



HAL
open science

Détection automatique des œdèmes aigus pulmonaires de surcharge post-transfusionnels dans les dossiers patients informatisés

Sébastien Cossin, Maryse Puntous, Sophie Pujol, Luc Lebrun, Thomas Ferté,
Vanessa Augis, Isabelle Roger, Vianney Jouhet

► To cite this version:

Sébastien Cossin, Maryse Puntous, Sophie Pujol, Luc Lebrun, Thomas Ferté, et al.. Détection automatique des œdèmes aigus pulmonaires de surcharge post-transfusionnels dans les dossiers patients informatisés. TALMED 2019, Aug 2019, Lyon, France. hal-02288139

HAL Id: hal-02288139

<https://hal.science/hal-02288139>

Submitted on 13 Sep 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Détection automatique des œdèmes aigus pulmonaires de surcharge post-transfusionnels dans les dossiers patients informatisés

Sébastien Cossin^{ab}, Maryse Puntous^c, Sophie Pujol^c, Luc Lebrun^a, Thomas Ferté^a,
Vanessa Augis^c, Isabelle Roger^c, Vianney Jouhet^{ab}

^aCHU de Bordeaux, Pôle de santé publique, Service d'information médicale, Unité Informatique et Archivistique Médicales, F-33000 Bordeaux, France

^bUniversité de Bordeaux, Bordeaux Population Health Research Center, équipe ERIAS, UMR 1219, F-33000 Bordeaux, France

^cCHU de Bordeaux, Pôle de santé publique, Unité de sécurité transfusionnelle et hémovigilance, F-33000 Bordeaux, France

Résumé

Les œdèmes aigus pulmonaires de surcharge post-transfusionnels (TACO) sont des événements indésirables graves consécutifs à la transfusion d'un produit sanguin labile. Bien que ces événements doivent être signalés, les sous-déclarations sont fréquentes. Dans cet article, nous décrivons l'implémentation d'un système de surveillance semi-automatisé basé sur le traitement automatique de la langue des données textuelles des dossiers patients informatisés. Un algorithme détecte les concepts de transfusion et d'œdème pulmonaire dans une même phrase et génère une alerte transmise aux hémovigilants. Plusieurs cas de TACO non signalés ont été détectés par cette approche puis confirmés après enquête d'hémovigilance. Cette approche innovante est susceptible d'aider les unités d'hémovigilance à détecter et surveiller des événements indésirables post-transfusionnels dans les établissements.

Mots clefs:

Transfusion Reaction
Natural Language Processing
Blood Safety

Introduction

L'œdème aigu pulmonaire de surcharge post-transfusionnel est un événement indésirable grave et fréquent d'une transfusion [1,2]. Il se définit comme la survenue d'une détresse respiratoire aigüe survenant dans les 6 heures après une transfusion [3,4]. L'acronyme anglais TACO (pour Transfusion-Associated Circulatory Overload) est souvent utilisé pour les désigner. Les TACO sont responsables de la majorité des décès post-transfusionnels.

Les produits sanguins labiles (PSL) sont issus de dons de sang de donneurs bénévoles et sont destinés à être transfusés à des patients, les receveurs. En France, l'établissement français du sang (EFS) assure le prélèvement, la préparation et la distribution aux établissements des PSL. L'ANSM est chargée de la mise en œuvre de l'hémovigilance au niveau national, de la collecte de sang jusqu'à la transfusion de PSL et aussi du suivi des receveurs [5]. Les événements indésirables liés aux transfusions doivent être déclarés à l'ANSM qui publie chaque année un rapport d'activité. Dans le dernier rapport portant sur l'année 2017, 235 TACO ont été déclarés à l'ANSM soit 7,6 TACO pour 100 000 PSL ou 4,5 pour 10 000 patients transfu-

sés. Les concentrés de globules rouges (CGR) sont les PSL les plus souvent incriminés avec 214 TACO (91%) en 2017. Dans 36% des cas, les recommandations n'avaient pas été respectées. Plusieurs stratégies de prévention existent pour diminuer le risque de survenue comme de ralentir le débit de la transfusion ou de prescrire unité par unité chez un patient à risque. Ces recommandations relatives à la transfusion des CGR ont été publiées dans un rapport de l'ANSM en septembre 2013 [6].

