Data Article

Association model data learned from clinicians stratified by patient mortality outcomes at a Tertiary Academic Center

Jason K. Wang, Jason Hom, Santhosh Balasubramanian, Jonathan H. Chen

Mathematical and Computational Science Program, Stanford University, Stanford, CA, USA
Department of Medicine, Stanford University, Stanford, CA, USA
Center for Biomedical Informatics Research, Stanford University, Stanford, CA, USA
Division of Hospital Medicine, Stanford University, Stanford, CA, USA

Article info

Article history:
Received 8 September 2018
Received in revised form
26 October 2018
Accepted 30 October 2018

Abstract

In this data article, we learn clinical order patterns from inpatient electronic health record (EHR) data at a tertiary academic center from three different cohorts of providers: (1) Clinicians with lower-than-expected patient mortality rates, (2) clinicians with higher-than-expected patient mortality rates, and (3) an unfiltered population of clinicians. We extract and make public these order patterns learned from each clinician cohort associated with six common admission diagnoses (e.g. pneumonia, chest pain, etc.). We also share a reusable reference standard or benchmark for evaluating automatically-learned clinical order patterns for each admission diagnosis, based on a manual review of clinical practice literature. The data shared in this article can support further study, evaluation, and translation of data-driven CDS systems. Further interpretation and discussion of this data can be found in Wang et al. (2018).

© 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Specifications table

| Subject area                           | Medical informatics                                |
|----------------------------------------|----------------------------------------------------|
| More specific subject area             | Clinical decision support, machine-learning, patient mortality |
| Type of data                           | Tables, spreadsheets                               |
| How data was acquired                  | Stanford University Medical Center Clinical Data Warehouse (Epic) electronic health record |
| Data format                            | Analyzed data                                      |
| Experimental factors                   | Clinical orders with prevalence < 1% among 2010–2013 patient hospitalizations were excluded from analysis. |
| Experimental features                  | Clinician cohorts were stratified based on observed vs. expected patient mortality outcomes. Association models were trained based on clinical order data [2] generated by low-mortality clinicians, high-mortality clinicians, and an unfiltered clinician crowd. |
| Data source location                   | Stanford University Medical Center, Stanford, CA, USA |
| Data accessibility                     | Tables are within this article; spreadsheets are attached as supplementary material. |
| Related research article               | Wang JK, Hom J, Balasubramanian S, et al. An Evaluation of Clinical Order Patterns Machine-Learned From Clinician Cohorts Stratified by Patient Mortality Outcomes. Journal of Biomedical Informatics. 2018;86:109-119. |

Value of the data

- Association model data can enable investigation of medical decision making patterns associated with low-mortality and high-mortality clinicians.
- Association model data can streamline the manual order set curation process by providing committees with real-time information regarding which orders are most commonly associated with which admission diagnoses.
- Association model data can be used to train or prototype data-driven clinical decision support tools that can ultimately provide clinicians with point-of-care guidance [3–5].
- Reference order lists curated from published practice guideline literature can serve as reference standards or benchmarks for evaluating automatically-learned clinical order patterns.

1. Data

Here we stratify clinicians in a tertiary academic hospital into "low-mortality" and "high-mortality" subgroups based on observed vs. expected patient mortality rates. We then train three distinct association models using clinical order data generated by the "low-mortality" and “high-mortality” clinician populations as well as an unfiltered crowd of all clinicians. We provide association data (Appendix A–F, see Table 1 for overview of association data spreadsheets) learned from each clinician cohort for six common admission diagnoses: Altered mental status (ICD9: 780.97), chest pain (ICD9: 786.5), gastrointestinal (GI) hemorrhage (ICD9: 578), heart failure (ICD9: 428), pneumonia (ICD9: 486), and syncope and collapse (ICD9: 780.2).

In this data article, we also share practice guideline-based reference standards that can be used to evaluate the "correctness" of automatically learned clinical order patterns (Appendix G, see Table 2 for overview of reference standard spreadsheet). These lists of reference orders were manually curated by physicians reviewing clinical practice literature for each admission diagnosis.
2. Experimental design, materials, and methods

2.1. Data source and preparation

We extracted deidentified, structured patient data from the (Epic) EHR for inpatient hospitalizations from 2008–2013 via the Stanford University Medical Center (SUMC) Clinical Data Warehouse [6]. Patient data was pre-processed to reduce complexity [8] across medication [7], lab result, and diagnosis coding. A complete description of data preparation can be found in Section 3.1 and 3.2 of Wang et al. [1].

