attestation de sorties. Cette proportion diminue dès avril. À chaque période, des terminologies médicales sont identifiées dans environ 10% des tweets. Les catégories les plus représentées sont regroupées dans le tableau 2. Les terminologies liées à la mort proviennent des bilans journaliers sur l'évolution de l'épidémie et relayées par les internautes. Si peu de symptômes du coronavirus sont observés, des termes évocateurs de troubles psychologiques sont identifiés : Anxiété, Ennui, et Dépression. Des comportements à risque sont également identifiables par les catégories Dépendance et Usage abusif de substances. Ces résultats suggèrent un effet du climat de crise et du confinement sur les individus.

CONCLUSION

L'étude du contenu échangé sur Twitter de mars à mai 2020 permet de qualifier l'usage dont en ont fait les français pendant la crise sanitaire. Les utilisateurs partagent leurs inquiétudes, leurs opinions, et partagent des supports d'information. Ils partagent également des troubles, notamment psychologiques et imputables au confinement. Chaque période analysée contient une ou plusieurs décisions de santé publique majeures (confinement, prolongation du confinement, déconfinement) ayant influencé le contenu des échanges. Ces résultats souffrent de limites inhérentes à toutes études réalisées à partir des réseaux sociaux. La représentativité d'abord, puisque seuls 34% des internautes français utilisent activement Twitter [3]. Dans un contexte d'épidémie, les réseaux sociaux sont néanmoins des indicateurs du niveau d'information de la population. Les analyser permet donc d'identifier et de répondre à ses attentes et inquiétudes et d'orienter les politiques publiques sanitaires.

RÉFÉRENCES

- [1] Yan, X., Guo, J., Lan, Y., & Cheng, X. (2013, May). A bitem topic model for short texts. In Proceedings of the 22nd international conference on World Wide Web (pp. 1445-1456).
- [2] Abdellaoui, R., Schück, S., Texier, N., & Burgun, A. (2017). Filtering entities to optimize identification of adverse drug reaction from social media: how can the number of words between entities in the messages help? JMIR public health and surveillance, 3(2), e36.
- [3] Hootsuite & We Are Social. Digital 2020: Global Digital Overview. DataReportal – Global Digital Insights <https://datareportal.com/reports/digital-2020-global-digital-overview>



Studying fake news on Twitter during the COVID-19 pandemic in France



Gedik Anais; Foulquié Pierre; Renner Simon; Mebarki Adel; Texier Nathalie;
Schück Stéphane

Studying fake news on Twitter during the COVID19 pandemic in France

Gedik A¹; Foulquié P¹; Renner S¹; Mebarki A¹; Texier N²; Schück S¹

¹Kap Code, Paris, France

INTRODUCTION

From the beginning of the global Covid-19 pandemic, misinformation have risen and have been spreaded on different social media platforms such as Facebook or Twitter. In particular, the latter has gathered an 8% increase of its daily users over a global scale between the end of 2019 and the beginning of 2020¹. The epidemic of misinformation is a global issue and disrupts public health. "We're not just fighting an epidemic: we're fighting an infodemic", said WHO Director-General Tedros Adhanom Ghebreyesus at the Munich Security Conference on February 15.

Social networks, especially Twitter, are a well-known source of real-world and real-time data [2]. The nature of shared information on these platforms can be of all kinds, true or false, and the accounts sharing it are also diverse, belonging to official sources and lambda citizens.

Covid-19 has been an important topic covered on Twitter that has given French people the chance to share their opinions on various matters related to it. Even though this social network has helped to create links between the users, it led to a propagation and a spread, consciously or not, of misinformation, commonly known as fake news.

The goal of our study was to identify and characterize fake news shared on Twitter related to Covid-19 pandemic.

METHODOLOGY

The dataset of this study is constituted of active French accounts during the pandemic. Through the Twitter API, we were able to extract French tweets related to Covid-19 between March and June 2019 using popular hashtags and key words such as "#coronavirusfrance" or "#confinement". The dataset was cleaned to keep only tweets that have been retweeted and/or quoted at least once.

A specific list of word combinations (unigram or n-grams) enabled the identification of tweets reporting false information. Elements of this list were gathered from words, expressions and tweet contents found in fact-checking resources released by media associations such as *Les décodeurs du monde* or *AFP*². A primary filter was applied using this list, however with the detection of some false positives being biased, especially those concerning miracle cures or treatments such as hydroxychloroquine which can help fight the virus. We applied a second filter based on disinformation lexical fields ("fake", "conspiracy", etc.). In other words, a tweet spreading a fake news will absolutely contain at least one element of our two lists mentioned above.

Once the tweets relating to fake news were identified, fake news were grouped manually into clusters depending on the subject covered in the tweets. With this step, an analysis of tweets contents was performed. These clusters also made it possible to determine twitter accounts involved in the transmission of each fake news. Every twitter account was weighted on its popularity and retweeting rate. To detect users' communities in each fake news group the Louvain clustering algorithm [3] was used.

For each fake news, a propagation model has been built in order to observe and analyze the exchanges between the different user accounts and detected communities. Retweets and quotes between users were represented by a network graph where each node represented a user and each edge an interaction (retweets or quote).

In this study, the analysis of the propagation network was done on the biggest group of fake news. Finally, a categorization of the different clusters of users was conducted by analyzing the nature of their exchanges as well as the tweet contents (articles, hashtags, shared URLs, videos..).

RESULTS

2.5 million tweets related to Covid-19 were extracted and analyzed, of which twenty fake news were deducted. These latter were identified and clustered into 5 groups (figure 1) thanks to the association rules of lexical fields such as "invox/Buzyn/chloroquine" or "creation/virus/lab".

The first group, corresponding to the biggest proportion of fake news (39%) (figure 1), refers to the former French Minister of Health, Agnès Buzyn, and her husband, the former director of the National Institute of Health and Medical Research, Yves Lévy, the decision makers of not to generalize the massive prescription of hydroxychloroquine as a way of preventing serious cases and deaths. Tweets of this group accuse them of sabotaging French infectiologist Didier Raoult's work on hydroxychloroquine with suspicion of conflicts of interest. Analyzing tweet contents, especially shared URLs, we found that more than half of the tweets deny this fake news by sharing articles from *Les décodeurs du monde*⁴.

The second group identified (21%) (figure 1) mentions the creation of the virus in laboratories. This group is composed by several subgroups covering the potential origins of the virus. More than half of the tweets mention the creation of the virus with the laboratory P4 in the Institute of virology in Wuhan-China, built with the help of France, particularly Yves Lévy. Then comes that the Pasteur Institute would have launched the virus and would own a patent, and finally that of a Harvard researcher, Charles Lieber, who allegedly sold the virus to China.

The category "Other Fake News" (figure 1) corresponds to fake news represented respectively with a proportion of less than 5%, such as the fact that the heat can eradicate the virus, with the heat waves of summer or by drinking hot beverages (figure 3).

Focusing on the first group of fake news, more than 150 users' communities were found using the Louvain algorithm. Only accounts retweeted or quoted more than 5 times in each community were kept to build the network graph. Analyzing the propagation network of fake news and the tweet contents enabled the categorization of accounts and different communities. The transmission network (Figure 2) reveals that the two protagonists of this fake news, Agnès Buzyn and Didier Raoult, hold the most retweeted accounts surrounded by media organizations accounts ("lemonde", "décodeurs"). These latter interact with lambda citizens' accounts and would be part of the community denying the fake news relying on the fact checking article URLs shared on their tweets. Many users' clusters are built around two central accounts belonging either to a journalist or a media source such as BFM. Small communities are on the periphery and seem to have no interaction with the media sphere. We also detected small communities politically involved in spreading the fake news coming from certain political parties and politically engaged users. These different communities, after analyzing the content of their tweets are separated into two categories. Those that are at the center of the network like the "Pcrouquet" accounts, supporting the Professor Raoult, posting hashtags and mentions related to Didier Raoult, and mainly made of people that share the same nationalist ideologies analyzing its account description. And those that are on the periphery such as the community composed by the account *UPR_Rhone*.

If the typology of each cluster's twitter accounts is quite heterogeneous in terms of influence (number of post retweets, number of followers, etc.), it is homogeneous in terms of the tweet's content (hashtags, shared URLs, etc.)

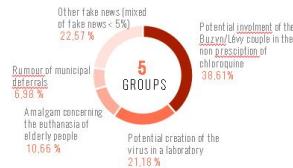


Figure 1 : Proportion of fake news groups detected



Figure 3 : Example of fake news tweets

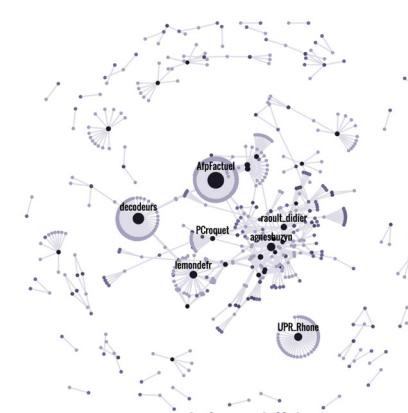


Figure 2: Propagation network of the first group of fake news

CONCLUSION AND PERSPECTIVES

Analyzing the proportion and model the propagation network of fake news relayed on Twitter can help fight misinformation. The study of networks has highlighted different clusters and accounts spreading fake news, with various degrees of influence.

It is then possible to quantify the presence of misinformation on social networks and its impact. These methods also allow to reveal the form and the nature of the fake news posts by analyzing their contents. Thus, we could determine the profile of accounts spreading fake news (bots...) and all the rhetorical elements used.

This kind of algorithm, which would prospectively detect fake news could help health authorities fight against them.

REFERENCES

- [1] Brennen S., Simon F.M., Howard P., Nielsen R. Types, Sources, and Claims of COVID-19 Misinformation. Reuters Institute for the Study of Journalism; Oxford, UK, 2020.
- [2] Ghosh P., Schwartz G., Narouze S. Twitter as a powerful tool for communication between pain physicians during COVID-19 pandemic. Reg. Anesth. Pain Med. .2020. doi: 10.1136/ramp-2020-101530.
- [3] V. D. Blondel, J. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," Journal of Statistical Mechanics, vol. 2008, no. 10, Article ID P10008, 2008.

¹ Le réseau des médias, INA, "Le confinement n'a pas révolutionné l'espace public numérique, mais il a accéléré sa transformation", 14 May 2020

² List of fake news shared by AFP fact checking media: AFP Factuel + Covid-19 : retrouvez ici tous nos articles de vérification , 17 February 2020

³ LeMonde, « Coronavirus et hydroxychloroquine : le couple Buzyn-Lévy cible de publications mensongères », 27 Mars 2020



Detection of Quality of Life Impact from Health-related Messages in Social Media



Tom Marty, Mickaïl Khadar, Simon Renner, Pierre Foulquié, Pamela Voillot, Adel Mebarki, Nathalie Texier and Stéphane Schück

14/06/2022 17:05

ML3 Detection of Quality of Life IMPACT in Health-Related Messages in Social MEDIA - ScienceDirect



ScienceDirect

Value in Health

Volume 23, Supplement 2, December 2020, Page S404

ML3 Detection of Quality of Life IMPACT in Health-Related Messages in Social MEDIA

T. Marty, M. Khadhar, S. Renner, P. Foulquié, P. Voillot, A. Mebarki, N. Texier, S. Schück

Show more ▾

Outline

Share

Cite

<https://doi.org/10.1016/j.jval.2020.08.043>

Under an Elsevier user license

[Get rights and content](#)

[Open archive](#)



Previous



Next

Objectives

Every day, people share health-related messages and opinions online. Understanding and analyzing these insights could be useful for medical applications. Quality of life is medically defined as individuals' perception of their place in life, in the cultural context and the value system in which they live, according to their objectives, standards and concerns. Mining social medias to understand the burden of pathologies and treatments is a new approach to better understand patients' concerns and unmet needs. This study proposes a NLP algorithm able to detect health-related quality of life (HRQoL) impacts on social media posts.

Methods

Messages relating a patient experience with a drug or disease were collected on French social medias, using a Web crawler. Two reviewers manually annotated messages following guidelines based on EQ-5D and SF-36 QoL measurements. Focus was put on physical, psychic, activity, relational and financial dimension of HRQoL. A first model detected impact presence. Then each impact was identified by a specific model. Impact's expressions were used to generate features. Other features were based on the message content, such as expressed sentiment, grammar and conjugation.



Étude des réseaux sociaux pour identifier les motifs de l'hésitation vaccinale anti-HPV : une étude infodémiologique



Simon Renner, Tom Marty, Pamela Voillot, Pierre Foulquie, Adel Mebarki, and Stéphane Schück

Kap Code

ÉTUDE DES RÉSEAUX SOCIAUX POUR IDENTIFIER LES MOTIFS DE L'HÉSITATION VACCINALE ANTI-HPV : UNE ÉTUDE INFODÉMILOGIQUE

Simon Renner¹, Tom Marty¹, Pamela Voillot¹, Pierre Foulquie¹, Adel Mebarki¹, et Stéphane Schück¹

¹Kap Code, 28 rue d'Englebert 75010 Paris, France

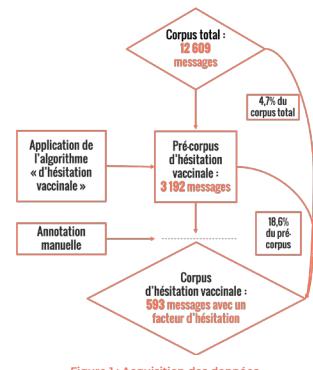
Evanex

INTRODUCTION

La recommandation vaccinale contre l'infection à papillomavirus humain (HPV) est un enjeu mondial de santé publique. En 2019, la couverture vaccinale de 21% place la France en dernière position européenne. Parallèlement, les réseaux sociaux sont un espace privilégié pour plus de 32 millions d'internautes actifs. Ces plateformes occupent une place grandissante comme lieu d'expression ou les internautes échangent sur leur santé, leurs préoccupations et souvent sur leur position vis-à-vis de la vaccination. Analyser et comprendre ces messages permettrait d'identifier les leviers associés à une amélioration de la couverture vaccinale anti-HPV.

MATÉRIEL ET MÉTHODES

L'étude a été menée via l'observatoire de la vaccination **EVANEX**. Les messages associés à la vaccination anti-HPV et écrits en français (dataset HPV) ont été extraits à partir de 23 forums médicaux et de réseaux sociaux, entre 2006 et 2019. Les termes d'extraction faisaient soit référence à un acte de vaccination contre le HPV « vaccin contre le papillomavirus » soit directement à un nom de produit (Gardasil® et Cervarix®). Une annotation manuelle de messages (présence ou non d'une hésitation) a permis la création d'un Gold standard visant à développer un algorithme de détection d'une hésitation vaccinale. Cet algorithme a ensuite été appliqué sur l'ensemble du dataset HPV. Un premier corpus d'analyse (pré-corpus d'hésitation vaccinale, Figure 1) a pu être constitué. Une annotation manuelle de l'ensemble de ce corpus d'intérêt a été effectuée afin d'identifier les différents facteurs d'hésitation (corpus d'hésitation vaccinale, Figure 1) et les regrouper par grandes thématiques d'hésitation.



