

New and Increasing Rates of Adverse Events Can be Found in Unstructured Text in Electronic Health Records using Natural Language Processing Tools

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

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

INTRODUCTION

Our goal was to 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 an AE.

We chose to explore clinical care notes from 2001-2008, to see if we could detect contaminated heparin AEs.

- Heparin contamination AEs in the US were recognized by FDA in early 2008, although some newspaper articles suggested contamination as early as 2005.
- Heparin is an anticoagulant used in surgeries, blood transfusions, and catheterizations.

METHODS



- ### Clinical Notes
- MIMIC III
 - 7/2001-6/2008
 - Teaching hospital
 - Critical care
 - Dates within several weeks of real dates

- ### Text Notes Preparation
- Concatenated notes in time order for each admission.
 - De-duplicated notes using *Bloatectomy*.
 - Vectorized each admission's remaining notes.

- ### Group Creation & Term Extraction
- Separated admissions by time period:
- Period 1, 7/2001-6/2006 (14,441).
 - Period 2, 7/2006-6/2007 (3,581).
 - Period 3, 7/2007-6/2008 (3,296).

Used ensemble of supervised classification methods and statistical rules to filter notes to unusual terms that distinguish period 3 from period 2.

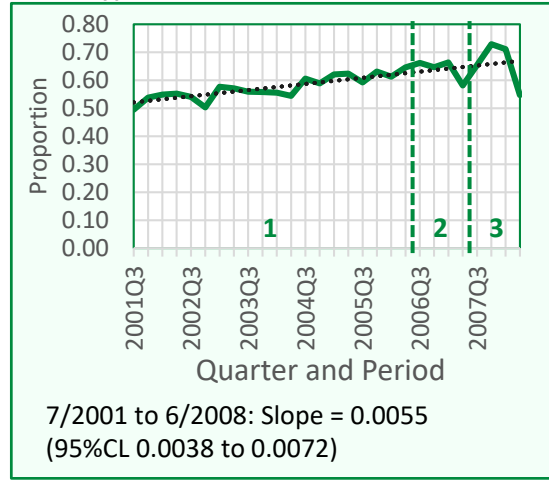
- ### Topic Analysis of Period 3
- Leveraged Latent Dirichlet Allocation topic modeling to organize notes into 20 topics.
 - Read admissions with top topic scores.
 - Interpreted topics within context of frequency analysis across periods 1-3.

RESULTS

Period 3 topics (showed increases over periods 1-3) for conditions that are:

Uncommon (examples)	Common (examples)
<ul style="list-style-type: none"> • Hypotension right after heparin exposure • Attributed drug AEs • Admissions for attributed therapy/ food AEs • Rare conditions • Continuing drainage after surgery 	<ul style="list-style-type: none"> • Long hospital stays • Heart attacks • Brain trauma • Brain bleed • Brain ischemia

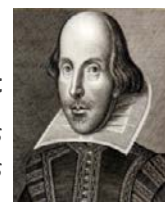
Proportion of invasive cardiovascular procedure admissions (heparin presumed) with "hypotension" word:



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!

New and Increasing Rates of Adverse Events Can be Found in Unstructured Text in Electronic Health Records using Natural Language Processing Tools

REFERENCES

Introduction

Baxter Issues Urgent Nationwide Voluntary Recall of Heparin 1,000 Units/ml 10 and 30ml Multi-Dose Vials NDC NUMBERS 0641-2440-45, 0641-2440-41, 0641-2450-45 and 0641-2450-41; LOTS: 107054, 117085, 047056, 097081, 107024, 107064, 107066, 107074, 107111. Food and Drug Administration. 2008 January 25. <http://wayback.archive-it.org/7993/20170111131710/http://www.fda.gov/Safety/Recalls/ArchiveRecalls/2008/default.htm?Page=5>. Information on heparin. Food and Drug Administration. 2017. <https://wayback.archive-it.org/7993/20170722214801/https://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/UCM112597>.

Acute Allergic-Type Reactions Among Patients Undergoing Hemodialysis — Multiple States, 2007–2008. MMWR. 2008, February 8; 57(5): 124-125.

Lyn TE. China pig disease caused by new strain: experts. Reuters. 2007 June 26. <https://www.reuters.com/article/us-china-disease-pig-idUSHKG26819620070626>.

Barboza D. Virus Spreading Alarm and Pig Disease in China. New York Times. 2007, August 16. <http://www.nytimes.com/2007/08/16/business/worldbusiness/16pigs.html>.

