

Potential Blood Transfusion Adverse Events Can be Found in Unstructured Text in Electronic Health Records using Natural Language Processing Tools

Roselie A Bright, ScD (OC/OCS/OHI)
 Summer K Rankin, PhD (Booz Allen Hamilton)
 Katherine Dowdy (Booz Allen Hamilton)
 Serge Blok, PhD (Booz Allen Hamilton)
 Susan J. Bright-Ponte, DVM, MPH (CVM/OSC/DS)
 Lee Anne Palmer, VMD, MPH (CVM/OSC/VPS)

Supported by FDA contract. Presentation is not necessarily official view of FDA or Booz Allen Hamilton. We gratefully acknowledge our colleagues' support.

INTRODUCTION

Goal: Develop a new method to identify adverse event (AE) signals using the unstructured text of electronic health records (EHRs):

- Independent of vocabulary used.
- Even if AE is new.
- Even if writer does not attribute event as AE.

We chose potential transfusion adverse events (PTAEs) as a proof-of-concept because:

- Real dates in MIMIC-III are obscured.
- Emerging recognition of new TAE during study data period.
- FDA received relatively few reports.

METHODS



Clinical notes

- MIMIC III
- 2001-2012
- Teaching hospital
- Critical care

Text Notes Preparation

- Concatenated each admission's notes in chronological order.
- De-duplicated notes (Bloatectomy).
- Vectorized each admission's remaining notes.

Group Creation & Term Extraction

- Transfused (T) (21,443 admissions)
- Comparison (C) (27,888 admissions)

Used ensemble of supervised classification methods and statistical rules to filter notes to unusual terms that distinguish T from C.

Topic Analysis of T

Leveraged Latent Dirichlet Allocation topic modeling to derive 45 topics.

Reviewed admissions:

- Top topic scores.
- Random selection from T.

RESULTS

Most PTAE in the notes were not attributed to be related to transfusion.

Admissions with a top-scoring cardiovascular topic:

- Heart valve repair
- Tapped pericardial effusion
- Coronary artery bypass graft
- Heart attack
- Vascular repair



... were more likely than random T admissions (proportion difference = 0.47, $p = 0.022$) to have at least one **heart PTAE**:

- *heart rhythm changes*
- *hypotension*

Admissions with a top-scoring pulmonary topic:

- Mechanical ventilation
- Acute respiratory distress syndrome
- Inhaled nitric oxide



... were more likely than random T admissions (proportion difference = 0.37, $p = 0.049$) to have at least one **lung PTAE**:

- *hypoxia*
- *mechanical ventilation*
- *bilateral pulmonary effusion*
- *pulmonary edema*

SUCCESSSES

- Could be a useful supplemental post-marketing surveillance method for generating hypotheses to be studied by finding unattributed AEs
- Unlocks clinical text notes
- Uses open source software
- Needs relatively few computing resources
- Adaptable to other settings

The *Shakespeare* Project

Literature big data tools applied to clinical records



FUTURE

- Automated support of review step
- We invite new members to the Shakespeare Project Team!

Potential Blood Transfusion Adverse Events Can be Found in Unstructured Text in Electronic Health Records using Natural Language Processing Tools

REFERENCES

Introduction

Carson JL, Grossman BJ, Kleinman S, et al. Red blood cell transfusion: a clinical practice guideline from the AABB. *Ann Intern Med*. 2012; 157: 49-58.

Holness L, Knippen MA, Simmons L, Lachenbruch PA. Fatalities caused by TRALI. *Transfusion Medicine Reviews*. 2004; 18(3): 184-188. <https://doi.org/10.1016/j.tmr.2004.03.004>.

Menis M, Anderson SA, Forshee RA, et al. Transfusion-associated circulatory overload (TACO) and potential risk factors among the inpatient US elderly as recorded in Medicare administrative databases during 2011. *Vox Sang*. 2014 Feb; 106(2): 144-52. doi: 10.1111/vox.12070.

Menis M, Anderson SA, Forshee RA, et al. Transfusion-related acute lung injury and potential risk factors among the inpatient US elderly as recorded in Medicare claims data, during 2007 through 2011. *Transfusion*. 2014 Sep; 54(9): 2182-93. doi: 10.1111/trf.12626.

Methods

Johnson AEW, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. *Sci Data*. 2016; 3: 160035. <https://doi.org/10.1038/sdata.2016.35>.

MIMIC-III Critical Care Database. <https://mimic.physionet.org/about/mimic/>.

Rankin SK, Bright R, Dowdy K. Bloatectomy (Version v0.0.12). Zenodo. 2020, June 26. <http://doi.org/10.5281/zenodo.3909030>.

Mikolov T, Sutskever I, Chen K, et al. Distributed Representations of Words and Phrases and their Compositionality. arXiv. 2013; 1310.4546. <https://arxiv.org/abs/1310.4546>.

Radim R, Sojka P. Software Framework for Topic Modelling with Large Corpora. Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks. 2010; 45-50. <http://is.muni.cz/publication/884893/en>.