RÉSULTATS

Le dataset HPV contenait 12 609 messages uniques rédigés par 5 117 internautes différents. Un pré-corpus de 3 192 messages propres à l'hésitation vaccinale a été déterminé algorithmiquement. L'analyse manuelle de ces 3 192 messages a permis de caractériser 593 messages d'hésitation vaccinale (18.6% du pré-corpus, 4.7% du dataset HPV) (Figure 1).

[L]e sa fait a peu près 1 an et demi que je ne suis plus vierge ma mère veut absolument que je fasse le vaccin HPV mais je ne ve pas lui dire que je ne suis plus vierge eske cela comporte un risque que je me fasse vacciner ?

Bonjour, j'ai 17 ans et j'ai déjà eu des rapports avec une seule personne, cette personne n'a eu des rapports avec moi aussi. J'en n'est pas été vacciné contre le cancer du col de l'utérus et j'aimerais savoir si c'est toujours possible ?

Figure 2 : Exemples de messages extraits

Les messages (Figure 2) ont été regroupés en 3 principaux groupes, selon les facteurs d'hésitation vaccinale exprimés. Le premier a pour motif l'influence de la vie sexuelle sur la vaccination (339 messages, 7.5% pré-corpus). Au sein de celui-ci, les adolescentes s'interrogent à la fois sur la nécessité de se faire vacciner mais aussi sur une éventuelle dangerosité du vaccin une fois sexuellement active (n=113; Figure 3) ou si un rapport advient entre deux injections (n=70). La crainte de l'apparition d'effets indésirables en raison de leur activité sexuelle est importante.

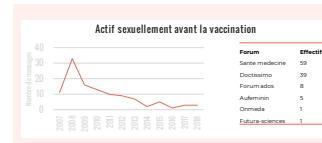


Figure 3 : Actif sexuellement - Volumétrie

être global (n=60, Figure 5) sur les modalités vaccinales : obligation vaccinale, efficacité, utilité, remboursement ...

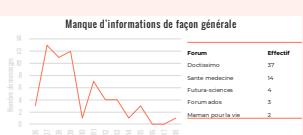
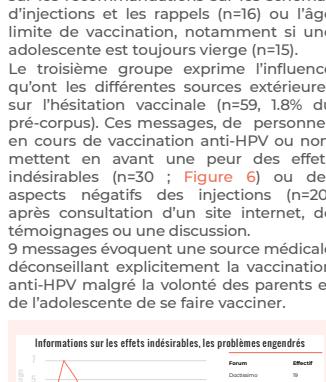


Figure 5 : Manque d'informations générale – Volumétrie



Patients Comments on Social Networks about Paracetamol Misuses



Avesta Roustamal; Anais Gedik; Pierre Foulquié; Paméla Voillot; Adel Mebarki; Nathalie Texier; Stéphane Schück

PATIENTS' COMMENTS ON SOCIAL NETWORKS ABOUT PARACETAMOL MISUSES

Avesta Roustamal¹; Anais Gedik¹; Pierre Foulquié¹; Paméla Voillot¹; Adel Mebarki¹; Nathalie Texier¹; Stéphane Schück¹

¹Kap Code, Paris, France

INTRODUCTION

Paracetamol is one of the Essential Medicines according to WHO. It is one of the most common painkiller but misuses and overdoses occur.

Recently, the ANSM has launched a public consultation in order to raise awareness of the toxicity risks linked to paracetamol in case of misuse. The social media are an important way for patients to share experiences regarding the misuse of paracetamol (Foulquié et al, 2019). These discussions on social media could be a potential new source of data to detect early potential signals. Within this context, Kap Code has developed Detec't an automatic tool for analyzing social media based on artificial intelligence and text mining methods that appears useful to analyze posts concerning paracetamol.

The main objective is to realize an infodemiology study of patients expressing on social media about paracetamol-containing products. More specifically, the goal is to detect misuse cases including overdose, unapproved indications and dependence discussed by web users about paracetamol.

METHODOLOGY

The posts in french language from social media sources were extracted with Detec't tool by detecting in social media discussions key words related to paracetamol containing products such as «Paracetamol», «Doliprane», «Actifed»... The posts were then, collected and aggregated. The project does not aim to collect adverse events as this data were not treated individually. The data were cleaned in order to apply the algorithms available with Detec't tools. These methods of analysis were used by ANSM in cooperation with Kap Code(Foulquié et al, 2018).

The number of posts was calculated by applying a Markov chain-based algorithm on posts in order to qualify the volume of activity (number of posts published by web users that can be weak, moderate or high) for each time period. This methodology allowed the identification of unusual activity period (that could correspond to alerts). The age and gender of web users were determined with the detection of regular expressions within the posts. A drug-intake algorithm based on a random forest method was applied on the corpus of posts in order to identify whether each drug identified in the post had been taken by the web user. The three next algorithms of misuse detection were applied on posts where the drug intake was detected. The overdose results were obtained with the detection of words from overdose lexical field and the digits in the posts were compared to the daily dose recommended. The identification of the reason of paracetamol use was performed by the detection of medical concepts discussed through MedDRA which was enriched with web user vocabulary. The unapproved indications were obtained by comparing the medical concepts (MedDRA) detected as the indication found in the posts and the approved indications for paracetamol products(described in the drug summary¹). Dependence was studied by the identification in the posts of dependence lexical field.

RESULTS

The number of posts extracted with the key words of only paracetamol containing products is 33 196 posts and published from 2003 and 2019 (March) on 18 social media sources (*), 17 070 web users dealt with paracetamol in their messages. The proportion of web users is 70% of women, 12% of men and 18% of unknown gender (FIGURE 1). The mean age of web users is 39 years old ± 20. The variable activity seen in the post volume over the study period shows a peak in post-volume activity between 2009 and 2012 (FIGURE 2). This spike may correspond with the withdrawal of dextropropoxyphene-containing medicines from the market in France. The number of posts tended to be higher during the winter months and lower during the summer and can be due to the increase of colds during this period. Regarding the misuse cases, the algorithm of unapproved indication detection identified 190 posts. The unapproved indications the most represented in the posts are Fatigue and Anxiety (0.28% of posts dealing with drug intake and quoting at least one medical concept) and Suicide (0.18%) as shown on the figure 3. The dependence of paracetamol was detected in 171 posts (1.33% of the posts detected with drug intake) and provided by 136 users (0.79% of users). Most of the posts detected with paracetamol dependence, deal with long term paracetamol treatment which induces the dependence and the associated risks. Some examples of posts of dependence are presented on figure 4. Reference to paracetamol overdose was made by 245 web users (which represents 1.45% of total web users) across 291 posts (2.26% of total corpus). The chart on the figure 5 shows the percentage of overdose above daily dose detected in the posts and the table introduces the drugs quoted for paracetamol overdose. The range of overdose described in the posts varies between 4.8 grams and 60 grams of paracetamol taken by web users. They describe an overdose for different reasons such as maximum daily dose unknown by the patients or intentionally substance abuse for recreational use or suicide attempt.

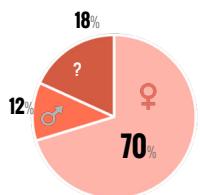


Figure 1 : Proportion of web users by genders

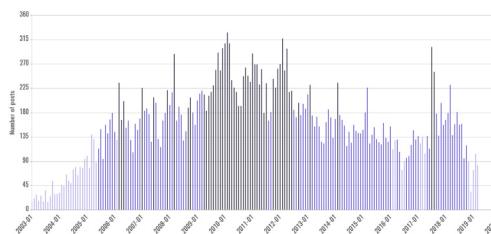


Figure 2 : Monitoring posts volume

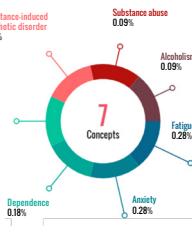
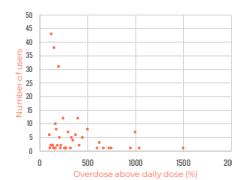


Figure 3 : Unapproved indications



Figure 4 : Examples of posts detected with dependence



«Hello, last night I swallowed 18g of Doliprane (I weight around 60kg and I'm a teenager). I have a slight abdominal cramp but nothing worst. What should I do ? What are the next symptoms ? Is it serious ?»

«Hello sunday I took at least 12g of Dafalgan, I can't stand I feel totally depressed. I know you are taking a risk but which one ? Thank you for your help »

Figure 5 : Paracetamol overdose

CONCLUSION AND PERSPECTIVES

This retrospective study permitted to highlight the importance to consider social media as a source for detecting early potential signals from real life data. Real communities are constituted, patients mostly sharing their experience, seeking for support or discussing drug misuse such as the use of paracetamol to reduce fatigue or anxiety (both 0.28% of the messages). Web users share also information about their drug addiction (1.33% of posts) and their overdose experience (291 posts). Thanks to these analyses, we will be able to identify the drug consumer behavior and the misuses as soon as possible. These analyses will be applied prospectively in order to identify the drug consumer behavior and the misuses as soon as possible and thus, to set up mitigation plan and to monitor risks.

REFERENCES

- Kürzinger ML, Schück S, Texier N, et al. Web-based signal detection using medical forums data in France. J Med Internet Res (2018)
- Van Stekelenborg J, Ellenuis J, Maskell S, et al. Recommendations for the Use of Social Media in Pharmacovigilance: Lessons from IMI WEB-RADR Drug Safety (2019)
- P. Foulquié, P. Voillot, C. Faviez, A. Mebarki, X. Chen, N. Texier, S. Schück, Étude de l'usage du méthylphénidate sur les réseaux sociaux, Revue d'Épidémiologie et de Santé Publique, Volume 66, Supplement 4, (2018)



Physicians' Perceptions of the Use of a Chatbot for Information Seeking: Qualitative Study



Koman J, Fauvette K, Schuck S, Texier N, Mebarki A

JOURNAL OF MEDICAL INTERNET RESEARCH

Koman et al

Original Paper

Physicians' Perceptions of the Use of a Chatbot for Information Seeking: Qualitative Study

Jason Koman^{1,2,3*}, PhD; Khristina Fauvette⁴, MD; Stéphane Schuck^{1*}, MD; Nathalie Texier¹, MD; Adel Mebarki¹, MD

¹Kap Code, Paris, France

²CNRS, PASSAGES, Bordeaux, France

³Bordeaux Population Health Research Center, University of Bordeaux, Inserm, Bordeaux, France

⁴Sanofi Aventis, Gentilly Cedex, France

*these authors contributed equally

Corresponding Author:

Khristina Fauvette, MD

Sanofi Aventis

82, avenue Raspail

Gentilly Cedex, 94255

France

Phone: 33 772219558

Email: kchristina.fauvette@sanofi.com

Abstract

Background: Seeking medical information can be an issue for physicians. In the specific context of medical practice, chatbots are hypothesized to present additional value for providing information quickly, particularly as far as drug risk minimization measures are concerned.

Objective: This qualitative study aimed to elicit physicians' perceptions of a pilot version of a chatbot used in the context of drug information and risk minimization measures.

Methods: General practitioners and specialists were recruited across France to participate in individual semistructured interviews. Interviews were recorded, transcribed, and analyzed using a horizontal thematic analysis approach.

Results: Eight general practitioners and 2 specialists participated. The tone and ergonomics of the pilot version were appreciated by physicians. However, all participants emphasized the importance of getting exhaustive, trustworthy answers when interacting with a chatbot.

Conclusions: The chatbot was perceived as a useful and innovative tool that could easily be integrated into routine medical practice and could help health professionals when seeking information on drug and risk minimization measures.

(*J Med Internet Res* 2020;22(11):e15185) doi: [10.2196/15185](https://doi.org/10.2196/15185)

KEYWORDS

health; digital health; innovation; conversational agent; decision support system; qualitative research; chatbot; bot; medical drugs; prescription; risk minimization measures



Concerns Discussed on Chinese and French Social Media During the COVID-19 Lockdown : Comparative Infodemiology Study Based on Topic Modelling



Stéphane Schück; Pierre Foulquié; Adel Mebarki; Carole Faviez; Mickaïl Khadhar; Nathalie Texier; Sandrine Katsahian; Anita Burgun; Xiaoyi Chen

JMIR FORMATIVE RESEARCH

Schück et al

Original Paper

Concerns Discussed on Chinese and French Social Media During the COVID-19 Lockdown: Comparative Infodemiology Study Based on Topic Modeling

Stéphane Schück¹, MD; Pierre Foulquié¹, MSc; Adel Mebarki¹, MSc; Carole Faviez², MSc; Mickaïl Khadhar¹, MSc; Nathalie Texier¹, PharmD; Sandrine Katsahian^{2,3}, PhD, MD; Anita Burgun^{2,4,5,6}, PhD, MD; Xiaoyi Chen², PhD

¹Kap Code, Paris, France

²Centre de Recherche des Cordeliers, INSERM, Sorbonne Université, Université de Paris, Paris, France

³Unité d'Épidémiologie et de Recherche Clinique, Hôpital européen Georges Pompidou, Assistance Publique - Hôpitaux de Paris, Paris, France

⁴Département d'informatique médicale, Hôpital européen Georges Pompidou, Assistance Publique - Hôpitaux de Paris, Paris, France

⁵Département d'informatique médicale, Hôpital Necker-Enfants Malades, Assistance Publique - Hôpitaux de Paris, Paris, France

⁶Paris Artificial Intelligence Research Institute, Paris, France

Corresponding Author:

Xiaoyi Chen, PhD

Centre de Recherche des Cordeliers

INSERM

Sorbonne Université, Université de Paris

15 Rue de l'école de médecine

Paris, F-75006

France

Phone: 33 171196369

Email: xiaoyi.chen@inserm.fr

Abstract

Background: During the COVID-19 pandemic, numerous countries, including China and France, have implemented lockdown measures that have been effective in controlling the epidemic. However, little is known about the impact of these measures on the population as expressed on social media from different cultural contexts.

Objective: This study aims to assess and compare the evolution of the topics discussed on Chinese and French social media during the COVID-19 lockdown.

Methods: We extracted posts containing COVID-19-related or lockdown-related keywords in the most commonly used microblogging social media platforms (ie, Weibo in China and Twitter in France) from 1 week before lockdown to the lifting of the lockdown. A topic model was applied independently for three periods (prelockdown, early lockdown, and mid to late lockdown) to assess the evolution of the topics discussed on Chinese and French social media.

Results: A total of 6395; 23,422; and 141,643 Chinese Weibo messages, and 34,327; 119,919; and 282,965 French tweets were extracted in the prelockdown, early lockdown, and mid to late lockdown periods, respectively, in China and France. Four categories of topics were discussed in a continuously evolving way in all three periods: *epidemic news and everyday life*, *scientific information*, *public measures*, and *solidarity and encouragement*. The most represented category over all periods in both countries was *epidemic news and everyday life*. *Scientific information* was far more discussed on Weibo than in French tweets. Misinformation circulated through social media in both countries; however, it was more concerned with the virus and epidemic in China, whereas it was more concerned with the lockdown measures in France. Regarding *public measures*, more criticisms were identified in French tweets than on Weibo. Advantages and data privacy concerns regarding tracing apps were also addressed in French tweets. All these differences were explained by the different uses of social media, the different timelines of the epidemic, and the different cultural contexts in these two countries.