2.2. Clinician stratification and patient cohort assembly

Clinicians who saw patients between 2010–2013 (n = 1,822) were stratified into low-mortality (21.8%, n = 397) and high-mortality (6.0%, n = 110) extremes using a two-sided P-value score quantifying deviation of observed vs. expected 30-day patient mortality rates. Expected per-patient mortality probabilities were predicted for patients seen in 2010–2013 based on 2008–2009 patient and mortality data (see Section 3.3 of Wang et al. [1] for full-length discussion of clinician stratification methodology). Defining physician–patient attribution using History and Physical Examination notes signed upon admission, three patient cohorts were assembled: Patients seen by low-mortality clinicians, high-mortality clinicians, and an unfiltered crowd of all clinicians. After balancing covariates between patient populations using common-referent 1:1:K propensity score matching [10], we obtained cohorts of size 1,046, 1,046, and 5,230 patients, respectively (see reference [1] for pre- and post-matching covariate distributions).

Table 1
Association model output overview.

| Feature              | Description                                                                 |
|----------------------|-----------------------------------------------------------------------------|
| clinical_item_id     | Unique clinical order identifier                                             |
| description          | Clinical order name or description                                           |
| score                | Negative log of the P-value computed by Yates’ chi-squared statistic       |
| PPV                  | Positive predictive value: nAB/nA                                           |
| PPV_95CI_low/high    | Positive predictive value 95% confidence intervals                          |
| OR                   | Odds ratio: (nAB/nB)/[(nA-nAB)/(N-nB)]                                      |
| OR_95CI_low/high     | Odds ratio 95% confidence intervals                                         |
| prevalence           | Prevalence: nB/N                                                            |
| prevalence_95CI_low/high | Prevalence 95% confidence interval                                        |
| RR                   | Relative risk: (nAB/nA)/[(nB – nAB)/(N – nA)]                              |
| RR_95CI_low/high     | Relative risk 95% confidence intervals                                      |
| P_YatesChi2          | P-value computed by Yates’ chi-squared statistic                            |
| N                    | Number of times any clinical order co-occurred within 24 h of the given admission diagnosis order |
| nAB                  | Number of times the specified clinical order co-occurred within 24 h of the given admission diagnosis order |
| nA                   | Number of times the admission diagnosis order occurred in general           |
| nB                   | Number of times the specified clinical order occurred in general            |

Table 2
Guideline reference standard overview.

| Feature              | Description                                                                 |
|----------------------|-----------------------------------------------------------------------------|
| icd9                 | Admission diagnosis classification code                                     |
| admission_diagnosis  | Admission diagnosis description                                              |
| clinical_item_id     | Unique clinical order identifier                                             |
| category             | Type of order (e.g. lab test, procedure, medication, etc.)                  |
| description          | Clinical order name or description                                           |
2.3. Association rule episode mining

We trained three distinct association models using patient encounters from the balanced low-mortality, high-mortality, and crowd patient cohorts, each reflecting clinical order patterns from the corresponding clinician population. We then generated order lists (Appendix A–F) from each association model for the six aforementioned admission diagnoses. Further discussion of association model (“clinical recommender engine”) training can be found in Section 3.6 of Wang et al. [1] and additional reading [9,11–14]. To assess similarity among predicted order lists, we can calculate agreement by Rank Biased Overlap [15], which accounts for rank-order (Table 3).

2.4. Practice guideline reference standard

We can evaluate each predicted order list against clinical practice guidelines as a proxy for “good” medical decision making (see Results of reference [1]). Two board-certified internal medicine physicians curated reference lists of clinical orders based on published clinical practice literature sourced from the National Guideline Clearinghouse (www.guideline.gov) and PubMed. After independently curating their lists, the two physicians resolved disagreements (items included in one physician’s list but not the other) by consensus to produce a final reference standard for each admission diagnosis. In this data article, we make available reference standards for the six aforementioned admission diagnoses (Appendix G). To assess pre-consensus agreement between the two clinicians, we computed Cohen’s Kappa statistics (Table 4).

Table 3
Rank Biased Overlap (RBO) computed between each pair of predicted order lists, score-ranked by PPV for the six example admission diagnoses. RBO computes the average fraction of top items in common between two order lists, geometrically weighting all ~20K candidate clinical order items, and ranges from 0.0 (no correlation or random list order) to 1.0 (perfect agreement). RBO is characterized by a “persistence” parameter p, the probability that an observer reviewing the top k items will continue to observe the (k+1)-th items. For our calculations, we used a default implementation parameter p of 0.98. This has the effect of geometrically weighting emphasis to the top of each list. RBO values of ~0.7 indicate strong overlap between order lists generated by two cohorts.

| Diagnosis                          | Low-mortality vs. high-mortality | Low-mortality vs. crowd | High-mortality vs. crowd |
|-----------------------------------|---------------------------------|-------------------------|-------------------------|
| Altered mental status (780.97)    | 0.64                            | 0.79                    | 0.64                    |
| Chest pain (786.5)                | 0.64                            | 0.77                    | 0.70                    |
| Gastrointestinal hemorrhage (578) | 0.65                            | 0.74                    | 0.67                    |
| Heart failure (428)               | 0.58                            | 0.67                    | 0.55                    |
| Pneumonia (486)                   | 0.66                            | 0.71                    | 0.67                    |
| Syncope and collapse (780.2)      | 0.61                            | 0.68                    | 0.63                    |