Pour permettre leur surveillance, améliorer les mesures de prévention et diminuer leur incidence dans un établissement de santé ces événements indésirables doivent être déclarés. D'après l'article R1221-49 du code de la santé publique, « tout professionnel de santé qui constate ou a connaissance d'un incident grave le signale sans délai au correspondant d'hémovigilance et de sécurité transfusionnelle de l'établissement ». Cependant, les systèmes de surveillance passive sont connus pour sous-estimer l'incidence des événements indésirables liés aux soins [7]. L'intérêt de méthodes de détection automatique d'événements indésirables à partir des dossiers patients informatisés a déjà été démontré [7,8]. Le traitement automatique de la langue a été utilisé pour la détection d'événements indésirables médicamenteux [9], d'infections associées aux soins [10] ou encore de complications chirurgicales [11]. Une méthode de détection automatique est moins chronophage qu'une détection par recherche manuelle dans les dossiers mais il est nécessaire de la développer et de réussir à l'intégrer aux processus existants pour être utilisée en routine dans un établissement.

Le CHU de Bordeaux a mis en œuvre en novembre 2017 un entrepôt de données de santé basé sur la solution open source i2b2 [12] pour faciliter la réutilisation des données à des fins de recherche, de prévention et d'amélioration de la qualité des soins. Le projet est piloté par le pôle de santé publique et est inscrit au projet d'établissement 2016-2020.

Dans cet article nous décrivons les premiers résultats d'un système d'alerte automatique de détection des TACO par traitement automatique de la langue à partir des données de l'entrepôt et son intégration au circuit de surveillance en place.

Méthodes

Le logiciel TraceLine™ de MAK-System est utilisé au CHU de Bordeaux pour la gestion des PSL. Il enregistre l'ensemble des PSL délivrés. Les PSL sont classifiés en trois grands types : les concentrés de globules rouges (CGR), les concentrés de plaquettes et les plasmas frais congelés. La classification se divise en sous-types, par exemple les CGR phénotypés, les CGR phénotypés étendus, les CGR irradiés etc... Pour chaque PSL délivré, le logiciel enregistre la date, le numéro du patient, le numéro de poche de sang (lié au don de sang) et le sous-type de PSL. Ces données transfusionnelles sont intégrées à l'entrepôt de données i2b2. Celui-ci contient l'ensemble des données structurées (données du Programme de Médicalisation des Systèmes d'Information – PMSI contenant les diagnostics et les actes chirurgicaux de chaque séjour, les administrations médicamenteuses, les résultats biologiques...), semi-structurées (formulaires de question-réponse du logiciel DxCare®) et non structurées (comptes rendus d'hospitalisation, de consultation, d'imagerie...) des patients venus au moins une fois depuis 2010 au CHU. L'entrepôt est alimenté quotidiennement par les données de soins du système d'information hospitalier.

Les données textuelles issues des formulaires et des comptes rendus sont susceptibles de contenir des informations relatives à la transfusion et d'éventuelles complications comme les TACO. Ces données sont produites par les professionnels de santé (médecins et infirmiers notamment) pour la prise en charge des patients et la continuité des soins.

Tous les séjours des patients ayant reçu une transfusion de CGR du 1^{er} janvier 2018 au 20 mars 2019 ont été identifiés. Un programme Java a été développé pour rechercher automatiquement la mention de TACO dans les données en texte libre. Le programme commence par segmenter le texte en phrases avec la bibliothèque Stanford CoreNLP [13]. Un signal de TACO est généré si une phrase contient à la fois un concept de transfusion et un concept d'œdème pulmonaire. Ces deux concepts sont définis par une liste prédéfinie de termes et d'abréviations. Par exemple, la notion de transfusion a été définie par la présence des termes « transfusion, transfusionnel, CGR, culot globulaire, concentré de globules rouges ». La détection de concepts est réalisée par l'algorithme IAMsystem [14] qui utilise une approche par dictionnaire pour la reconnaissance de termes dans le texte. Cet algorithme stocke un dictionnaire de termes dans un arbre après une étape de normalisation (transformation du texte en minuscule, retrait des accents et des mots vides) et de tokénisation (découpage en mots). Il prend en compte les différentes variations possibles d'un mot et ses abréviations qui lui sont fournies en entrées. La distance de Levenshtein est utilisée pour la détection de fautes d'orthographe dans un mot.