Conclusions: This study is the first to compare the social media content in eastern and western countries during the unprecedented COVID-19 lockdown. Using general COVID-19-related social media data, our results describe common and different public reactions, behaviors, and concerns in China and France, even covering the topics identified in prior studies focusing on specific



Assessing Patient Perceptions and Experiences of Paracetamol in France: An Infodemiology Study Using Social Media Data Mining



Stéphane Schück; Avesta Roustamal; Anaïs Gedik; Paméla Voillot; Pierre Foulquié; Catherine Penfornis; Bernard Job

JOURNAL OF MEDICAL INTERNET RESEARCH

Schück et al

Original Paper

Assessing Patient Perceptions and Experiences of Paracetamol in France: Infodemiology Study Using Social Media Data Mining

Stéphane Schück¹, MSc, MD; Avesta Roustamal¹, MSc; Anaïs Gedik¹, MSc; Paméla Voillot¹, MSc; Pierre Foulquié¹, MSc; Catherine Penfornis², MD; Bernard Job², MD

¹Kap Code, Paris, France

²Sanofi, Paris, France

Corresponding Author:

Avesta Roustamal, MSc
Kap Code
28 Rue d'Enghien
Paris,
France
Phone: 33 9 72 60 57 64
Email: avesta.roustamal@kapcode.fr

Abstract

Background: Individuals frequently turn to social media to discuss medical conditions and medication, sharing their experiences and information and asking questions among themselves. These online discussions can provide valuable insights into individual perceptions of medical treatment, and increasingly, studies are focusing on the potential use of this information to improve health care management.

Objective: The objective of this infodemiology study was to identify social media posts mentioning paracetamol-containing products to develop a better understanding of patients' opinions and perceptions of the drug.

Methods: Posts between January 2003 and March 2019 containing at least one mention of paracetamol were extracted from 18 French forums in May 2019 with the use of the Detec't (Kap Code) web crawler. Posts were then analyzed using the automated Detec't tool, which uses machine learning and text mining methods to inspect social media posts and extract relevant content. Posts were classified into groups: Paracetamol Only, Paracetamol and Opioids, Paracetamol and Others, and the Aggregate group.

Results: Overall, 44,283 posts were analyzed from 20,883 different users. Post volume over the study period showed a peak in activity between 2009 and 2012, as well as a spike in 2017 in the Aggregate group. The number of posts tended to be higher during winter each year. Posts were predominantly made by women (14,897/20,883, 71.34%), with 12.00% (2507/20,883) made by men and 16.67% (3479/20,883) by individuals of unknown gender. The mean age of web users was 39 (SD 19) years. In the Aggregate group, pain was the most common medical concept discussed (22,257/37,863, 58.78%), and paracetamol risk was the most common discussion topic, addressed in 20.36% (8902/43,725) of posts. Doliprane was the most common medication mentioned (14,058/44,283, 31.74%) within the Aggregate group, and tramadol was the most commonly mentioned drug in combination with paracetamol in the Aggregate group (1038/19,587, 5.30%). The most common unapproved indication mentioned within the Paracetamol Only group was fatigue (190/616, with 16.32% positive for an unapproved indication), with reference to dependence made by 1.61% (136/8470) of the web users, accounting for 1.33% (171/12,843) of the posts in the Paracetamol Only group. Dependence mentions in the Paracetamol and Opioids group were provided by 6.94% (248/3576) of web users, accounting for 5.44% (342/6281) of total posts. Reference to overdose was made by 245 web users across 291 posts within the Paracetamol Only group. The most common potential adverse event detected was nausea (306/12843, 2.38%) within the Paracetamol Only group.

Conclusions: The use of social media mining with the Detec't tool provided valuable information on the perceptions and understanding of the web users, highlighting areas where providing more information for the general public on paracetamol, as well as other medications, may be of benefit.

(*J Med Internet Res* 2021;23(7):e25049) doi: [10.2196/25049](https://doi.org/10.2196/25049)

KEYWORDS

analgesic use; data mining; infodemiology; paracetamol; pharmacovigilance; social media; patient perception



Mapping and Modeling of Discussions Related to Gastrointestinal Discomfort in French-Speaking Online Forums: Results of a 15-Year Retrospective Infodemiology Study



Schäfer F, Faviez C, Voillot P, Foulquié P, Najm M, Jeanne JF, Fagherazzi G, Schück S, Le Nevé B

JOURNAL OF MEDICAL INTERNET RESEARCH

Schäfer et al

Original Paper

Mapping and Modeling of Discussions Related to Gastrointestinal Discomfort in French-Speaking Online Forums: Results of a 15-Year Retrospective Infodemiology Study

Florent Schäfer¹, MSc; Carole Faviez², MSc; Paméla Voillot², MSc; Pierre Foulquié², MSc; Matthieu Najm², MSc; Jean-François Jeanne¹, MSc; Guy Fagherazzi^{3,4}, MSc, PhD; Stéphane Schück², MSc, MD; Boris Le Nevé¹, PhD

¹Innovation Science and Nutrition, Danone Nutricia Research, Palaiseau, France

²Kap Code, Paris, France

³Deep Digital Phenotyping Research Unit, Department of Population Health, Luxembourg Institute of Health, Strassen, Luxembourg

⁴Center of Research in Epidemiology and Population Health, UMR 1018 Inserm, Institut Gustave Roussy, Paris-Sud Paris-Saclay University, Villejuif, France

Corresponding Author:

Florent Schäfer, MSc

Innovation Science and Nutrition

Danone Nutricia Research

RD 128 Avenue de la Vauve

Palaiseau, 91767

France

Phone: 33 1 69 35 70 00

Email: florent.schafer@danone.com

Abstract

Background: Gastrointestinal (GI) discomfort is prevalent and known to be associated with impaired quality of life. Real-world information on factors of GI discomfort and solutions used by people is, however, limited. Social media, including online forums, have been considered a new source of information to examine the health of populations in real-life settings.

Objective: The aims of this retrospective infodemiology study are to identify discussion topics, characterize users, and identify perceived determinants of GI discomfort in web-based messages posted by users of French social media.

Methods: Messages related to GI discomfort posted between January 2003 and August 2018 were extracted from 14 French-speaking general and specialized publicly available online forums. Extracted messages were cleaned and deidentified. Relevant medical concepts were determined on the basis of the Medical Dictionary for Regulatory Activities and vernacular terms. The identification of discussion topics was carried out by using a correlated topic model on the basis of the latent Dirichlet allocation. A nonsupervised clustering algorithm was applied to cluster forum users according to the reported symptoms of GI discomfort, discussion topics, and activity on online forums. Users' age and gender were determined by linear regression and application of a support vector machine, respectively, to characterize the identified clusters according to demographic parameters. Perceived factors of GI discomfort were classified by a combined method on the basis of syntactic analysis to identify messages with causality terms and a second topic modeling in a relevant segment of phrases.

Results: A total of 198,866 messages associated with GI discomfort were included in the analysis corpus after extraction and cleaning. These messages were posted by 36,989 separate web users, most of them being women younger than 40 years. Everyday life, diet, digestion, abdominal pain, impact on the quality of life, and tips to manage stress were among the most discussed topics. Segmentation of users identified 5 clusters corresponding to chronic and acute GI concerns. Diet topic was associated with each cluster, and stress was strongly associated with abdominal pain. Psychological factors, food, and allergens were perceived as the main causes of GI discomfort by web users.

Conclusions: GI discomfort is actively discussed by web users. This study reveals a complex relationship between food, stress, and GI discomfort. Our approach has shown that identifying web-based discussion topics associated with GI discomfort and its perceived factors is feasible and can serve as a complementary source of real-world evidence for caregivers.



Health-Related Quality Of Life (Hrql) Of Cancer Patients Treated With Immunotherapy: Use Of Social Media In French Language To Explore Relevance Of Concepts Covered By Cancer Generic QoL Measures



C. Faviez, B. Bennett, P. Voillot, S. Schück, B. Falissard, C.T. Tzourio, P. Foulquié, A.F. Gaudin, H. Lemasson, V. Grumberg, L. McDonald, F.E. Cotte

Objectives

Immune-checkpoint inhibitors (ICI) are emerging as the standard of care for many cancers. However, HRQoL of patients treated with ICI in the real-world remain largely unknown. Social media is increasingly used by patients to express their views about their illness and treatment experience. The objective was to assess the conceptual similarity between cancer generic HRQoL measures and cancer patients' experience described in social media.

Methods

Patient messages were retrieved from 19 different French social media sources between Jan-2011 and Aug-2018. Messages of interest were extracted using automatic processes and manual searches. Extracted messages were analysed by two independent reviewers. HRQoL statements were classified according to predefined HRQoL dimensions in the *European Organisation for Research and Treatment of Cancer* (EORTC) QLQ-C30 and FACT-G.

Results

Overall, 137 ICI-treated patients posted HRQoL-related messages on social media. Cited dimensions of HRQoL were global health (115 patients), symptoms (76; mean: 2.1 per patient), emotion (49), role (22), physical functioning (13), professional situation (2) and cognitive state (2). Both the QLQ-C30 and FACT-G cover global health dimension expressed by patients in their messages. A total of 13 symptoms were identified from messages; the QLQ-C30 captured 5, whereas the FACT-G captured 3. Ten themes related to emotional functioning were identified, 2 were covered by the QLQ-C30 and 3 by the FACT-G. Seven themes were retrieved describing patients' role; the QLQ-C30 captured 5, whereas the FACT-G captured 3. In terms of physical functioning, 4 themes were identified; the QLQ-C30 captured all themes, whereas the FACT-G captured 1.

Conclusions

Conceptual similarity between HRQoL themes from social media and the most commonly used generic HRQoL measures was generally suboptimal. Careful interpretation is required due to the relatively low sample size. A larger study on social media would be needed to assess correlation with tumor specific questionnaires of HRQoL.



JMIR Publications
Advancing Digital Health & Open Science



Exploring the Health-Related Quality of Life of Patients Treated With Immune Checkpoint Inhibitors: Social Media Study



Cotté FE, Voillot P, Bennett B, Falissard B, Tzourio C, Foulquié P, Gaudin AF, Lemasson H, Grumberg V, McDonald L, Faviez C, Schück S

Original Paper

Exploring the Health-Related Quality of Life of Patients Treated With Immune Checkpoint Inhibitors: Social Media Study

François-Emery Cotté¹, PharmD, PhD, MPH; Paméla Voillot², MSc; Bryan Bennett³, PhD; Bruno Falissard^{4,5,6,7}, MD, PhD; Christophe Tzourio^{8,9}, MD, PhD; Pierre Foulquié², MSc; Anne-Françoise Gaudin¹, PharmD; Hervé Lemasson¹, MSc; Valentine Grumberg¹⁰, MSc; Laura McDonald¹¹, PhD; Carole Faviez², MSc; Stéphane Schück², MD

¹Bristol-Myers Squibb, Rueil-Malmaison, France

²Kap Code, Paris, France

³Bristol-Myers Squibb, Stockport, United Kingdom

⁴Paris-Sud University, Paris, France

⁵Paris-Descartes University, Paris, France

⁶AP-HP, Paris, France

⁷INSERM U1178, Paris, France

⁸Bordeaux University, Bordeaux, France

⁹Inserm U1219, Bordeaux, France

¹⁰UFR Pharmacy, Grenoble Alpes University, Grenoble, France

¹¹Bristol-Myers Squibb, London, United Kingdom

Corresponding Author:

Paméla Voillot, MSc

Kap Code

28 rue d'Enghien

Paris, 75010

France

Phone: 33 624058742

Email: pamela.voillot@kapcode.fr

Abstract

Background: Immune checkpoint inhibitors (ICIs) are increasingly used to treat several types of tumors. Impact of this emerging therapy on patients' health-related quality of life (HRQoL) is usually collected in clinical trials through standard questionnaires. However, this might not fully reflect HRQoL of patients under real-world conditions. In parallel, users' narratives from social media represent a potential new source of research concerning HRQoL.

Objective: The aim of this study is to assess and compare coverage of ICI-treated patients' HRQoL domains and subdomains in standard questionnaires from clinical trials and in real-world setting from social media posts.

Methods: A retrospective study was carried out by collecting social media posts in French language written by internet users mentioning their experiences with ICIs between January 2011 and August 2018. Automatic and manual extractions were implemented to create a corpus where domains and subdomains of HRQoL were classified. These annotations were compared with domains covered by 2 standard HRQoL questionnaires, the EORTC QLQ-C30 and the FACT-G.

Results: We identified 150 users who described their own experience with ICI (89/150, 59.3%) or that of their relative (61/150, 40.7%), with 137 users (91.3%) reporting at least one HRQoL domain in their social media posts. A total of 8 domains and 42 subdomains of HRQoL were identified: Global health (1 subdomain; 115 patients), Symptoms (13; 76), Emotional state (10; 49), Role (7; 22), Physical activity (4; 13), Professional situation (3; 9), Cognitive state (2; 2), and Social state (2; 2). The QLQ-C30 showed a wider global coverage of social media HRQoL subdomains than the FACT-G, 45% (19/42) and 29% (12/42), respectively. For both QLQ-C30 and FACT-G questionnaires, coverage rates were particularly suboptimal for Symptoms (68/123, 55.3% and 72/123, 58.5%, respectively), Emotional state (7/49, 14% and 24/49, 49%, respectively), and Role (17/22, 77% and 15/22, 68%, respectively).

Conclusions: Many patients with cancer are using social media to share their experiences with immunotherapy. Collecting and analyzing their spontaneous narratives are helpful to capture and understand their HRQoL in real-world setting. New measures

Étude de l'utilisation de l'inhalateur connecté Connect'inh en vie réelle et de sa pertinence pour répondre aux besoins des patients asthmatiques et atteints de bronchopneumopathie chronique obstructive.



S.Renner, A.Nigar, A.Mebarki, J.Olivier, V.Marie-Joseph, N.Texier, J.Koman, S.Schück

Kap•Code

ÉTUDE DE L'UTILISATION DE L'INHALATEUR CONNECTÉ CONNECT'INH EN VIE RÉELLE ET DE SA PERTINENCE POUR RÉPONDRE AUX BESOINS DES PATIENTS ASTHMATIQUES ET ATTEINTS DE BPCO

Simon Renner¹, Archange Nigar¹, Adel Mebarki¹, Juliette Olivier¹, Vanessa Marie-Joseph¹, Nathalie Texier¹, Jason Koman¹, Stéphane Schück¹

¹Kap Code, Paris, France

INTRODUCTION

Connect'inh est un dispositif connecté développé par la start-up Kap Code et destiné aux patients atteints de pathologies respiratoires chroniques comme l'asthme ou la BPCO. Cette étude a été réalisée en condition réelle d'utilisation afin de s'assurer de la pertinence de l'objet pour répondre aux besoins des patients. L'étude a aussi pour but d'évaluer le degré d'appropriation de l'objet par les utilisateurs cibles de Connect'inh au cours du temps.



MATÉRIEL & MÉTHODES

L'étude a été conduite auprès de 62 patients atteints de pathologies respiratoires chroniques et recrutés via les réseaux sociaux (Twitter, Facebook, LinkedIn, Google). Les patients ont été suivis pendant 1 semaine et l'étude s'est déroulée en 2 étapes :

- Un premier questionnaire avait pour but d'explorer le profil des malades et leurs attentes concernant un objet connecté, avant la prise en main de l'objet Connect'inh.