Table 4
Cohen’s Kappa values to assess pre-consensus agreement between reference standards independently curated by two board-certified clinicians from clinical practice guidelines. Values range from −1 to +1, with values < 0 indicating poor agreement and values > 0.6 indicating substantial agreement [16].

| Diagnosis                          | Pre-consensus Cohen's Kappa statistic |
|-----------------------------------|---------------------------------------|
| Altered Mental Status (780.97)    | 0.82                                  |
| Chest Pain (786.5)                | 0.66                                  |
| Gastrointestinal Hemorrhage (578) | 0.64                                  |
| Heart Failure (428)               | 0.75                                  |
| Pneumonia (486)                   | 0.72                                  |
| Syncope and Collapse (780.2)      | 0.72                                  |
Acknowledgements

This research was supported by the NIH Big Data 2 Knowledge initiative via the National Institute of Environmental Health Sciences under Award Number K01ES026837. Patient data were extracted and de-identified by Stanford Medicine’s Research IT department as part of the Stanford Medicine Research Data Repository (StaRR) project. The data repository project described was supported by the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health through grant UL1 RR025744. A Stanford Undergraduate Advising and Research Travel Grant supported JKW in presenting a portion of this work at the 2017 American Medical Informatics Association Annual Symposium. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or Stanford Healthcare.

Transparency document. Supporting information

Transparency data associated with this article can be found in the online version at https://doi.org/10.1016/j.dib.2018.10.163.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.dib.2018.10.163.

References

[1] J.K. Wang, J. Hom, S. Balasubramanian, et al., An evaluation of clinical order patterns machine-learned from clinician cohorts stratified by patient mortality outcomes, J. Biomed. Inform. 86 (2018) 109–119.
[2] R. Kaushal, K.G. Shojania, D.W. Bates, Effects of computerized physician order entry and clinical decision support systems on medication safety: a systematic review, Arch. Intern Med 163 (2003) 1409–1416.
[3] D.F. Sittig, A. Wright, J.A. Osheroff, et al., Grand challenges in clinical decision support, J Biomed. Inform. 41 (2008) 387–392.
[4] C.A. Longhurst, R.A. Harrington, N.H.A. Shah, ‘Green Button’ for using aggregate patient data at the point of care, Health Aff. 33 (2014) 1229–1235.
[5] J.K. Wang, A. Schuler, N.H. Shah, et al., Inpatient clinical order patterns machine-learned from teaching versus attending-only medical services, AMIA Jt. Summits Transl. Sci. Proc. 2017 (2018) 226–235.
[6] Stanford Medicine Research IT. Stanford Medicine Research Data Repository. 2017.(Available at) 〈https://med.stanford.edu/researchit.html〉.
[7] P. Hernandez, T. Podchiyska, S. Weber, T. Ferris, H. Lowe, Automated mapping of pharmacy orders from two electronic health record systems to RxNorm within the STRIDE clinical data warehouse, AMIA Annu Symp. Proc. 2009 (2009) 244–248.
[8] A. Wright, D.W. Bates, Distribution of problems, medications and lab results in electronic health records: the pareto principle at work, Appl. Clin. Inform. 1 (2010) 32–37.
[9] J.H. Chen, A. Muthuraman, M.K. Goldstein, et al., Decaying relevance of clinical data towards future decisions in data-driven inpatient clinical order sets, Int. J. Med. Inform. 102 (2017) 71–79.
[10] E. Sugawara, H. Nikaido, Properties of AdeABC and AdeIJK efflux systems of Acinetobacter Baumannii compared with those of the AcrAB-ToLC system of Escherichia coli, Antimicrob. Agents Chemother. 58 (2014) 7250–7257.
[11] J.H. Chen, R.B. Altman, Automated physician order recommendations and outcome predictions by data-mining electronic medical records, Proc. AMIA Summit Transl. Sci. 2014 (2014) 206–210.
[12] J. Chen, M. Goldstein, S. Asch, et al. Dynamically evolving clinical practices and implications for predicting medical decisions. Pacific Symposium of Biocomputing, 2016.
[13] J.H. Chen, M.K. Goldstein, S.M. Asch, et al., Predicting inpatient clinical order patterns with probabilistic topic models vs conventional order sets, JAMA 24 (2017) 472–480.
[14] A.K. Manrai, G. Bhatia, J. Strymish, et al., Medicine’s uncomfortable relationship with math: calculating positive predictive value, JAMA Intern. Med. (2016) 991–993.
[15] W. Webber, A. Moffat, J. Zobel, A similarity measure for indefinite rankings, ACM Trans. Inf. Syst. (2010) 1–38.
[16] J.R. Landis, G.G. Koch, The measurement of observer agreement for categorical data, Biometrics (1977) 159–174.