Par exemple, en indiquant que le terme « concentrés » est le pluriel de « concentré » et que « GR » est l'abréviation de « globules rouges » l'algorithme est capable de reconnaître que « concentrés de GR » correspond au terme du dictionnaire « concentré de globules rouges ». Le fonctionnement de cet algorithme open source est décrit en détail dans un autre article [14].

Le processus de génération automatique de signaux et de validation est présenté figure 1.

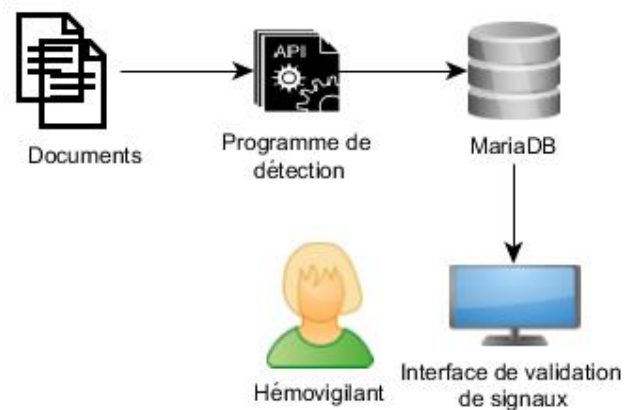


Figure 1– Un programme détecte automatiquement des suspicions de TACO à partir de données textuelles des dossiers patients informatisés. Les signaux sont stockés dans une base de données (MariaDB) et validés par l'unité d'hémovigilance via une interface.

Tous les documents textuels d'un séjour d'un patient transfusé sont extraits de l'entrepôt de données par le langage R. Ces derniers sont ensuite transmis au programme de détection des TACO via une interface de programmation applicative (API). Si ce programme détecte un concept de transfusion et d'œdème pulmonaire dans la même phrase d'un document, la phrase et l'identifiant du document sont enregistrés dans MariaDB, une base de données relationnelles issue de MySQL. Cette base de données contient uniquement les phrases détectées par le programme et l'identifiant du document permettant de retourner au dossier du patient dans le système d'information hospitalier. Les hémovigilants du CHU de Bordeaux se connectent à une interface pour valider chaque phrase détectée par l'algorithme afin de confirmer ou infirmer que celle-ci mentionne la survenue d'un cas incident de TACO (figure 2). Si la phrase est validée, l'interface affiche les identifiants du patient et du séjour. Les hémovigilants peuvent alors retourner au dossier patient informatisé via le logiciel de soins (DxCare®), échanger avec les cliniciens et vali-

Signaux détectésSignaux validés

Signaux d'OAP post-transfusionnel détectés :

- [Signal n°11 détecté le 21/02/2018](#)
- [Signal n°12 détecté le 16/02/2018](#)
- [Signal n°13 détecté le 03/01/2018](#)
- [Signal n°14 détecté le 22/01/2018](#)
- [Signal n°15 détecté le 23/04/2018](#)
- [Signal n°17 détecté le 23/04/2018](#)
- [Signal n°18 détecté le 17/04/2018](#)
- [Signal n°24 détecté le 15/05/2018](#)

Patiente entrée pour OAP post transfusion chez une patiente GIR 2 souffrant d Alzheimer et vivant en EHPAD

Figure 2– Interface de validation des signaux. Les hémovigilants valident ou invalident chaque phrase détectée par l'algorithme pour confirmer que la phrase évoque la survenue d'un cas incident de TACO (œdème pulmonaire post-transfusionnel). Quand l'utilisateur clique sur le bouton Valider, les identifiants du patient et du séjour s'affichent pour permettre un retour au dossier de soins.

der l'évènement indésirable qui sera communiqué à l'ANSM.

Résultats

Du 1^{er} janvier 2018 au 20 mars 2019, 7 877 patients du CHU de Bordeaux ont reçu 80 439 transfusions de CGR. Le programme a généré 319 signaux dans les documents textuels de 169 patients, 53 signaux de 23 dossiers médicaux ont été validés via l'interface de validation. Les principales raisons d'invalidation dans l'interface étaient la présence d'un antécédent de TACO (« Compte tenu antécédent d'OAP sur transfusion de CGR ... ») et la présence d'une négation devant le signe clinique (« transfusion 1 CGR, pas de dyspnée », « nette amélioration après la transfusion, plus de dyspnée »).