Tian K, Yu X, Zhao T, et al. Emergence of fatal PRRSV variants: unparalleled outbreaks of atypical PRRS in China and molecular dissection of the unique hallmark. PLoS ONE 2(6): e526. doi:10.1371/journal.pone.0000526.

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>.

LINEST function. Microsoft Support. 2020. <https://support.microsoft.com/en-us/office/linest-function-84d7d0d9-6e50-4101-977a-fa7abf772b6d>.

Altman DG, Bland JM. How to obtain the P value from a confidence interval. *BMJ*. 2011; 343. <https://doi.org/10.1136/bmj.d2304>.

Invasive cardiovascular procedure codes were ICD9CM 3891, 3961, 3965-3966, from MIMIC-III definitions table.

Interpretation

Kern M. Hypotension in the Cath Lab? Think vagal reaction early. *Cath Lab Digest*. HMP CardioVascular. 2020. <https://www.cathlabdigest.com/articles/Hypotension-Cath-Lab-Think-Vagal-Reaction-Early>.

Information on heparin. Food and Drug Administration. 2017. <https://wayback.archive-it.org/7993/20170722214801/https://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/UCM112597>.

Acute Allergic-Type Reactions Among Patients Undergoing Hemodialysis — Multiple States, 2007–2008. MMWR. 2008, February 8; 57(5): 124-125.

Lyn TE. China pig disease caused by new strain: experts. Reuters. 2007 June 26.

<https://www.reuters.com/article/us-china-disease-pig-idUSHKG26819620070626>.

Barboza D. Virus Spreading Alarm and Pig Disease in China. New York Times. 2007, August 16. <http://www.nytimes.com/2007/08/16/business/worldbusiness/16pigs.html>.

Tian K, Yu X, Zhao T, et al. Emergence of fatal PRRSV variants: unparalleled outbreaks of atypical PRRS in China and molecular dissection of the unique hallmark. PLoS ONE 2(6): e526. doi:10.1371/journal.pone.0000526.

Von Hippel-Lindau disease. GARD Genetic and Rare Diseases Information Center. National Center for Advancing Translational Sciences, National Institutes of Health. 2016 February 2.

<https://rarediseases.info.nih.gov/diseases/7855/von-hippel-lindau-disease>.

Levy P. The Harvard medical system. Not Running a Hospital. 2007, January 14.

<http://runningahospital.blogspot.com/2007/01/harvard-medical-system.html>.

Cunha JP. Vancomycin. RxList. 2020. https://www.rxlist.com/consumer_vancomycin/drugs-condition.htm.

Reliable 1 Laboratories LLC. Aspirin 325 MG- aspirin tablet. Daily Med, National Library of Medicine, National Institutes of Health. 2020.

<https://dailymed.nlm.nih.gov/dailymed/drugInfo.cfm?setid=28c48336-c238-0141-e054-00144ff8d46c>.

Bristol-Myers Squibb Pharma Company. COUMADIN- warfarin sodium tablet. Daily Med, National Library of Medicine, National Institutes of Health. 2020.

<https://dailymed.nlm.nih.gov/dailymed/drugInfo.cfm?setid=d91934a0-902e-c26c-23ca-d5acc4151b6>.

Actavis Pharma, Inc. DOXORUBICIN HYDROCHLORIDE injection, solution. Daily Med, National Library of Medicine, National Institutes of Health. 2020.

<https://dailymed.nlm.nih.gov/dailymed/drugInfo.cfm?setid=33ddfdb1-b4bc-464d-a186-49c4b156279f>.

Misal US, Joshi SA, Shaikh MM. Delayed recovery from anesthesia: A postgraduate educational review. *Anesth Essays Res*. 2016 May-Aug; 10(2): 164–172. doi: 10.4103/0259-1162.165506.

Shah T, Van Dam T, Madamangalam. Epidural pump malfunction. *Anaesthesia*. 2017; 72(2). <https://doi.org/10.1111/anae.13797>.

Grau D, Clarivet B, Lotthé A, et al. Complications with peripherally inserted central catheters (PICCs) used in hospitalized patients and outpatients: a prospective cohort study. *Antimicrob Resist Infect Control*. 2017; 6: 18. doi: 10.1186/s13756-016-0161-0.

A Study of ARQ 171 in Patients With Advanced Solid Tumors. ClinicalTrials.gov Identifier: NCT00398840. National Library of Medicine, National Institutes of Health. 2020.

<https://clinicaltrials.gov/ct2/show/NCT00398840>.

Graphics

<http://wesharepics.info/imagesgkl-sick-hospital-patient-in-bed.asp>