- Un deuxième questionnaire était administré une semaine après la prise en main de l'objet.

Des analyses statistiques ont ensuite permis de comparer dans un contexte de vie réelle les attentes des patients quant à l'amélioration de leur prise en charge aux réponses apportées par Connect'inh.

RÉSULTATS

Parmi les 62 patients (fig.1), 97% étaient atteints d'asthme, associé dans 61% des cas à des allergies. Plus de 74% d'entre eux étaient malades depuis plus de 5 ans.

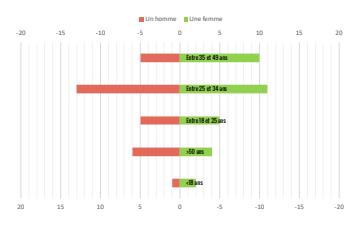


Figure 1 – Pyramides des âges des patients

L'analyse des questionnaires a permis de montrer que 40% des participants à l'étude rencontraient des difficultés au niveau de leur prise en charge thérapeutique. Parmi les patients sous traitement (tab.1), 89% utilisaient un inhalateur à chaque crise.

TRAITEMENT	EFFECTIFS
De crise et de fond	37
De crise	17
De fond	8

Tableau 1 – Effectifs par traitements (crise ou fond)

40% des patients utilisaient des objets connectés en santé (dont 4% pour leur pathologie respiratoire). L'analyse du premier questionnaire met en évidence que 100% des patients estimaient qu'un inhalateur connecté comme Connect'inh pourrait répondre à leurs besoins et les aiderait à améliorer leur prise en charge (fig.2).

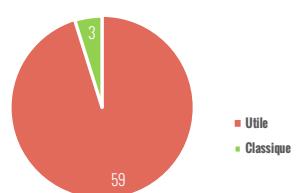


Figure 2 – Qualification de Connect'inh avant sa réception

L'analyse du second questionnaire, après 1 semaine d'utilisation de Connect'inh, a montré que dans 63% des cas, ce dernier répondait mieux à leurs besoins que d'autres alternatives (fig. 3 et 4). 78% recommanderaient le produit.

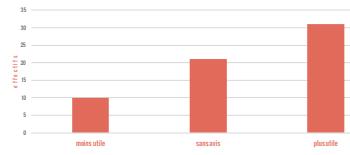


Figure 3 – Utilité du dispositif

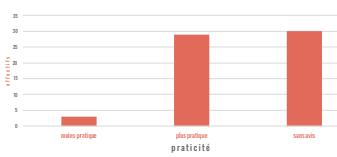


Figure 4 – Pratique du dispositif

Ces patients étaient satisfaits après utilisation du produit dans plus de 76% des cas (fig. 5).

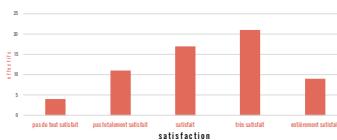


Figure 5 – Satisfaction générale des patients utilisateurs de Connect'inh

De façon générale, les utilisateurs apprécient avant tout d'avoir la possibilité d'un suivi au long terme de leur pathologie via une solution pratique et simple d'utilisation (fig. 6)

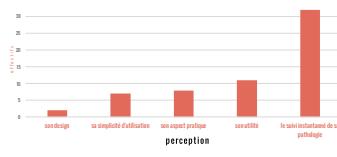


Figure 6 – Perception générale après 1 semaine d'utilisation

CONCLUSION

Cette étude a permis de s'assurer de la bonne appropriation d'un objet connecté et sa pertinence en condition réelle d'utilisation par des patients atteints de pathologies respiratoires.

L'analyse des données met en évidence que ces patients sont aujourd'hui en majorité demandeurs de davantage de suivi, et qu'un objet connecté pourrait être un moyen d'y parvenir.

Dans les conditions réelles d'utilisation, les patients ont estimé que Connect'inh pouvait répondre à leurs besoins. Un suivi à plus long terme est en cours pour s'assurer de l'appropriation de l'inhalateur connecté dans la durée et s'intéresser à l'impact d'un tel dispositif sur la fréquence et la gravité des crises.



Mapping and Modelling of Discussions Related to Gastrointestinal Discomfort in French-Speaking Online Forums : Results of a 15-Year Retrospective Infodemiology Study.



Florent Schäfer, Carole Faviez, Paméla Voillot, Pierre Foulquié, Matthieu Najm, Jean-François Jeanne, Guy Fagherazzi, Stéphane Schück, Boris Le Nevé, G.Fagherazzi, S.Schück.

JOURNAL OF MEDICAL INTERNET RESEARCH

Schäfer et al

Original Paper

Mapping and Modeling of Discussions Related to Gastrointestinal Discomfort in French-Speaking Online Forums: Results of a 15-Year Retrospective Infodemiology Study

Florent Schäfer¹, MSc; Carole Faviez², MSc; Paméla Voillot², MSc; Pierre Foulquié², MSc; Matthieu Najm², MSc; Jean-François Jeanne¹, MSc; Guy Fagherazzi^{3,4}, MSc, PhD; Stéphane Schück², MSc, MD; Boris Le Neve¹, PhD

¹Innovation Science and Nutrition, Danone Nutricia Research, Palaiseau, France

²Kap Code, Paris, France

³Deep Digital Phenotyping Research Unit, Department of Population Health, Luxembourg Institute of Health, Strassen, Luxembourg

⁴Center of Research in Epidemiology and Population Health, UMR 1018 Inserm, Institut Gustave Roussy, Paris-Sud Paris-Saclay University, Villejuif, France

Corresponding Author:

Florent Schäfer, MSc

Innovation Science and Nutrition

Danone Nutricia Research

RD 128 Avenue de la Vauve

Palaiseau, 91767

France

Phone: 33 1 69 35 70 00

Email: florent.schafer@danone.com

Abstract

Background: Gastrointestinal (GI) discomfort is prevalent and known to be associated with impaired quality of life. Real-world information on factors of GI discomfort and solutions used by people is, however, limited. Social media, including online forums, have been considered a new source of information to examine the health of populations in real-life settings.

Objective: The aims of this retrospective infodemiology study are to identify discussion topics, characterize users, and identify perceived determinants of GI discomfort in web-based messages posted by users of French social media.

Methods: Messages related to GI discomfort posted between January 2003 and August 2018 were extracted from 14 French-speaking general and specialized publicly available online forums. Extracted messages were cleaned and deidentified. Relevant medical concepts were determined on the basis of the Medical Dictionary for Regulatory Activities and vernacular terms. The identification of discussion topics was carried out by using a correlated topic model on the basis of the latent Dirichlet allocation. A nonsupervised clustering algorithm was applied to cluster forum users according to the reported symptoms of GI discomfort, discussion topics, and activity on online forums. Users' age and gender were determined by linear regression and application of a support vector machine, respectively, to characterize the identified clusters according to demographic parameters. Perceived factors of GI discomfort were classified by a combined method on the basis of syntactic analysis to identify messages with causality terms and a second topic modeling in a relevant segment of phrases.

Results: A total of 198,866 messages associated with GI discomfort were included in the analysis corpus after extraction and cleaning. These messages were posted by 36,989 separate web users, most of them being women younger than 40 years. Everyday life, diet, digestion, abdominal pain, impact on the quality of life, and tips to manage stress were among the most discussed topics. Segmentation of users identified 5 clusters corresponding to chronic and acute GI concerns. Diet topic was associated with each cluster, and stress was strongly associated with abdominal pain. Psychological factors, food, and allergens were perceived as the main causes of GI discomfort by web users.

Conclusions: GI discomfort is actively discussed by web users. This study reveals a complex relationship between food, stress, and GI discomfort. Our approach has shown that identifying web-based discussion topics associated with GI discomfort and its perceived factors is feasible and can serve as a complementary source of real-world evidence for caregivers.



ScienceDirect



Étude de l'utilisation de l'inhalateur connecté Connect'inh en vie réelle et de sa pertinence pour répondre aux besoins des patients asthmatiques et atteints de bronchopneumopathie chronique obstructive.



S.Renner, A.Nigar, A.Mebarki, J.Olivier, V.Marie-Joseph, N.Texier, J.Koman, S.Schück



ScienceDirect

Revue d'Épidémiologie et de Santé Publique

Volume 67, Supplement 4, June 2019, Pages S185-S186

P11

Étude de l'utilisation de l'inhalateur connecté Connect'inh en vie réelle et de sa pertinence pour répondre aux besoins des patients asthmatiques et atteints de bronchopneumopathie chronique obstructive

S. Renner A. Nigar, A. Mebarki, J. Olivier, V. Marie-Joseph, N. Texier, J. Koman, S. Schück

Show more

Outline

Share

Cite

<https://doi.org/10.1016/j.respe.2019.04.040>

[Get rights and content](#)

Objectifs

Connect'inh est un dispositif connecté développé par la start-up Kap Code et destiné aux patients atteints de pathologies respiratoires chroniques comme l'asthme ou la bronchopneumopathie chronique obstructive (BPCO). Cette étude a été réalisée en condition réelle d'utilisation afin de s'assurer de la pertinence de l'objet pour répondre aux besoins des patients mais aussi pour évaluer au cours du temps le degré d'appropriation de l'objet par les utilisateurs finaux de Connect'inh.

Méthode

L'étude a été menée avec des patients atteints de pathologies respiratoires (asthme et BPCO) recrutés via les réseaux sociaux (Twitter, Facebook, LinkedIn, Google). Les patients ont été suivis pendant trois mois. Au total, 42 personnes (18 hommes et 24 femmes) ont participé à l'étude : 14 % avaient entre 18 et 25 ans, 43 % entre 25 et 34 ans, 26 % entre 35 et 49 ans et 17 % plus de 50 ans. Trois questionnaires ont été administrés en ligne : le premier questionnaire avait pour but d'explorer le profil des malades et leurs attentes concernant un objet connecté, avant la prise en main de l'objet Connect'inh. Un deuxième questionnaire a été administré une semaine après la prise en main de l'objet puis, un troisième après trois mois d'utilisation. Une analyse statistique a ensuite permis de classifier et mettre en regard les attentes des patients et la réponse aux besoins en vie réelle grâce à l'objet Connect'inh.



ScienceDirect



Utilisation des médias sociaux dans l'étude de la qualité de vie des patients atteints de cancer et traités par immunothérapie : une étude infodémiologique



S.Schück, F.-E.Cotté, P.Voillot, B.Falissard, C.Tzourio, P.Foulquié, A.-F.Gaudin, H.Lemasson, C.Faviez

14/06/2022 14:45

Utilisation des médias sociaux dans l'étude de la qualité de vie des patients atteints de cancer et traités par immunothérapie ...



ScienceDirect

Revue d'Épidémiologie et de Santé Publique

Volume 67, Supplement 4, June 2019, Pages S176-S177

4

Utilisation des médias sociaux dans l'étude de la qualité de vie des patients atteints de cancer et traités par immunothérapie : une étude infodémiologique

S. Schück ^a, F.-E. Cotté ^b ☰✉, P. Voillot ^a, B. Falissard ^c, C. Tzourio ^d, P. Foulquié ^a, A.-F. Gaudin ^b, H. Lemasson ^b, C. Faviez ^a

Show more ▾

Outline

Share

Cite

<https://doi.org/10.1016/j.respe.2019.04.005>

[Get rights and content](#)

Objectifs

Dans de nombreux cancers, l'immunothérapie est devenue un standard de prise en charge au profil d'efficacité et de tolérance différent de celui des chimiothérapies. Le vécu et la qualité de vie (QdV) des patients traités par immunothérapie en vie réelle restent méconnus. Les médias sociaux sont de plus en plus un moyen d'expression et d'échange des patients sur leur maladie et leurs préoccupations. L'objectif était de décrire dans les médias sociaux les thématiques de discussion des patients traités ou ayant été traités par immunothérapie et en particulier d'étudier les messages mentionnant leur QdV.

Méthode

L'analyse portait sur les messages publiés par des patients ou leurs proches entre janvier 2011 et août 2018 sur de nombreux médias sociaux généralistes ou spécialisés en santé. Les messages d'intérêt ont été identifiés à l'aide d'une liste exhaustive de mots clés relatifs à



Livre blanc : Les chatbots en santé



Sanofi, Kap Code et Orange business service

Les chatbots en santé



ScienceDirect



How to improve vaccine acceptability (evaluation, pharmacovigilance, communication, public health, mandatory vaccination, fears and beliefs).



Dutilleul A, Morel J, Schilte C, Launay O, Autran B, Béhier JM, Borel T, Bresse X, Chêne G, Courcier S, Dufour V, Faurisson F, Gagneur A, Gelpi O, Gérald F, Kheloufi F, Koeck JL, Lamarque-Garnier V, Lery T, Ménin G, Molimard M, Opinel A, Roger C, Rouby F, Schuck S, Simon L, Soubeyrand B, Truchet MC.

