Unsupervised Machine Learning of Topics Documented by Nurses about Hospitalized Patients Prior to a Rapid-Response Event

Zfania Tom Korach[1]   Kenrick D. Cato[2]   Sarah A. Collins[2,3]   Min Jeoung Kang[1]   Christopher Knaplund[2]
Patricia C. Dykes[1]   Liqin Wang[1]   Kumiko O. Schnock[1]   Jose P. Garcia[1]   Haomiao Jia[2]   Frank Chang[1]
Jessica M. Schwartz[2]   Li Zhou[1]

[1] Division of General Internal Medicine and Primary Care, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts, United States
[2] School of Nursing, Columbia University, New York, New York, United States
[3] Department of Biomedical Informatics, Columbia University, New York, New York, United States

Address for correspondence Zfania Tom Korach, MD, Division of General Internal Medicine and Primary Care, Department of Medicine, Brigham and Women's Hospital, 399 Revolution Dr., Somerville, Massachusetts 02145, United States (e-mail: zkorach@bwh.harvard.edu).

Abstract

Background In the hospital setting, it is crucial to identify patients at risk for deterioration before it fully develops, so providers can respond rapidly to reverse the deterioration. Rapid response (RR) activation criteria include a subjective component ("worried about the patient") that is often documented in nurses' notes and is hard to capture and quantify, hindering active screening for deteriorating patients.

Objectives We used unsupervised machine learning to automatically discover RR event risk/protective factors from unstructured nursing notes.

Methods In this retrospective cohort study, we obtained nursing notes of hospitalized, nonintensive care unit patients, documented from 2015 through 2018 from Partners HealthCare databases. We applied topic modeling to those notes to reveal topics (clusters of associated words) documented by nurses. Two nursing experts named each topic with a representative Systematized Nomenclature of Medicine–Clinical Terms (SNOMED CT) concept. We used the concepts along with vital signs and demographics in a time-dependent covariates extended Cox model to identify risk/protective factors for RR event risk.

Results From a total of 776,849 notes of 45,299 patients, we generated 95 stable topics, of which 80 were mapped to 72 distinct SNOMED CT concepts. Compared with a model containing only demographics and vital signs, the latent topics improved the model's predictive ability from a concordance index of 0.657 to 0.720. Thirty topics were found significantly associated with RR event risk at a 0.05 level, and 11 remained significant after Bonferroni correction of the significance level to 6.94E-04, including physical examination (hazard ratio [HR] = 1.07, 95% confidence interval [CI], 1.03–1.12), informing doctor (HR = 1.05, 95% CI, 1.03–1.08), and seizure precautions (HR = 1.08, 95% CI, 1.04–1.12).

Conclusion Unsupervised machine learning methods can automatically reveal interpretable and informative signals from free-text and may support early identification of patients at risk for RR events.

Background and Significance

Rapid response (RR) teams are charged with responding to nonintensive care unit patients at risk for rapid deterioration, with the goal of preventing further deterioration and changing the deterioration's course. RR systems usually include two components: (1) identification of a clinical deterioration while it *develops* (as opposed to cardiac arrest teams that respond after the actual deterioration has occurred), and (2) provision of effective and timely interventions, aimed at treating the deterioration. The identification of imminent clinical deterioration and prompt clinical intervention were demonstrated to reduce mortality.[1,2]

Typically, RR systems are reactive in the sense of responding to calls made by staff noticing concerning findings. Active (prospective) surveillance of triggers for patient deterioration has achieved mixed results so far.[3,4] To facilitate active detection, the Clinical Decision Support Communication for Risky Patient States (CONCERN) study investigates nurses' judgment that a patient's clinical state may be deteriorating, in both narrative and structured information in acute and critical care.[5] While RR triggers are mostly objective measurements (e.g., heart rate, blood pressure, and alertness), they typically also include a subjective component such as "Staff member is worried about the patient" or "Any patient you are seriously worried about."[6,7] Such subjective measures have been shown to capture cases that would be missed by the objective criteria.[8] From the perspective of missed cases, an independent review of 118 inpatient cardiac arrest cases in a public hospital found that in 35% of avoidable arrest cases, communication of the nurses' concern about the patient's deterioration to the physician was delayed.[9]

While clinically important, the subjective criterion encompasses a multitude of clinical entities, and the reporting clinician might even lack a clear culprit finding.[8] Such variability and implicitness may hinder the ability to actively survey patients for subtle or subjective signs of deterioration due to challenges to formalize the criteria. By nature, these signs are expressed only in unstructured data such as notes. However, free-text notes cannot be used as-is for statistical modeling and risk prediction. Rather, they need to be transformed to numerical values in a process called "feature engineering." Hand-crafting feature is a labor-intensive process (e.g., deciding what signs and symptoms to extract from the notes), especially for exploratory studies looking to elucidate new associations as in the case of nurses' concern for RR event. To overcome these challenges, we applied topic modeling, an unsupervised machine learning approach, to explore the content of the notes prior to RR events and discover potential associations between the topics mentioned in nurses' documentation and RR event incidence.