Plusieurs cas ont été validés dans l'interface puis invalidés après retour au dossier car des critères cliniques manquaient pour le diagnostic de TACO ou que celui-ci était présent dans une section 'antécédents' du dossier par exemple.

Au total, 11 TACO non déclarés par les cliniciens ont été confirmés après enquêtes d'hémovigilance. Ces cas non déclarés étaient plus nombreux que les cas déclarés sur cette période.

Deux cas non déclarés étaient survenus dans le même service évoquant un manque de mesures préventives avant les transfusions.

L'algorithme a été modifié durant cette période car deux signalements n'avaient pas été détectés tandis que l'information était bien présente dans le dossier. Plusieurs termes comme « désaturation », « surcharge », « décompensation cardiaque » ont été ajoutés à la liste des termes permettant de détecter les œdèmes pulmonaires.

Discussion

Plusieurs cas non déclarés d'œdème aigu pulmonaire post-transfusionnel (TACO) ont été détectés par traitement automatique de la langue. Cette étude démontre une nouvelle fois l'intérêt des méthodes de détection automatique pour la surveillance des évènements indésirables dans les établissements de santé.

Cet évènement indésirable post-transfusionnel nécessite un signalement obligatoire au correspondant d'hémovigilance de l'établissement qui n'est pas systématiquement réalisé par les professionnels de santé faute de rapprochement avec la transfusion ou par manque de connaissance du dispositif en place ou par manque de temps [15]. La détection de cas non signalés permet d'améliorer la surveillance, de rappeler les mesures de prévention et contribue à la sécurité des soins.

Ce nouveau système d'alerte automatique est complémentaire avec le signalement d'évènements indésirables réalisé par les professionnels de santé. L'outil est simple d'utilisation et très peu chronophage car une seule phrase est présentée à l'utilisateur.

Limites

Bien que le système d'alerte ait démontré un intérêt pour l'hémovigilance, il présente de nombreuses limites et peut être amélioré.

Premièrement, ses performances en termes de sensibilité et spécificité n'ont pas pu être évaluées en l'absence de gold standard. Une évaluation précise des performances serait possible par une surveillance active pour mesurer l'incidence des TACO. Celle-ci nécessite la mise en œuvre de moyens supplémentaires pour évaluer et objectiver, par une méthode rigoureuse, les complications de chaque transfusion [16]. Des cas déclarés par les professionnels de santé non détectés par le

programme ont mis en évidence un manque de sensibilité. Le programme a été modifié et amélioré par l'ajout de nouveaux termes.

La méthode de traitement automatique de la langue utilisée est relativement simple et présente plusieurs limites. La segmentation par phrases est une étape fréquente dans l'analyse de documents médicaux [17] mais restreindre l'analyse au niveau de chaque phrase entraîne très probablement une diminution de la sensibilité. Ce choix a été réalisé pour faciliter le travail de validation par la lecture d'un texte court qui nécessite quelques secondes pour valider ou invalider un signal. D'autres paramétrages pour la génération de signaux pourraient être testés comme fixer une distance de chaînes de caractères entre le concept de transfusion et celui d'œdème pulmonaire ou classifier automatique la relation entre ces deux concepts.

La spécificité serait nettement améliorable par la détection de la négation et des antécédents, des algorithmes existent pour la langue française [18,19, 20].

Utiliser uniquement les données en texte libre présente aussi une limite pour la détection des TACO. La survenue de l'évènement indésirable n'est pas toujours retranscrite dans le dossier patient informatisé et d'autres sources de données du système information hospitalier pourraient être utilisées. Pour détecter les TACO dans les dossiers patients informatisés, Clifford et al. [16] ont développé deux algorithmes, le premier à base de règles et le second par apprentissage automatique, à partir des données structurées suivantes : la saturation artérielle en oxygène, la fréquence respiratoire, des marqueurs sanguins (niveau de BNP), et des signes radiologiques sur une radiographie pulmonaire 8 heures après la transfusion. Ils obtiennent une sensibilité et une spécificité de 86,5% et 92,3% respectivement. Ces données de surveillance au lit du patient ne sont pas présentes dans l'entrepôt du CHU de Bordeaux.

Une autre limite de ce travail est son caractère unicentrique. La capacité à détecter automatiquement des signaux non déclarés dépend du niveau d'informatisation de l'établissement, du niveau de remplissage des dossiers patients informatisés et du taux de signalement des évènements indésirables par les professionnels de santé.