Version of Record: <https://www.sciencedirect.com/science/article/pii/S0040595718302610>
 Manuscript_75e93d65f00f3614130ba685bd4b7019

THERAPIE

RUBRIQUE : GIENS WORKSHOPS 2018 / SOCIAL PHARMACOLOGY

EPUB AHEAD OF PRINT PUIS N°1 (jan-fév) 2019

How to improve vaccine acceptability (evaluation, pharmacovigilance, communication, public health, mandatory vaccination, fears and beliefs)*

How to improve the acceptability of vaccines

Alain Dutilleul^a, Jacques Morel^b, Clémentine Shilte^c, Odile Launay^{d,*}, the participants of Giens XXXIV Round Table “hot topic N°1”, Brigitte Autran^e, Jehan-Michel Béhier^f, Thomas Borel^g, Xavier Bresse^h, Geneviève Chêneⁱ, Soizic Courcier^j, Véronique Dufour^k, François Faurisson^l, Arnaud Gagneur^m, Odile Gelpi^c, Franck Géraldⁿ, Farid Kheloufi^o, Jean-Louis Koeck^p, Véronique Lamarque-Garnier^q, Telma Lery^r, Géraldine Ménin^s, Mathieu Molimard^t, Annick Opinel^c, Claire Roger^j, Franck Rouby^o, Stéphane Schuck^u, Laurent Simon^v, Benoît Soubeyrand^w, Marie-Christine Truchet^x

^a Sanofi France, 69007 Lyon, France

^b CHU de Montpellier, 34295 Montpellier, France

^c Institut Pasteur, 75015 Paris, France

The Adverse Drug Reactions From Patient Reports in Social Media Project: Protocol for an Evaluation Against a Gold Standard



Armelle Arnoux-Guenegou, MSc, PhD ; Yannick Girardeau, MSc, MD ; Xiaoyi Chen, MSc, PhD ; Myrtille Deldossi, MSc ; Rim Aboukhamis, M Pharm ; Carole Faviez, MSc ; Badisse Dahamna, MSc ; Pierre Karapetiantz, MSc ; Sylvie Guillemin-Lanne, MSc ; Agnès Lillo-Le Louët, MD, Prof Dr ; Nathalie Texier, M Pharm ; Anita Burgun, MD, PhD ; Sandrine Katsahian, MD, PhD, Prof Dr

JMIR RESEARCH PROTOCOLS

Arnoux-Guenegou et al

Protocol

The Adverse Drug Reactions From Patient Reports in Social Media Project: Protocol for an Evaluation Against a Gold Standard

Armelle Arnoux-Guenegou¹, MSc, PhD; Yannick Girardeau^{1,2}, MSc, MD; Xiaoyi Chen¹, MSc, PhD; Myrtille Deldossi³, MSc; Rim Aboukhamis⁴, M Pharm; Carole Faviez⁵, MSc; Badisse Dahamna⁶, MSc; Pierre Karapetiantz¹, MSc; Sylvie Guillemin-Lanne³, MSc; Agnès Lillo-Le Louët⁴, MD, Prof Dr; Nathalie Texier⁵, M Pharm; Anita Burgun^{1,2,7}, MD, PhD; Sandrine Katsahian^{1,7,8,9}, MD, PhD, Prof Dr

¹INSERM U1138 - Team 22, Information Sciences to Support Personalized Medicine, Centre de Recherche des Cordeliers, Paris, France

²Département d'Informatique Médicale, Hôpital Européen Georges-Pompidou, Assistance Publique - Hôpitaux de Paris, Paris, France

³Innovative Projects - Text Mining, Expert System, Paris, France

⁴Centre Régional de Pharmacovigilance, Hôpital Européen Georges-Pompidou, Assistance Publique - Hôpitaux de Paris, Paris, France

⁵Kappa Santé, Paris, France

⁶Service d'Informatique Biomédicale, D2IM, Centre Hospitalier Universitaire de Rouen, Rouen, France

⁷INSERM U1138 - Team 22, Information Sciences to Support Personalized Medicine, Paris Descartes University, Sorbonne Paris Cité, Paris, France

⁸Clinical Research Unit Hôpitaux Universitaires Paris Ouest, Hôpital Européen Georges-Pompidou, Assistance Publique - Hôpitaux de Paris, Paris, France

⁹INSERM CIC1418, Clinical Epidemiology, Hôpital Européen Georges-Pompidou, Paris, France

Corresponding Author:

Xiaoyi Chen, MSc, PhD

INSERM U1138 - Team 22

Information Sciences to Support Personalized Medicine

Centre de Recherche des Cordeliers

15 Rue de l'école de Médecine

Esc D, 1e étage

Paris, 75006

France

Phone: 33 71196369

Email: chenxiaoyi619@gmail.com

Abstract

Background: Social media is a potential source of information on postmarketing drug safety surveillance that still remains unexploited nowadays. Information technology solutions aiming at extracting adverse reactions (ADRs) from posts on health forums require a rigorous evaluation methodology if their results are to be used to make decisions. First, a gold standard, consisting of manual annotations of the ADR by human experts from the corpus extracted from social media, must be implemented and its quality must be assessed. Second, as for clinical research protocols, the sample size must rely on statistical arguments. Finally, the extraction methods must target the relation between the drug and the disease (which might be either treated or caused by the drug) rather than simple co-occurrences in the posts.

Objective: We propose a standardized protocol for the evaluation of a software extracting ADRs from the messages on health forums. The study is conducted as part of the Adverse Drug Reactions from Patient Reports in Social Media project.

Methods: Messages from French health forums were extracted. Entity recognition was based on *Racine Pharma* lexicon for drugs and Medical Dictionary for Regulatory Activities terminology for potential adverse events (AEs). Natural language processing-based techniques automated the ADR information extraction (relation between the drug and AE entities). The corpus of evaluation was a random sample of the messages containing drugs and/or AE concepts corresponding to recent pharmacovigilance alerts. A total of 2 persons experienced in medical terminology manually annotated the corpus, thus creating the gold standard, according to an annotator guideline. We will evaluate our tool against the gold standard with recall, precision, and f-measure. Interannotator agreement, reflecting gold standard quality, will be evaluated with hierarchical kappa. Granularities in the terminologies will be further explored.



Web-Based Signal Detection Using Medical Forums Data in France: Comparative Analysis



Marie-Laure Kürzinger, MSc; Stéphane Schück, MSc, MD; Nathalie Texier, PharmD; Redhouane Abdellaoui, MSc; Carole Faviez, MSc; Julie Pouget, MSc; Ling Zhang, MSc; Stéphanie Tcherny-Lessenot, MSc, MD; Stephen Lin, MD; Juhaeri Juhaeri, PhD

Web-Based Signal Detection Using Medical Forums Data in France: Comparative Analysis

Marie-Laure Kürzinger¹, MSc; Stéphane Schück², MSc, MD; Nathalie Texier², PharmD; Redhouane Abdellaoui³, MSc; Carole Faviez³, MSc; Julie Pouget⁴, MSc; Ling Zhang⁵, MSc; Stéphanie Tcherny-Lessenot¹, MSc, MD; Stephen Lin⁵, MD; Juhaeri Juhaeri⁶, PhD

¹Epidemiology and Benefit Risk Evaluation, Sanofi, Chilly-Mazarin, France

²Kappa Santé, Paris, France

³Kap Code, Paris, France

⁴Information Technology and Solutions, Sanofi, Lyon, France

⁵Global Pharmacovigilance, Sanofi, Bridgewater, NJ, United States

⁶Epidemiology and Benefit Risk Evaluation, Sanofi, Bridgewater, NJ, United States

Corresponding Author:

Marie-Laure Kürzinger, MSc
Epidemiology and Benefit Risk Evaluation
Sanofi
1 avenue Pierre Brossolette
Chilly-Mazarin, 91385
France
Phone: 33 1 69 74 59 42
Email: marie-laure.kurzinger@sanofi.com

Abstract

Background: While traditional signal detection methods in pharmacovigilance are based on spontaneous reports, the use of social media is emerging. The potential strength of Web-based data relies on their volume and real-time availability, allowing early detection of signals of disproportionate reporting (SDRs).

Objective: This study aimed (1) to assess the consistency of SDRs detected from patients' medical forums in France compared with those detected from the traditional reporting systems and (2) to assess the ability of SDRs in identifying earlier than the traditional reporting systems.

Methods: Messages posted on patients' forums between 2005 and 2015 were used. We retained 8 disproportionality definitions. Comparison of SDRs from the forums with SDRs detected in VigiBase was done by describing the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, receiver operating characteristics curve, and the area under the curve (AUC). The time difference in months between the detection dates of SDRs from the forums and VigiBase was provided.

Results: The comparison analysis showed that the sensitivity ranged from 29% to 50.6%, the specificity from 86.1% to 95.5%, the PPV from 51.2% to 75.4%, the NPV from 68.5% to 91.6%, and the accuracy from 68% to 87.7%. The AUC reached 0.85 when using the metric empirical Bayes geometric mean. Up to 38% (12/32) of the SDRs were detected earlier in the forums than that in VigiBase.

Conclusions: The specificity, PPV, and NPV were high. The overall performance was good, showing that data from medical forums may be a valuable source for signal detection. In total, up to 38% (12/32) of the SDRs could have been detected earlier, thus, ensuring the increased safety of patients. Further enhancements are needed to investigate the reliability and validation of patients' medical forums worldwide, the extension of this analysis to all possible drugs or at least to a wider selection of drugs, as well as to further assess performance against established signals.

(*J Med Internet Res* 2018;20(11):e10466) doi:[10.2196/10466](https://doi.org/10.2196/10466)

KEYWORDS

adverse event; internet; medical forums; pharmacovigilance; signal detection; signals of disproportionate reporting; social media



LIVRE BLANC : Les réseaux sociaux et la santé, un enjeu pour le suivi des patients et la recherche scientifique



**HEALTHCARE
DATA INSTITUTE**

INTERNATIONAL THINK TANK
DEDICATED TO BIG DATA IN HEALTHCARE

.....

**LES RÉSEAUX SOCIAUX
ET LA SANTÉ :**

UN ENJEU POUR LE SUIVI
DES PATIENTS ET LA
RECHERCHE SCIENTIFIQUE

SEPTEMBRE 2018



Que nous apportent les réseaux sociaux quant à la crise sanitaire du Levothyrox d'Août 2017 ?



Stéphane Schück, Paméla Voillot, Pierre Foulquié, Carole Faviez,
Adel Mebarki, Nathalie Texier, Tristan Gauvin

QUE NOUS APPORTENT LES RÉSEAUX SOCIAUX QUANT À LA CRISE SANITAIRE DU LEVOTHYROX® D'AOÛT 2017 ?

Stéphane Schück¹, Paméla Voillot¹, Pierre Foulquié¹, Carole Faviez¹, Adel Mebarki¹, Nathalie Texier¹

¹Kap Code, Paris, France

KAPCODE

INTRODUCTION

Le Levothyrox® prescrit en France à plus de 3 millions de patients atteints de pathologies thyroïdiennes, a fait l'objet d'une polémique en août 2017 suite à un changement de formule en mars 2017. Un nombre important d'effets indésirables a été rapporté auprès des autorités sanitaires, des incertitudes persistent concernant la nature de ces effets ou leur spécificité par rapport à la nouvelle formule.

En parallèle, 32 millions d'internautes sont actifs sur les réseaux sociaux, favorisant les échanges à propos de leur santé.

Dans ce contexte, l'objectif de cette étude consistait à identifier via les réseaux sociaux ces échanges et les thématiques abordées par les patients concernant le Levothyrox®, ainsi que les effets indésirables mentionnés.

MATÉRIEL & MÉTHODES

Les messages évoquant le Levothyrox® publiés entre 2007 et 2017 ont été extraits à partir du forums français Doctissimo via l'outil **Detect¹** développé par la société Kap Code. L'identification des effets indésirables était effectuée via ce dernier.

L'application d'un algorithme de veille basé sur le volume de messages postés permettait la classification des observations d'une période selon leur activité, plus ou moins soutenue.

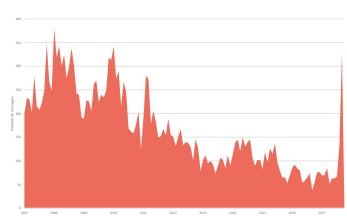
Les thématiques de discussion étaient modélisées à l'aide d'un modèle de sujet (*topic model*). L'application d'un partitionnement k-moyennes (*k-means clustering*) permettait de catégoriser les auteurs des messages selon leur comportement.

Enfin, la détection de signal consistait en l'application de la méthode du PRR Composite* sur trois périodes : 2007-2015, 2007-2016 et 2007-2017.

RÉSULTATS

Le corpus d'analyse Levothyrox® contenait près de 30 000 messages publiés entre 2007 et 2017 par plus de 8 000 internautes différents.

L'évolution temporelle (fig.1) a permis de mettre en évidence un pic de messages en août et septembre 2017, concomitant au bruit médiatique lié au changement de formule du Levothyrox® en mars 2017.



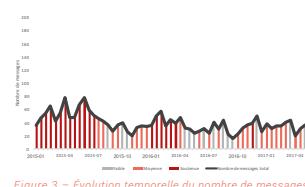
Plus de 34 000 effets secondaires ont été détectés au sein des messages entre 2007 et 2017. L'évolution du nombre d'effets suit l'évolution du nombre de messages. Ces effets sont regroupés en catégories (fig.2).

Quatre de ces catégories de troubles regroupent plus de 75% des effets détectés.



Figure 2 - Catégories de troubles identifiés

L'algorithme de veille a permis de classer les observations de l'été 2017 en activité soutenue (fig.3).



Les thématiques de discussions (fig.4) identifiées couvraient les différents aspects des pathologies thyroïdiennes tels que les actes médicaux relatifs à la maladie, les modalités de la prise du traitement et le parcours de soin.

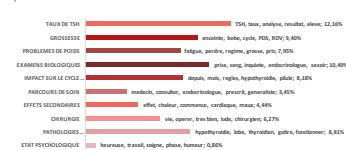


Figure 4 - Thématiques de discussions estimées sur les messages associés au Levothyrox®

La catégorisation des utilisateurs (fig.5) a fait émerger un groupe de nouveaux internautes n'ayant quasiment pas écrit depuis 2017. Ils étaient les auteurs de 77% des messages postés cette année là.

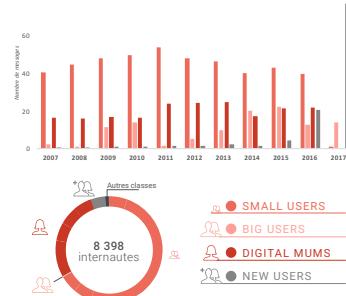


Figure 5 - Répartition annuelle des messages par typologie d'utilisateurs

SMALL USERS

Ces utilisateurs n'ont posté que très peu (2 messages en moyenne), sur une petite période et avant 2017.

BIG USERS

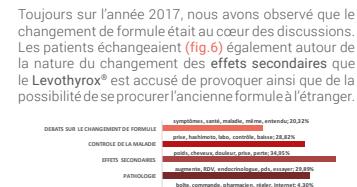
Ils forment une communauté active. Ils postent beaucoup (plus de 340 messages en moyenne) et depuis presque 10 ans.

DIGITAL MUMS

Cette classe est constituée d'utilisatrices postant à propos de l'impact du Levothyrox® sur le cycle menstruel et sur la grossesse.

NEW USERS

Ces utilisatrices ont commencé à poster en 2017 et leurs sujets de prédilection sont les actualités et détails autour de la nouvelle formule et les effets secondaires.



Toujours sur l'année 2017, nous avons observé que le changement de formule était au cœur des discussions. Les patients échangeaient (fig.6) également autour de la nature du changement des effets secondaires que le Levothyrox® est accusé de provoquer ainsi que de la possibilité de se procurer l'ancienne formule à l'étranger.



La détection de signaux (fig.7) a mis en avant des signaux déjà connus tels que les chutes de cheveux ou les vertiges. Les deux nouveaux signaux observés en 2017 étaient la froidure des extrémités et la fatigue musculaire.

CONCLUSION

Les messages publiés en 2017 concernaient principalement le changement de formule et ses conséquences. Ils étaient majoritairement publiés par de nouveaux utilisateurs. Les effets secondaires médiatisés étaient déjà connus mais ont augmenté en 2017. La surveillance des réseaux sociaux et des effets indésirables rapportés par de nouveaux utilisateurs apparaît comme un outil d'alerte supplémentaire dans la chaîne de surveillance du médicament.

RÉFÉRENCES

[1] Abdellaoui Retal. Filtering Entities to Optimize Identification of Adverse Drug Reaction From Social Media: How can the Number of Words Between Entities in the Messages Help?. JMIR public health and surveillance, 2017, vol. 3, no 2.