Blei D, Ng A Jordan M. Latent Dirichlet Allocation. *Journal of Machine Learning Research* 2003; 3: 993-1022.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 2011; 12: 2825-2830. <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>.

sklearn.feature_extraction.text.CountVectorizer. Scikit-learn Machine Learning in Python. Scikit-learn developers. 2020. http://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html.

Tang B, Kay S and He H. Toward optimal feature selection in naive Bayes for text categorization. arXiv. 2016; 1602.02850. <https://arxiv.org/abs/1602.02850>.

Azizi S, Liu Y, Ju S, Wang J, Su C. A New Feature Selection Method for Text Classification Based on Independent Feature Space Search. *Mathematical Problems in Engineering*. 2020 May 12. 2020. <https://doi.org/10.1155/2020/6076272>.

Ng, A. Feature Selection, L1 vs. L2 regularization and rotational invariance. Proceedings of the 21st International Conference on Machine Learning. 2004. <https://icml.cc/Conferences/2004/proceedings/papers/354.pdf>.

International Society of Blood Transfusion Working Party on Haemovigilance. Proposed standard definitions for surveillance of non infectious adverse transfusion reactions: Incorporating correction to TRALI definition (as adopted June 2013). International Haemovigilance Network. 2011, July. http://www.isbtweb.org/fileadmin/user_upload/files-2015/haemovigilance/definitions/Proposed%20definitions%202011%20surveillance%20non%20infectious%20adverse%20reactions%20haemovigilance%20incl%20TRALI%20correction%202013.pdf.

Sahu S, Hemlata, Verma A. Adverse events related to blood transfusion. *Indian J Anaesth*. 2014; 58(5): 543-551. Doi: 10.4103/0019-5049.144650.

Juffermans NP, Wals, h. TS. Introduction. IN: Juffermans NP, Walsh TS, eds. *Transfusion in the Intensive Care Unit*. Springer, New York. 2015; 1-4. DOI 10.1007/978-3-319-08735-1_1.

Preacher, KJ, Briggs, NE. Calculation for Fisher's Exact Test: An interactive calculation tool for Fisher's exact probability test for 2 x 2 tables. *Quantpsy*. 2001, May. <http://quantpsy.org/fisher/fisher.htm>.

Interpretation

Halamka J. Dispatch from the digital health frontier/2007/. 2007. <http://geekdoctor.blogspot.com>.

Krishnamoorthy P, Mukherjee D, Chatterjee S. Red blood cell transfusion trigger in cardiac disease. IN: Juffermans NP, Walsh TS, eds. *Transfusion in the Intensive Care Unit*. Springer, New York. 2015; 25-34. DOI 10.1007/978-3-319-08735-1_4.

Snyder MJ, Bepko J, White M. Acute pericarditis: diagnosis and management. *Am Fam Physician*. 2014 Apr 1; 89(7): 553-560. <https://www.aafp.org/afp/2014/0401/p553.html>.

Murphy GJ, Patel NN, Sterne JAC. Red blood cell transfusion trigger in cardiac surgery. IN: Juffermans NP, Walsh TS, eds. *Transfusion in the Intensive Care Unit*. Springer, New York. 2015; 35-44. DOI 10.1007/978-3-319-08735-1_5.

Tracheostomy. National Heart Lung and Blood Institute. US Department of Health and Human Services. <https://www.nhlbi.nih.gov/health-topics/tracheostomy>.

Acute respiratory distress syndrome. National Heart Lung and Blood Institute. US Department of Health and Human Services. 2019. <https://www.nhlbi.nih.gov/health-topics/acute-respiratory-distress-syndrome>.

Hemodialysis. Health Information Center. National Institute of Diabetes and Digestive and Kidney Diseases, US National Institutes of Health. 2018. <https://www.niddk.nih.gov/health-information/kidney-disease/kidney-failure/hemodialysis>.

Davenport RD. Hemolytic transfusion reactions. IN: Simon TL, Snyder EL, Solheim BG, et al, eds. *Rossi's principles of transfusion medicine*, 4th ed. Blackwell Publishing Ltd. 2009; 811- 825.

Butcher BW, Liu KD. Fluid Overload in AKI - Epiphenomenon or Putative Effect on Mortality? *Curr Opin Crit Care*. 2012 Dec; 18(6): 593-598. doi: 10.1097/MCC.0b013e32835a1c44.

International Society of Blood Transfusion (ISBT), International Haemovigilance Network. *Transfusion-associated circulatory overload (TACO): draft revised criteria*. ISBT. 2017. http://www.isbtweb.org/fileadmin/user_upload/Proposed_definitions_2011_surveillance_non_infectious_adverse_reactions_haemovigilance_incl_TRALI_correction_2013.pdf.

Graphics

<http://wesharepics.info/imagesgkl-sick-hospital-patient-in-bed.asp>
 Heart. <https://pngimg.com/download/74903>
 Lungs. <https://pngimg.com/download/74903>