Perspectives

La détection de la négation et des antécédents devrait permettre d'améliorer la spécificité de l'algorithme.

La validation manuelle de signaux a généré un jeu de données annoté. Celui est utile pour entraîner et ajouter un algorithme de classification supervisée. Ces algorithmes pourraient prendre en compte certains facteurs de risque de TACO pour améliorer leur performance. Ils devront cependant conserver une excellente sensibilité car les faux positifs sont déjà rapidement écartés et les faux négatifs ne sont pas acceptables.

Ce premier algorithme pourra servir de référence pour l'évaluation d'autres algorithmes plus sophistiqués. Cette approche de détection automatique de signaux associée à une interface de validation pourrait s'appliquer à d'autres évènements comme les effets indésirables médicamenteux, les infections nosocomiales ou certaines complications chirurgicales.

Le système développé comprenant la détection de signaux et l'interface de validation est publié open source¹ pour per-

¹ <https://github.com/scossin/TACO>

mettre son installation locale dans un autre établissement. Son déploiement nécessite la capacité d'extraire les données textuelles des séjours de patients transfusés dans le système d'information hospitalier. Son implémentation au CHU de Bordeaux a été facilitée par la présence d'un entrepôt de données qui regroupe l'ensemble de ces données. Ce système d'alerte automatique doit s'intégrer à un circuit de sécurité transfusionnelle déjà existant.

Conclusion

Le traitement automatique de la langue a permis de détecter des cas non déclarés d'œdème pulmonaire post-transfusionnel mentionné dans les données en texte libre des dossiers patients informatisés. Le système d'alerte automatique proposé dans cet article utilise une approche simple par dictionnaire de termes pour détecter ces événements et générer des alertes. Les alertes sont ensuite validées ou invalidées via une interface par les hémovigilants.

Les premiers résultats donnent satisfaction même si plusieurs pistes d'amélioration de l'algorithme ont été mises en évidence. Ce système permet d'améliorer la surveillance de ces événements indésirables dont le signalement est obligatoire. Son implémentation au CHU de Bordeaux a été facilitée par la présence d'un entrepôt de données de santé et le système est publié en open source pour tester sa transposabilité à d'autres établissements.