Mining Patients' Narratives in Social Media for Pharmacovigilance: Adverse Effects and Misuse of Methylphenidate



Xiaoyi Chen, Carole Faviez, Stéphane Schuck, Agnès Lillo-Le-Louët, Nathalie Texier, Badisse Dahamna, Charles Huot, Pierre Foulquié, Suzanne Pereira, Vincent Leroux, Pierre Karapetiantz, Armelle Guenegou-Arnoux, Sandrine Katsahian, Cédric Bousquet and Anita Burgun

OPEN ACCESS

Edited by:

Iñaki Gutiérrez-Ibarluzea,
Basque Office for Health Technology
Assessment (OSTEBA), Spain

Reviewed by:

Bryan Martin Bennett,
Adelphi (United Kingdom),
United Kingdom
Ana Paula Martins,
Universidade de Lisboa, Portugal

*Correspondence:

Xiaoyi Chen
xiao.yi.chen@inserm.fr
Carole Faviez
carole.faviez@kapcode.fr

[†]These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Pharmaceutical Medicine and
Outcomes Research,
a section of the journal
Frontiers in Pharmacology

Received: 15 December 2017

Accepted: 04 May 2018

Published: 24 May 2018

Citation:

Chen X, Faviez C, Schuck S, Lillo-Le-Louët A, Texier N, Dahamna B, Huot C, Foulquié P, Pereira S, Leroux V, Karapetiantz P, Guenegou-Arnoux A, Katsahian S, Bousquet C and Burgun A (2018) Mining Patients' Narratives in Social Media for Pharmacovigilance: Adverse Effects and Misuse of Methylphenidate. *Front. Pharmacol.* 9:541. doi: 10.3389/fphar.2018.00541

Mining Patients' Narratives in Social Media for Pharmacovigilance: Adverse Effects and Misuse of Methylphenidate

Xiaoyi Chen ^{1*†}, Carole Faviez ^{2*†}, Stéphane Schuck ², Agnès Lillo-Le-Louët ³, Nathalie Texier ², Badisse Dahamna ^{4,5}, Charles Huot ⁶, Pierre Foulquié ², Suzanne Pereira ⁷, Vincent Leroux ⁸, Pierre Karapetiantz ¹, Armelle Guenegou-Arnoux ¹, Sandrine Katsahian ^{1,9}, Cédric Bousquet ¹⁰ and Anita Burgun ^{1,9}

¹ UMRS 1138, équipe 22, Institut National de la Santé et de la Recherche Médicale, Centre de Recherche des Cordeliers, Université Paris Descartes, Paris, France, ² Kappa Santé, Paris, France, ³ Centre Régional de Pharmacovigilance, Hôpital Européen Georges-Pompidou, AP-HP, Paris, France, ⁴ Service d'Informatique Biomédicale, Centre Hospitalier Universitaire de Rouen, Rouen, France, ⁵ Laboratoire d'Informatique, du Traitement de l'Information et des Systèmes-TIBS EA 4108, Rouen, France, ⁶ Expert System, Paris, France, ⁷ Vidal, Issy Les Moulineaux, France, ⁸ Institut de Santé Urbaine, Saint-Maurice, France, ⁹ Département d'Informatique Médicale, Hôpital Européen Georges Pompidou, Paris, France, ¹⁰ Sorbonne Université, Inserm, université Paris 13, Laboratoire d'informatique médicale et d'ingénierie des connaissances en e-santé, LIMICS, Paris, France

Background: The Food and Drug Administration (FDA) in the United States and the European Medicines Agency (EMA) have recognized social media as a new data source to strengthen their activities regarding drug safety.

Objective: Our objective in the ADR-PRISM project was to provide text mining and visualization tools to explore a corpus of posts extracted from social media. We evaluated this approach on a corpus of 21 million posts from five patient forums, and conducted a qualitative analysis of the data available on methylphenidate in this corpus.

Methods: We applied text mining methods based on named entity recognition and relation extraction in the corpus, followed by signal detection using proportional reporting ratio (PRR). We also used topic modeling based on the Correlated Topic Model to obtain the list of the matics in the corpus and classify the messages based on their topics.

Results: We automatically identified 3443 posts about methylphenidate published between 2007 and 2016, among which 61 adverse drug reactions (ADR) were automatically detected. Two pharmacovigilance experts evaluated manually the quality of automatic identification, and a f-measure of 0.57 was reached. Patient's reports were mainly neuro-psychiatric effects. Applying PRR, 67% of the ADRs were signals, including most of the neuro-psychiatric symptoms but also palpitations. Topic modeling showed that the most represented topics were related to *Childhood and Treatment initiation*, but also *Side effects*. Cases of misuse were also identified in this corpus, including recreational use and abuse.

Détection automatique du mésusage des neuroleptiques dans le trouble anxieux et la démence à partir des réseaux sociaux



Stéphane Schück, Pierre Foulquié, Adel Mebarki, Paméla Voillot, Nathalie Texier, Carole Faviez

DÉTECTION AUTOMATIQUE DU MÉSUSAGE DES NEUROLEPTIQUES DANS LE TROUBLE ANXIEUX ET LA DÉMENCE À PARTIR DES RÉSEAUX SOCIAUX

Stéphane Schück¹, Pierre Foulquié¹, Adel Mebarki¹, Paméla Voillot¹, Nathalie Texier¹, Carole Faviez¹
¹Kap Code, Paris, France

KAP CODE

INTRODUCTION

Les neuroleptiques sont des psychotropes utilisés dans le traitement des troubles psychotiques, tels que la schizophrénie et les troubles bipolaires. Dès 2011, l'étude réalisée par Alexander et al. met en évidence un usage des neuroleptiques en dehors de leurs autorisations de mise sur le marché (AMM) grandissant aux Etats-Unis [1].

En parallèle, on estime que 50% des français sont actifs sur les réseaux sociaux, un grand nombre d'entre eux échangeant à propos de leur santé.

Dans ce contexte, l'objectif de cette étude consistait à (1) évaluer s'il était possible d'identifier des usages hors AMM des neuroleptiques sur les réseaux sociaux français dans le cadre de la démence et des troubles anxieux et (2) mettre en place une méthodologie de détection automatique de ces cas.

MATÉRIEL & MÉTHODES

Les messages associés à deux neuroleptiques, aripiprazole et rispéridone, ont été extraits à partir de forums médicaux généralistes français via l'outil DeteC't [2] développé par la société Kap Code.

La méthode présentée ici s'organisait en deux étapes. Dans un premier temps, des échantillons des deux corpus ont été annotés manuellement afin d'identifier les messages associés à une prise médicamenteuse dans le cadre d'un des deux mésusages recherchés.

Dans un deuxième temps, des champs lexicaux correspondant à ces mésusages ont été créés. Ils ont été initialement constitués à partir du dictionnaire d'effets indésirables MedDRA (*Medical Dictionary for Regulatory Activities*). Un modèle de sujet (*topic model*) a ensuite permis l'identification, au sein du corpus aripiprazole, de messages contenant un vocabulaire patient évocateur du mésusage. Ces messages ont été revus et le vocabulaire identifié a été utilisé pour compléter les champs lexicaux. Des synonymes issus du langage courant ont enfin été ajoutés afin de parfaire les champs lexicaux.

La recherche des éléments de ces champs lexicaux au sein des corpus avait pour objectif d'identifier automatiquement les messages associés aux mésusages en question.

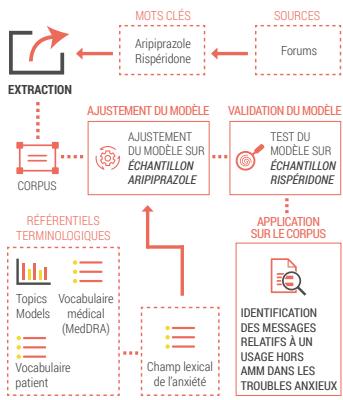


Figure 1 - Synoptique général

Cette méthode de détection a été appliquée aux deux échantillons préalablement annotés, dans un premier temps à l'échantillon aripiprazole pour ajustement du champ lexical et de la méthode, puis dans un deuxième temps à l'échantillon rispéridone pour test du modèle.

RÉSULTATS

Les corpus associés à aripiprazole et rispéridone contenaient, respectivement, 9 528 et 3 868 messages.

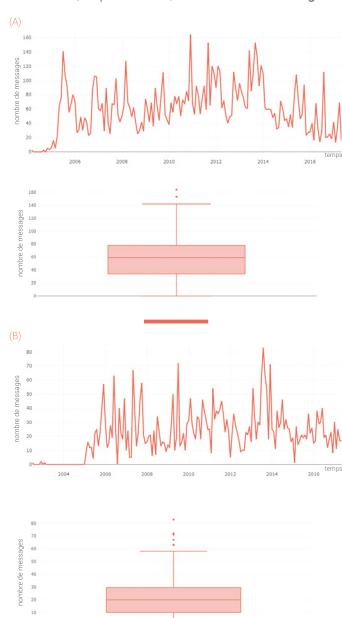


Figure 2 - Distribution des observations mensuelles.
(A) - Corpus aripiprazole / (B) - Corpus rispéridone

Annotation des messages

1 000 messages issus du corpus aripiprazole et 400 messages issus du corpus rispéridone ont été annotés manuellement.

Au sein du corpus aripiprazole (resp. rispéridone), 35 messages associés au trouble anxieux (resp. 15), soit 3,5% (IC 95% : 2,4%-4,6%) des messages de l'échantillon (resp. 3,8%, IC 95% : 2,0%-5,6%), ont été identifiés.

Aucun message associé à la démence n'a pu être identifié dans aucun des deux corpus.

Tableau 1 - Effectifs en nombre de messages au sein des échantillons annotés.

TRAITEMENT	Échantillon	Troubles anxieux	Démences
Aripiprazole	1 000 messages	35 messages	0 message
Rispéridone	400 messages	15 messages	0 message

Constitution des champs lexicaux

La recherche de termes évocateurs du trouble anxieux au sein du dictionnaire MedDRA a permis l'identification de 1 116 termes. En parallèle, la revue de messages et l'ajout de vocabulaire patient a permis d'identifier 227 termes supplémentaires.

Ces 1 343 termes ont été répartis en trois classes en fonction de leur spécificité par rapport au trouble anxieux ; la classe 1 contenait des termes très spécifiques du trouble anxieux (tels que *angoisse*, *anxiété*, etc.), la classe 2 des termes évocateurs du trouble anxieux mais qui pouvaient également être associés à d'autres pathologies (*agitation*, *insomnie*, etc.) et la classe 3 des termes encore moins spécifiques du trouble anxieux.

Tableau 2 - Répartition des termes du champ lexical du trouble anxieux

ORIGINE	Classe 1	Classe 2	Classe 3	Total
MedDRA	205	889	22	1 116
Termes supplémentaires	54	172	1	227
Total	259	1 061	23	1 343

Test du modèle

Le test de différents modèles utilisant les différentes classes de ces champs lexicaux sur l'échantillon aripiprazole a permis d'identifier le modèle possédant les meilleures performances.

Ce modèle était uniquement construit à partir de la classe 1 du champ lexical, c'est à dire les termes les plus spécifiques (soit 205 termes issus de MedDRA et 54 termes issus de l'annotation des messages du corpus aripiprazole, identifiés par le *topic model*).

Le modèle a alors été réappliqué sur l'échantillon rispéridone. Les performances obtenues ont été équivalentes sur les deux échantillons, le rappel et la précision allant respectivement de 97 à 100% et de 19 à 20%.

Tableau 3 - Performances du modèle trouble anxieux sur les échantillons

TRAITEMENT	Sensibilité	Spécificité	VPP	VNP
Aripiprazole	97%	86%	20%	100%
Rispéridone	100%	83%	19%	100%

CONCLUSION

Une prise de neuroleptiques dans le cadre des troubles anxieux a été identifiée sur les réseaux sociaux. Le mésusage dans le cadre de la démence n'a pas été identifié. La méthodologie de détection automatique de ces cas de mésusages a obtenu de bonnes performances pour détecter ces messages d'intérêt.

RÉFÉRENCES

[1] Alexander, G., Caleb, et al. «Increasing off-label use of antipsychotic medications in the United States, 1995–2008.» *Pharmacoepidemiology and drug safety* 20.2 (2011): 177-184.

[2] Abdellaoui Retal. Filtering Entities to Optimize Identification of Adverse Drug Reaction From Social Media: How Can the Number of Words Between Entities in the Messages Help? *JMIR public health and surveillance*, 2017, vol. 3, no 2.



The Adverse Drug Reactions from Patient Reports in Social Media Project: Five Major Challenges to Overcome to Operationalize Analysis and Efficiently Support Pharmacovigilance Process



Cedric Bousquet, PharmD, PhD; Badisse Dahamna, MSc; Sylvie Guillemin-Lanne, MSc; Stefan J Darmoni^{1,3}, MD, PhD; Carole Faviez⁵, MSc; Charles Huot⁴, PhD; Sandrine Katsahian⁶, MD, PhD; Vincent Leroux⁷, MD; Suzanne Pereira⁸, PhD; Christophe Richard⁹, MD; Stéphane Schück, MPH, MD; Julien Souvignet, MSc; Agnès Lillo-Le Louët, MD; Nathalie Texier, PharmD

The Adverse Drug Reactions from Patient Reports in Social Media Project: Five Major Challenges to Overcome to Operationalize Analysis and Efficiently Support Pharmacovigilance Process

Cedric Bousquet^{1,2}, PharmD, PhD; Badisse Dahamna³, MSc; Sylvie Guillemin-Lanne⁴, MSc; Stefan J Darmoni^{1,3}, MD, PhD; Carole Faviez⁵, MSc; Charles Huot⁴, PhD; Sandrine Katsahian⁶, MD, PhD; Vincent Leroux⁷, MD; Suzanne Pereira⁸, PhD; Christophe Richard⁹, MD; Stéphane Schück⁵, MPH, MD; Julien Souvignet¹, MSc; Agnès Lillo-Le Louët¹⁰, MD; Nathalie Texier⁵, PharmD

¹Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé, U1142, Institut National de la Santé et de la Recherche Médicale, Paris, France

²Service de Santé Publique et de l'Information Médicale, Centre Hospitalier Universitaire de Saint Etienne, Saint-Etienne, France

³Department of Biomedical Informatics, Rouen University Hospital, Rouen, France

⁴Expert System, Paris, France

⁵Kappa Santé, Paris, France

⁶Unité mixte de recherche 1138, équipe 22, Institut National de la Santé et de la Recherche Médicale, Centre de Recherche des Cordeliers, Paris, France

⁷Institut de Santé Urbaine, Saint Maurice, France

⁸Vidal, Issy Les Moulineaux, France

⁹Santeos, Paris, France

¹⁰Assistance Publique-Hôpitaux de Paris, Hôpital Européen Georges Pompidou, Centre Régional de Pharmacovigilance, Paris, France

Corresponding Author:

Cedric Bousquet, PharmD, PhD

Service de Santé Publique et de l'Information Médicale

Centre Hospitalier Universitaire de Saint Etienne

Chemin de la Marandière

Bâtiment CIM42 - Hôpital Nord

Saint-Etienne,

France

Phone: 33 477 27974

Email: cedric.bousquet@chu-st-etienne.fr

Abstract

Background: Adverse drug reactions (ADRs) are an important cause of morbidity and mortality. Classical Pharmacovigilance process is limited by underreporting which justifies the current interest in new knowledge sources such as social media. The Adverse Drug Reactions from Patient Reports in Social Media (ADR-PRISM) project aims to extract ADRs reported by patients in these media. We identified 5 major challenges to overcome to operationalize the analysis of patient posts: (1) variable quality of information on social media, (2) guarantee of data privacy, (3) response to pharmacovigilance expert expectations, (4) identification of relevant information within Web pages, and (5) robust and evolutive architecture.