Références

- [1] N. Roubinian, TACO and TRALI: biology, risk factors, and prevention strategies, *Hematology Am Soc Hematol Educ Program*. **2018** (2018) 585–594. doi:10.1182/asheducation-2018.1.585.
- [2] G. Li, M. Kojicic, M.K. Reriani, E.R. Fernández Pérez, L. Thakur, R. Kashyap, C.M. Van Buskirk, and O. Gajic, Long-Term Survival and Quality of Life After Transfusion-Associated Pulmonary Edema in Critically Ill Medical Patients, *Chest*. **137** (2010) 783–789. doi:10.1378/chest.09-0841.
- [3] G. Li, S. Rachmale, M. Kojicic, K. Shahjehan, M. Malinchoc, D.J. Kor, and O. Gajic, Incidence and transfusion risk factors for transfusion-associated circulatory overload among medical intensive care unit patients, *Transfusion*. **51** (2011) 338–343. doi:10.1111/j.1537-2995.2010.02816.x.
- [4] N.H. Roubinian, J.E. Hendrickson, D.J. Triulzi, J.L. Gottschall, D. Chowdhury, D.J. Kor, M.R. Looney, M.A. Matthay, S.H. Kleinman, D. Brambilla, and E.L. Murphy, Incidence and clinical characteristics of TACO using an active surveillance algorithm, *Vox Sang*. **112** (2017) 56–63. doi:10.1111/vox.12466.
- [5] Rapport annuel d'hémovigilance 2017 - Point d'information - ANSM : Agence nationale de sécurité du médicament et des produits de santé, (n.d.). <https://ansm.sante.fr/S-informer/Points-d-information-Points-d-information/L-ANSM-publie-le-rapport-annuel-d-hemovigilance-2017-Point-d-information>.
- [6] ANSM, Les œdèmes aigus pulmonaires de surcharge post-transfusionnels. Rapport. Septembre 2013.
- [7] D.W. Bates, R.S. Evans, H. Murff, P.D. Stetson, L. Piziferri, and G. Hripsak, Detecting Adverse Events Using Information Technology, *J Am Med Inform Assoc*. **10** (2003) 115–128. doi:10.1197/jamia.M1074.
- [8] C.M. Rochefort, D.L. Buckeridge, and A.J. Forster, Accuracy of using automated methods for detecting adverse events from electronic health record data: a re-search protocol, *Implementation Science*. **10** (2015) 5. doi:10.1186/s13012-014-0197-6.
- [9] Y. Luo, W.K. Thompson, T.M. Herr, Z. Zeng, M.A. Berendsen, S.R. Jonnalagadda, M.B. Carson, and J. Starren, Natural Language Processing for EHR-Based Pharmacovigilance: A Structured Review, *Drug Saf*. **40** (2017) 1075–1089. doi:10.1007/s40264-017-0558-6.
- [10] N. Tvardik, I. Kergourlay, A. Bittar, F. Segond, S. Darmoni, and M.-H. Metzger, Accuracy of using natural language processing methods for identifying healthcare-associated infections, *Int J Med Inform*. **117** (2018) 96–102. doi:10.1016/j.ijmedinf.2018.06.002.
- [11] H.J. Murff, F. FitzHenry, M.E. Matheny, N. Gentry, K.L. Kotter, K. Crimin, R.S. Dittus, A.K. Rosen, P.L. Elkin, S.H. Brown, and T. Speroff, Automated identification of postoperative complications within an electronic medical record using natural language processing, *JAMA*. **306** (2011) 848–855. doi:10.1001/jama.2011.1204.
- [12] S.N. Murphy, M.E. Mendis, D.A. Berkowitz, I. Kohane, and H.C. Chueh, Integration of clinical and genetic data in the i2b2 architecture, *AMIA Annu Symp Proc*. (2006) 1040.
- [13] C.D. Manning, M. Surdeanu, J. Bauer, J.R. Finkel, S. Bethard, and D. McClosky, The Stanford CoreNLP Natural Language Processing Toolkit, in: ACL, 2014. doi:10.3115/v1/P14-5010.
- [14] S. Cossin, V. Jouhet, F. Mougin, G. Diallo, and F. Thiessard, IAM at CLEF eHealth 2018: Concept Annotation and Coding in French Death Certificates, *ArXiv:1807.03674 [Cs]*. (2018). <http://arxiv.org/abs/1807.03674>.
- [15] C. Narick, D.J. Triulzi, and M.H. Yazer, Transfusion-associated circulatory overload after plasma transfusion, *Transfusion*. **52** (2012) 160–165. doi:10.1111/j.1537-2995.2011.03247.x.
- [16] L. Clifford, A. Singh, G.A. Wilson, P. Toy, O. Gajic, M. Malinchoc, V. Herasevich, J. Pathak, and D.J. Kor, Electronic health record surveillance algorithms facilitate the detection of transfusion-related pulmonary complications, *Transfusion*. **53** (2013) 1205–1216. doi:10.1111/j.1537-2995.2012.03886.x.
- [17] D. Griffis, C. Shivade, E. Fosler-Lussier, and A.M. Lai, A Quantitative and Qualitative Evaluation of Sentence Boundary Detection for the Clinical Domain, *AMIA Jt Summits Transl Sci Proc*. **2016** (2016) 88–97.
- [18] A. Abdaoui, A. Tchechmedjiev, W. Digan, S. Bringay, and C. Jonquet, French ConText: Détecter la négation, la temporalité et le sujet dans les textes cliniques Français, in: SIIM: Symposium Sur l'Ingénierie de l'Information Médicale, Toulouse, France, 2017. <https://hal-lirmm.ccsd.cnrs.fr/lirmm-01656834>
- [19] N. Garcelon, A. Neuraz, V. Benoit, R. Salomon, and A. Burgun, Improving a full-text search engine: the importance of negation detection and family history context to identify cases in a biomedical data warehouse, *J Am Med Inform Assoc*. **24** (2017) 607–613. doi:10.1093/jamia/ocw144.
- [20] W.W. Chapman, D. Hillert, S. Velupillai, M. Kvist, M. Skeppstedt, B.E. Chapman, M. Conway, M. Tharp, D.L. Mowery, and L. Deleger, Extending the NegEx lexicon for multiple languages, *Stud Health Technol Inform*. **192** (2013) 677–681.

Adresse de correspondance

sebastien.cossin@chu-bordeaux.fr