Objective: This article aims to describe the current state of advancement of the ADR-PRISM project by focusing on the solutions we have chosen to address these 5 major challenges.

Methods: In this article, we propose methods and describe the advancement of this project on several aspects: (1) a quality driven approach for selecting relevant social media for the extraction of knowledge on potential ADRs, (2) an assessment of ethical issues and French regulation for the analysis of data on social media, (3) an analysis of pharmacovigilance expert requirements when reviewing patient posts on the Internet, (4) an extraction method based on natural language processing, pattern based matching, and selection of relevant medical concepts in reference terminologies, and (5) specifications of a component-based architecture for the monitoring system.

Results: Considering the 5 major challenges, we (1) selected a set of 21 validated criteria for selecting social media to support the extraction of potential ADRs, (2) proposed solutions to guarantee data privacy of patients posting on Internet, (3) took into



Detection of Cases of Noncompliance to Drug Treatment in Patient Forum Posts: Topic Model Approach



Redhouane Abdellaoui, MSc; Pierre Foulquié, MSc; Nathalie Texier, PharmD; Carole Faviez, MSc; Anita Burgun, MD, PhD; Stéphane Schück², MSc, MD

Detection of Cases of Noncompliance to Drug Treatment in Patient Forum Posts: Topic Model Approach

Redhouane Abdellaoui¹, MSc; Pierre Foulquié², MSc; Nathalie Texier², PharmD; Carole Faviez², MSc; Anita Burgun^{1,3}, MD, PhD; Stéphane Schück², MSc, MD

¹Unité de Mixte de Recherche 1138 Team 22, Institut National de la Santé et de la Recherche Médicale / Université Pierre et Marie Curie, Paris, France

²Kappa Santé, Innovation (Kap Code), Paris, France

³Medical Informatics, Hôpital Européen Georges-Pompidou, Assistance Publique-Hôpitaux de Paris, Paris, France

Corresponding Author:

Redhouane Abdellaoui, MSc

Unité de Mixte de Recherche 1138 Team 22

Institut National de la Santé et de la Recherche Médicale / Université Pierre et Marie Curie

15 Rue de l'École de Médecine

Paris, 75006

France

Phone: 33 648094269

Email: redhouane.a@gmail.com

Abstract

Background: Medication nonadherence is a major impediment to the management of many health conditions. A better understanding of the factors underlying noncompliance to treatment may help health professionals to address it. Patients use peer-to-peer virtual communities and social media to share their experiences regarding their treatments and diseases. Using topic models makes it possible to model themes present in a collection of posts, thus to identify cases of noncompliance.

Objective: The aim of this study was to detect messages describing patients' noncompliant behaviors associated with a drug of interest. Thus, the objective was the clustering of posts featuring a homogeneous vocabulary related to nonadherent attitudes.

Methods: We focused on escitalopram and aripiprazole used to treat depression and psychotic conditions, respectively. We implemented a probabilistic topic model to identify the topics that occurred in a corpus of messages mentioning these drugs, posted from 2004 to 2013 on three of the most popular French forums. Data were collected using a Web crawler designed by Kappa Santé as part of the Detec't project to analyze social media for drug safety. Several topics were related to noncompliance to treatment.

Results: Starting from a corpus of 3650 posts related to an antidepressant drug (escitalopram) and 2164 posts related to an antipsychotic drug (aripiprazole), the use of latent Dirichlet allocation allowed us to model several themes, including interruptions of treatment and changes in dosage. The topic model approach detected cases of noncompliance behaviors with a recall of 98.5% (272/276) and a precision of 32.6% (272/844).

Conclusions: Topic models enabled us to explore patients' discussions on community websites and to identify posts related with noncompliant behaviors. After a manual review of the messages in the noncompliance topics, we found that noncompliance to treatment was present in 6.17% (276/4469) of the posts.

(*J Med Internet Res* 2018;20(3):e85) doi:[10.2196/jmir.9222](https://doi.org/10.2196/jmir.9222)

KEYWORDS

medication adherence; compliance; infodemiology; social media; text mining; depression; psychosis; peer-to-peer support; virtual community



Mining adverse drug reactions in social media with named entity recognition and semantic methods



Xiaoyi Chen, Myrtille Deldossi, Rim Aboukhamis, Carole Faviez, Badisse Dahamna, Pierre Karapetiantz, Armelle Guenegou-Arnoux, Yannick Girardeau, Sylvie Guillemin-Lanne, Agnès Lillo-Le-Louët, Nathalie Texier, Anita Burgun, Sandrine Katsahian

Mining adverse drug reactions in social media with named entity recognition and semantic methods

Xiaoyi Chen^a, Myrtille Deldossi^b, Rim Aboukhamis^c, Carole Faviez^d, Badisse Dahamna^{e,f,g}, Pierre Karapetiantz^a, Armelle Guenegou-Arnoux^a, Yannick Girardeau^{h,i}, Sylvie Guillemin-Lanne^b, Agnès Lillo-Le-Louët^c, Nathalie Texier^d, Anita Burgun^{a,h,i}, Sandrine Katsahian^{a,h,i}

^a INSERM, UMR S1138, équipe 22, Centre de Recherche des Cordeliers, Paris, France

^b Expert System, 75012 Paris, France

^c Centre Régional de Pharmacovigilance, Hôpital Européen Georges-Pompidou, AP-HP, Paris, France

^d Kappa Santé, 75002 Paris, France

^e Service d'Informatique Biomédicale, CHU de Rouen, France

^f LITIS-TIBS EA 4108, 76031 Rouen Cedex, France

^g INSERM, U1142, LIMICS, 75006 Paris, France

^h Université Paris Descartes, Sorbonne Paris Cité, UMR S1138, Centre de Recherche de Cordeliers, Paris, France

ⁱ Département d'Informatique Hospitalière, Hôpital Européen Georges-Pompidou, AP-HP, Paris, France

Abstract

Suspected adverse drug reactions (ADR) reported by patients through social media can be a complementary source to current pharmacovigilance systems. However, the performance of text mining tools applied to social media to discover ADRs needs to be evaluated. In this paper, we introduce the approach developed to mine ADR from French social media. A protocol of evaluation is highlighted, which includes a detailed sample size determination and corpus constitution. Our text mining approach provided very encouraging preliminary results with F-measures of 0.94, 0.81 and 0.70 for recognition of drugs, symptoms and ADRs respectively, thus this approach is promising for downstream pharmacovigilance analysis.

Keywords:

Social Media, Pharmacovigilance, Data Mining

Introduction

The rapid expansion of the Internet and social media is changing the way people gather information about disease and treatment, as well as how they share personal health experiences with others [1]. The *Digit in 2016* [2] reported that, in France, 86% of the population are active internet users. This proportion is higher than Western Europe's average of 83% and slightly lower than North America's average of 88%. Various questionnaire statistics [3]–[5] showed that a large proportion of French people (46% to 71%) use the Internet to seek medical or health related information. Many people also use social media, such as forums, to communicate with others with the same health concerns and share information related to their illnesses, feelings, medication use and many other aspects [6], which offers promising opportunities for public health surveillance with a rich internet-based, patient-generated source.

The World Health Organization (WHO) defines Pharmacovigilance as “the science relating to the detection, assessment, understanding, and prevention of adverse effects or any other drug-related problems”. It begins during clinical trials and continues after the drug is released onto the market. However a study [7] showed that 60% of potentially fatal ADRs were

not described in initial drug labels and 39% were not included in any report of randomized controlled trials. The main pharmacovigilance tools are spontaneous reporting systems, driven by drug agencies, like the U.S. FDA's (Food and Drug Administrations) Adverse Event Reporting System (FAERS) which gathers voluntary reports by healthcare professionals and consumers (59% by professionals Vs. 41% by consumers in 2006, and 46% Vs. 54% in Q1 2015 [8]). It can also include Phase IV clinical trials driven by pharmaceutical companies and governmental agencies [9]. Despite such systems, the underreporting of ADRs by the patients as well as by the health professionals remains a significant limitation [10][11].

Several studies have already demonstrated the value of mining ADR from social media posts [12]–[14]. However, in contrast to the numerous studies in social media, the potential of utilising this data for pharmacovigilance has not yet been fully exploited. It represents only 0.5% of publications with “social media” (SM) keyword query in the PubMed database (Figure 1A). Figure 1B shows that the number of publications with “SM + pharmacovigilance” as keywords has increased exponentially in the last five years.

A recent scoping review [11] outlined five complete steps that should be taken for processing ADR extraction in social media: (1) data collection, (2) preprocessing, (3) entity recognition (for drugs and symptoms), (4) identifying the relationship between drug and symptom, and (5) results evaluation. Since content and language of medical social media differ from those of general social media and of clinical documents, specific text mining methods or techniques based on Natural Language Processing (NLP) are necessary for step (3) and step (4) in order to identify medical concepts (such as drugs, symptoms, etc.) and relations among them [15]. It is evident that the performance of the text mining methods plays a decisive role in ADR signal detection.

From a text mining perspective, the key challenge is that internet users' expressions are usually informal and colloquial, especially when they describe their feelings and symptoms. However, researches have progressed using (i) various data sources, such as forum messages [16][17], Twitter micro blogs [18][19] and Yahoo Wellness Groups [12], in (ii) different languages, (iii) diverse approaches, such as Support Vector



Filtering Entities to Optimize Identification of Adverse Drug Reaction From Social Media: How Can the Number of Words Between Entities in the Messages Help?



Redhouane Abdellaoui, MSc; Stéphane Schück, MSc, MD; Nathalie Texier, PharmD; Anita Burgun, MD, PhD

Filtering Entities to Optimize Identification of Adverse Drug Reaction From Social Media: How Can the Number of Words Between Entities in the Messages Help?

Redhouane Abdellaoui^{1,2}, MSc; Stéphane Schück², MSc, MD; Nathalie Texier², PharmD; Anita Burgun^{1,3}, MD, PhD

¹INSERM, UMR 1138 Team 22, Université Pierre et Marie Curie, Paris, France

²Kappa Santé, Innovation, Paris, France

³Assistance Publique-Hôpitaux de Paris (AP-HP), Hôpital Européen Georges-Pompidou (HEGP), Medical Informatics, Paris, France

Corresponding Author:

Redhouane Abdellaoui, MSc

INSERM

UMRS 1138 Team 22

Université Pierre et Marie Curie

4 rue de Cléry

Paris, 75002

France

Phone: 33 1 44 82 74 74

Fax: 33 1 44 82 74 75

Email: redhouane.abdellaoui@kappasante.com

Abstract

Background: With the increasing popularity of Web 2.0 applications, social media has made it possible for individuals to post messages on adverse drug reactions. In such online conversations, patients discuss their symptoms, medical history, and diseases. These disorders may correspond to adverse drug reactions (ADRs) or any other medical condition. Therefore, methods must be developed to distinguish between false positives and true ADR declarations.

Objective: The aim of this study was to investigate a method for filtering out disorder terms that did not correspond to adverse events by using the distance (as number of words) between the drug term and the disorder or symptom term in the post. We hypothesized that the shorter the distance between the disorder name and the drug, the higher the probability to be an ADR.

Methods: We analyzed a corpus of 648 messages corresponding to a total of 1654 (drug and disorder) pairs from 5 French forums using Gaussian mixture models and an expectation-maximization (EM) algorithm.

Results: The distribution of the distances between the drug term and the disorder term enabled the filtering of 50.03% (733/1465) of the disorders that were not ADRs. Our filtering strategy achieved a precision of 95.8% and a recall of 50.0%.

Conclusions: This study suggests that such distance between terms can be used for identifying false positives, thereby improving ADR detection in social media.

(*JMIR Public Health Surveill* 2017;3(2):e36) doi:[10.2196/publichealth.6577](https://doi.org/10.2196/publichealth.6577)

KEYWORDS

pharmacovigilance; social media; text mining; Gaussian mixture model; EM algorithm; clustering; density estimation

Introduction

Background

Adverse drug reactions (ADRs) cause millions of injuries worldwide each year and require billions of Euros in associated costs [1,2]. Drug safety surveillance targets the detection, assessment, and prevention of ADRs in the postapproval period. A promise of augmenting drug safety with patient-generated

data drawn from the Internet was called for by several scientific committees related to pharmacovigilance in the United States and in Europe [3,4].

There are now sites for consumers that enable patients to report ADRs. Patients who experience ADRs want to contribute drug safety content, share their experience, and obtain information and support from other Internet users [5-8].



Web-based signal using medical forums in france from 2005-2015



Marie-Laure Kürzinger MSc, Nathalie Texier PharmD, Stéphane Schuck MD, MSc(2), Carole Faviez MSc(2), Thierry Deliens MSc(3), Ling Zhang MSc(4), Stéphanie Tcherny-Lessenot MD, MSc(1), Juhaeji Juhaeji PhD(4), Susan Welsh MD.

Web-based signal detection using medical forums in France from 2005-2015

Abstract #750716

Marie-Laure Kürzinger MSc(1), Nathalie Texier PharmD(2), Stéphane Schuck MD, MSc(2), Carole Faviez MSc(2), Thierry Deliens MSc(3), Ling Zhang MSc(4), Stéphanie Tcherny-Lessenot MD, MSc(1), Juhaeji Juhaeji PhD(4), Susan Welsh MD(4).

1 Sanofi, Global Pharmacovigilance and Epidemiology, Chilly-Mazarin, France

2 Kappa Santé, France

3 Sanofi, Information Technology & Solutions, Chilly-Mazarin, France

4 Sanofi, Global Pharmacovigilance and Epidemiology, Bridgewater, New Jersey, United States

Background

- ✓ Traditional signal detection methods in pharmacovigilance are based on individual case safety reports
- ✓ The use of web-based data (such as social media) is emerging among regulators, industry and academia.
- ✓ The strength of web-based data relies on their real time availability allowing early signal detection.

Objectives

- ✓ The study aims at assessing the ability of identifying early signals from web-based patient's medical forums in France on 3 products over the last 11 years:
- To compare the detected signals from the patient's medical forums in France to signals detected in VigiBase® (cumulative)
- To evaluate time difference in the detection of signals from the patient forums and from VigiBase®

Methods

✓ Data sources

- Data were extracted from the Detect database, composed of messages posted on patients' forums between 2005 and 2015
- 13 French patients forums: Atout, Docissimo, E-sante, SanteMedecine, Onmeda, Futura sciences, Psoriasis, AlarmAsso, Morphée, Albi, Renaloo, Allodocteurs, Forum Sclérose en plaques
- WHO adverse events reporting system (VigiBase®)

✓ 8 disproportionality methods/definitions:

- EB05 ≥2; EBMG ≥2; EBMG ≥4; PRR ≥2, N ≥3, Chi-square ≥4; PRR025 ≥1; ROR025≥1; IC025 > 0; RFET p-value ≤ 0.05

✓ Three single agent drugs

- teriflunomide, insulin glargin, zolpidem

- ✓ Comparison of signals detected from the forums to signals detected in VigiBase® was done by describing the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy and ROC curves.

- ✓ For expected signals, time difference in months between the detection date of signals from the patients' forums and date of signals from VigiBase® was provided.

Results

Figure 1. Number of intake messages for the three drugs

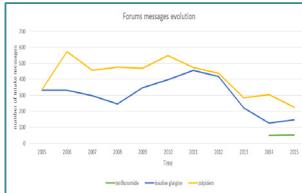


Figure 3. Number of drug-event pairs in VigiBase® and in the Forums

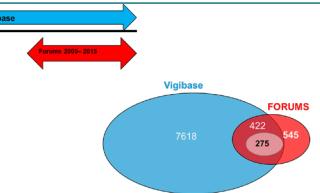


Table 1. Comparison between patient's forums and VigiBase® signals (all drugs together)

Definition	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)
EB05 ≥2	29.06	95.53	62.5	84.03	81.99
EBGM ≥2	48.24	89.28	62.5	82.33	78.19
EBGM ≥4	39.62	94.57	51.21	91.6	87.67
PRR ≥2, N ≥3, Chi-square ≥4	31.93	94.05	67.85	77.86	76.54
Lower 95% CI of PRR ≥1	37.34	87.5	64.13	70	68.72
Lower 95% CI of ROR ≥1	36.96	87.93	66.3	68.48	68
IC025 > 0	33.33	94.18	75.38	72.54	72.98
RFET: p-value ≤ 0.05	50.6	86.09	68.1	74.83	72.98

Figure 4. Specificity and Sensitivity whatever the methods

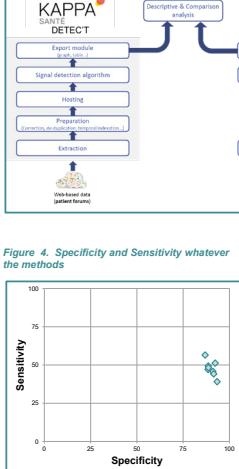
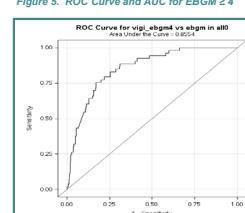


Figure 5. ROC Curve and AUC for EBMG ≥ 4



Conclusions

- ✓ Web-based signal detection using patients' medical forums in France appears to have good performance compared to signals detected in traditional sources. Half of the web-based signals from the French forums are detected or available earlier than in the traditional data sources. Those signals relate to serious medical events as well as patients related symptoms (stress, hunger ...).
- ✓ These results indicate that using patients' medical forums should be considered as a complementary source of data to traditional sources allowing signals to be detected earlier and thus ensuring increased safety of the patients.
- ✓ Further enhancements are needed to investigate the reliability and validation of patient's medical forums worldwide.

Conflict of Interest Disclosure

This study was entirely funded by Sanofi. Authors are employees of Sanofi and the other authors are employees at Kappa Santé which was the company doing the data processing and analysis.

Comparison of Web-based signal detection using medical forums data in France from 2005-2015 with signals from Vigibase®

Schück S, Kürzinger ML, Abdellaoui R, Texier N, Pouget J, Faviez C, Zhang L, Tcherny-Lessenot S, Juhaeri J, Welsh S



Comparison of Web-based signal detection using medical forums data in France from 2005-2015 with signals from Vigibase®



Schück S¹, Kürzinger ML², Abdellaoui R¹, Texier N¹, Pouget J³, Faviez C¹, Zhang L⁴, Tcherny-Lessenot S², Juhaeri J⁴, Welsh S⁴

¹Kappa Santé, Paris, France

²Global Pharmacovigilance and Epidemiology, Sanofi, Chilly-Mazarin, France

³Information Technology Solutions, Sanofi, Lyon, France

⁴Global Pharmacovigilance and Epidemiology, Sanofi, Bridgewater, NJ, USA

Background

Post-marketing drug surveillance is largely based on signals found from spontaneous reports from patients and health care providers. Rare adverse drug reactions and adverse events (AEs) which may develop after long-term exposure to a drug or from drug interactions may be missed. It has been proposed by the FDA and others [1, 2, 3, 4, 5] that web-based data could be mined as a resource to detect latent signals associated with adverse drug reactions. While traditional signal detection methods in PV are based on individual case safety reports, the use of Web-based data (such as, query logs and social media) is emerging among regulators (FDA and EMA), industry and academia. The strength of Web-based data relies on their real time availability allowing early signal detection - 6 months to 1 year earlier than traditional data sources used for signal detection (spontaneous reporting system, electronic medical records, and claims databases).

Objectives

While traditional signal detection methods in pharmacovigilance are based on individual case safety reports, the use of social media data is emerging among regulators and industry. This study aims at assessing the reliability of signals from web-based patient's forums in France on 3 products (teriflunomide, insulin glargine, zolpidem). Signals detected from this source were compared with those from adverse events reporting system.

Methods

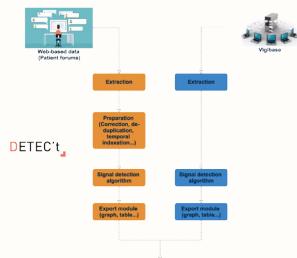


Figure: Extraction & methodology

Data sources

Data were extracted from the *Detect* database, composed of messages posted on patients' forums between 2005 and 2015. Thirteen French patients' forums: *Atoutc*, *Doctissimo*, *E-sante*, *SanteMedecine*, *Ommeda*, *Futura sciences*, *Pсориаз*, *AlarmAsso*, *Morphee*, *Albi*, *Renalo*, *Allodoc-teurs*, *Forum Sclérose en plaques* were screened. Another data source for this study was the WHO adverse events reporting system (VigiBase®) for drug-events comparison. Three drugs teriflunomide, insulin glargin, zolpidem were targeted.

Disproportionality approaches

Methods	Statistics	Criteria
EBGM	EB05	≥ 2
EBGM	EBGM	≥ 2
EBGM	EBGM	≥ 4
PRR Composite	PRR, N, χ^2	PRR ≥ 2 N ≥ 3 $\chi^2 \geq 4$
PRR	LB95(log(PR))	≥ 1
ROR	LB95(log(ROR))	≥ 1
BCPNN	IC025	≥ 1
RFET	p.value	$p \leq 0.05$

Figure: Signals detection strategy

Evaluation

Signals detected from forums were compared to signals detected in VigiBase® by describing sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy and ROC curves.

Results

The number of posts from the patients' forums containing the mention of intake were respectively over the 2005-2015 period: 102, 3326, 4584.

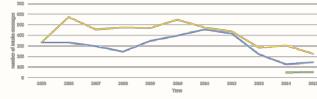


Figure: Messages' intake number for the three drugs

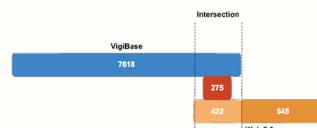


Figure: Drug-event pairs' frequencies

According to metrics used, the sensitivity ranges from 29.1 to 50.6%, the specificity from 86.1 to 95.5%, the PPV from 51.2 to 75.4%, the NPV from 68.5 to 91.6% and the accuracy from 68 to 87.7%.

Method	Sen (%)	Spe (%)	PPV (%)	NPV (%)	Acc (%)
EB05	29.06	95.53	62.5	84.03	81.99
EBGM	48.24	89.28	62.5	82.33	78.19
EBGM	39.62	94.57	51.21	91.6	87.67
Composite	31.93	94.05	67.85	77.86	76.54
PRR	37.34	87.5	64.13	70	68.72
ROR	36.96	87.93	66.3	68.48	68
BCPNN	33.33	94.18	75.38	72.54	72.98
RFET	50.6	86.09	68.1	74.83	72.98

Figure: Comparison between patient's forums and VigiBase® signals (all drugs together)

The AUC reaches 0.85 when considering EBGM ≥ 4

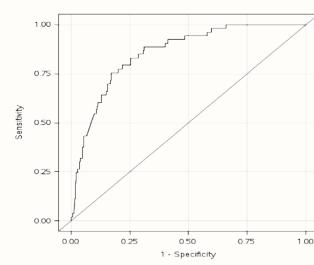


Figure: ROC Curve and AUC for EBGM ≥ 4

The time analysis shows that 30% of the signals are detected earlier (up to 128 months earlier) in the forums than in VigiBase® and 20% are detected at the same time but are available earlier.

Conclusion

Web-based signal detection using forums data in France appears to have good performance features compared to adverse events reporting system. Further analysis will explore time difference between the date of signals from the forums and date of signals from VigiBase®.

Conflict of Interest Disclosure

This study was entirely funded by Sanofi. Authors are employees of Sanofi and the other authors are employees at Kappa Santé which was the company doing the data processing and analysis.

References

- [1] Sloane R, Osanloo O, Lewis D, Bollegala D, Maskell S, Pirmohamed M. Social media and pharmacovigilance: A review of the opportunities and challenges. *Br J Clin Pharmacol.* 2015 Oct;80(4):910-20.
- [2] Sarker A, Ginn R, Nikfarjam A, O'Connor K, Smith K, Jayaraman S, et al. Utilizing social media data for pharmacovigilance: A review. *J Biomed Inform.* 2015 Apr;54:202-12.
- [3] Lardon J, Abdellaoui R, Bellet F, Afshari H, Souvignet J, Texier N, et al. Adverse Drug Reaction Identification and Extraction in Social Media: A Scoping Review. *J Med Internet Res.* 2017 Jul;17(7):e171.
- [4] White RW, Harpaz R, Shah NH, DuMouchel W, Horvitz E. Toward enhanced pharmacovigilance using patient-generated data on the internet. *Clin Pharmacol Ther.* 2014 Aug;96(2):239-46.
- [5] Harpaz R, Callahan A, Tamang S, Low Y, Odgers D, Finlayson S, et al. Text mining for adverse drug events: the promise, challenges, and state of the art. *Drug Saf.* 2014 Oct;37(10):777-90.



Evaluation of Internet Social Networks using Net scoring Tool: A Case Study in adverse drug reaction mining



Sandrine Katsahian, Erica Simond Moreau, Damien Leprovost, Jeremy Lardon, Cedric Bousquet, Gaétan Kerdelhué, Redhouane Abdellaoui, Nathalie Texier, Anita Burgun, Abdelali Boussadi, and Carole Faviez

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,e}, Gaétan Kerdelhué^g, Redhouane Abdellaouiⁱ, Nathalie Texierⁱ, Anita Burgun^{a,c,d,h}, Abdelali Boussadi^{j,a,c,d,h} and Carole Faviezⁱ

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

^b*INSERM, U1142, LIMICS, F-75006, Paris, France*

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

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

^e*University of Saint Etienne, Department of Public Health and Medical Informatics, Saint-Etienne, France.*

^f*Université Paris 13, Sorbonne Paris Cité, F-93430, Villetaneuse, France.*

^g*CISMeF, Rouen University Hospital, Cour Leschevin, Rouen, Cedex, France.*

^h*AP-HP, HEGP, Département d'Informatique Hospitalière, Paris, France.*

ⁱ*Kappa 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



Adverse Drug Reaction Identification and Extraction in Social Media: A Scoping Review



Jérémie Lardon, PhD; Redhouane Abdellaoui, MS; Florelle Bellet, PharmD; Hadyl Asfari, PharmD; Julien Souvignet, MS; Nathalie Texier, PharmD; Marie-Christine Jaulent, PhD; Marie-Noëlle Beyens, MD; Anita Burgun, MD, PhD; Cédric Bousquet, PharmD, PhD

Adverse Drug Reaction Identification and Extraction in Social Media: A Scoping Review

Jérémie Lardon^{1,2*}, PhD; Redhouane Abdellaoui^{3,4*}, MS; Florelle Bellet⁵, PharmD; Hadyl Asfari^{2,6}, PharmD; Julien Souvignet^{2,6}, MS; Nathalie Texier⁴, PharmD; Marie-Christine Jaulent⁶, PhD; Marie-Noëlle Beyens⁷, MD; Anita Burgun^{8,9}, MD, PhD; Cédric Bousquet^{2,6}, PharmD, PhD

¹Université Paris 13, Sorbonne Paris Cité, Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé (LIMICS), (Unité Mixte de Recherche en Santé, UMR_S 1142), F-93430, Villejuif, France, Sorbonne Universités, University of Pierre and Marie Curie (UPMC) Université Paris 06, Unité Mixte de Recherche en Santé (UMR_S) 1142, Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé (LIMICS), F-75006, Institut National de la Santé et de la Recherche Médicale (INSERM), U1142, Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé (LIMICS), F-75006, Paris, France

²Service de Santé Publique et de l'Information Médicale (SSPIM), Department of Public Health and Medical Informatics, Centre Hospitalier Universitaire (CHU) University Hospital of Saint Etienne, Saint-Etienne, France

³Institut National de la Santé et de la Recherche Médicale (INSERM), Unité Mixte de Recherche en Santé (UMR_S) 1138, équipe 22, Centre de Recherche des Cordeliers, Université Paris Descartes, Sorbonne Paris Cité, F-75006, Paris, France

⁴Kappa Santé, Paris, France

⁵Centre de Pharmacovigilance, Centre Hospitalier Universitaire (CHU) University Hospital of Saint Etienne, Saint-Etienne, France

⁶Université Paris 13, Sorbonne Paris Cité, Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé (LIMICS), (Unité Mixte de Recherche en Santé, UMR_S 1142), F-93430, Villejuif, France, Sorbonne Universités, University of Pierre and Marie Curie (UPMC) Université Paris 06, Unité Mixte de Recherche en Santé (UMR_S) 1142, Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé (LIMICS), F-75006, Paris, Institut National de la Santé et de la Recherche Médicale (INSERM), U1142, Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé (LIMICS), F-75006, Paris, France

⁷Centre de Pharmacovigilance, Centre Hospitalier Universitaire (CHU) University Hospital of Saint Etienne, Saint-Etienne, France

⁸Institut National de la Santé et de la Recherche Médicale (INSERM), Unité Mixte de Recherche en Santé (UMR_S) 1138, équipe 22, Centre de Recherche des Cordeliers, Université Paris Descartes, Sorbonne Paris Cité, F-75006, Paris, France

⁹Assistance Publique-Hôpitaux de Paris (AP-HP), Hôpital Européen Georges-Pompidou (HEGP), Department of Medical Informatics, Paris, France

*these authors contributed equally

Corresponding Author:

Jérémie Lardon, PhD

Université Paris 13, Sorbonne Paris Cité, Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé (LIMICS), (Unité Mixte de Recherche en Santé, UMR_S 1142), F-93430, Villejuif, France

Sorbonne Universités, University of Pierre and Marie Curie (UPMC) Université Paris 06, Unité Mixte de Recherche en Santé (UMR_S) 1142, Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé (LIMICS), F-75006

Institut National de la Santé et de la Recherche Médicale (INSERM), U1142, Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé (LIMICS), F-75006

Campus des Cordeliers, Institut National de la Santé et de la Recherche Médicale (INSERM) U 1142 - Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé (LIMICS) Esc D - 2ème étage

15 rue de l'École de Médecine

Paris, 75006

France

Phone: 33 144279190

Fax: 33 144279192

Email: jeremy.lardon@chu-st-etienne.fr

Abstract

Background: The underreporting of adverse drug reactions (ADRs) through traditional reporting channels is a limitation in the efficiency of the current pharmacovigilance system. Patients' experiences with drugs that they report on social media represent a new source of data that may have some value in postmarketing safety surveillance.

Objective: A scoping review was undertaken to explore the breadth of evidence about the use of social media as a new source of knowledge for pharmacovigilance.