Topic modeling, by applying statistical machine learning approaches, allows the revelation of latent patterns in the text without requiring manual annotation. Thus, it reduces the needed subject matter expert (SME) effort and is more suitable for exploratory analysis, where the relevant factors in the notes are not yet known. It has been extensively used for automatic feature engineering in text analysis and classification tasks.[10,11]

Briefly, topic modeling views documents as bags of words (i.e., the word's position does not matter, only its occurrence in the document). It assumes that each document was generated by picking a set of topics for this document and then for each selected topic, picking a set of words according to their association with the topic. Thus, topics are essentially distribution of words, where for a given topic each word has an association strength ranging from 0 to 1, and the sum of the association strength across all words equals 1. By observing the actual distribution of words in representative documents, the process can be reverse-engineered to unravel the original topics (i.e., their word distribution). The weight of each topic in a document can be calculated by counting the occurrence of each word in the document and applying the word-in-topic distribution. Various algorithms can be used to learn the topics from text, among them latent Dirichlet allocation (LDA) is commonly used.[12] In contrast to term-based approaches, topic modeling provides insights about themes appearing in a document even if the exact phrases vary from case to case. The resulting topics can be interpreted by SMEs and assigned to real-world clinical entities, to increase the interpretability of other statistical models based on the topics. Topic modeling has been widely applied to clinical text for various analyses including risk prediction, disease trajectory detection, and phenotyping.[13–15]

In the present study, we adopted topic modeling and survival analysis to discover potential factors in the clinical notes associated with the risk of RR event. We hypothesize that the topics revealed by topic modeling from nurses' documentation will be significantly associated with RR event incidence. The advancements in the identification of RR events therefore could facilitate earlier intervention and mitigation of preventable harms.

Methods

A high-level description of the study architecture is outlined in ►Fig. 1.

Data Collection

Study Population

The study population comprised hospitalized patients who were admitted to hospitals affiliated to Partners HealthCare, a large hospital system in the United States' northeastern region, between 2015 and 2018.

Inclusion criterion: Hospitalized patients who were admitted to at least one of the study units for 24 hours and longer. The study units were defined as "*A clinical general medical or surgical acute care or critical care unit,*" excluding pediatric or neonatal units, hospice units, emergency department, oncology units, obstetrics/labor and delivery units, behavioral/psychiatry units, observational units, operating room, preoperative, postoperative/postanesthesia care unit, same day surgical units, and plastic surgery units.

Exclusion criteria: (1) Patients less than 18 years old at the beginning of the study period, (2) patients who received hospice or palliative care, and (3) patients who lacked a hospital encounter.
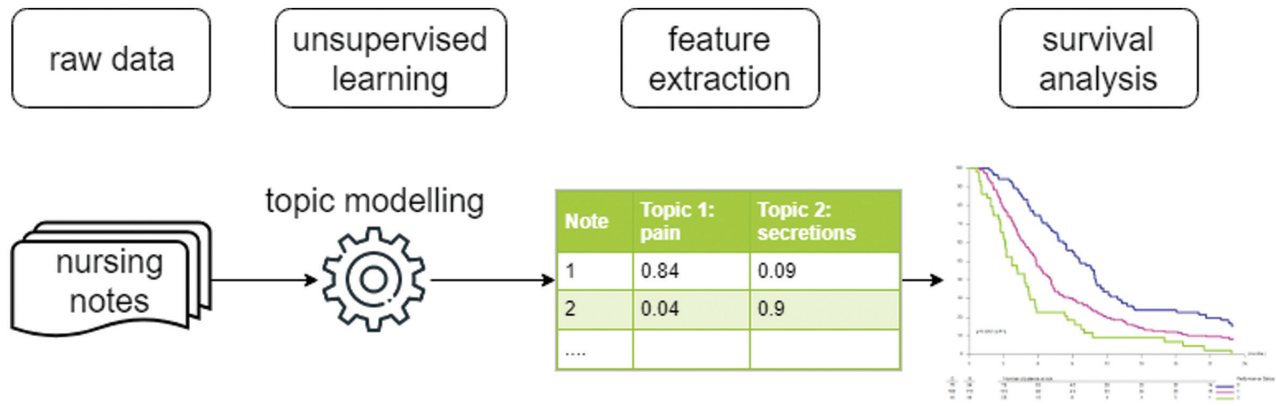
Fig. 1 A high-level description of the study architecture. The content of narrative texts is not numeric and cannot be used directly in statistical modeling (unlike, e.g., vital signs). Therefore, unsupervised machine learning was used to automatically learn numeric features that represent the content of nursing notes and correspond to known clinical concepts (from Systematized Nomenclature of Medicine–Clinical Terms [SNOMED CT]). These features were then used to develop a survival model of the clinical outcome, incidence of rapid response event, to discover clinical entities associated with the outcome.

Follow-Up Scope

Since patients may transfer between departments during their hospitalization, they were followed (for both exposure and outcome) only during their stay in study-included units and excluded when moved to a nonstudy unit. Transfers to radiology, procedures, operating room, and outpatient units were considered included or excluded based on the inclusion of the department from which they transferred. An example of an encounter's timeline and the corresponding inclusion status can be found in ►**Fig. 2**. Thus, each encounter included one or more contiguous time intervals during which the patient stayed only in included study units. For survival analysis, each interval is considered a separate case. Thus, the intervals of each encounter are correlated and will therefore require adjustment for the correlation. Since in practice over 99.8% of the encounters contained only a single interval, only the first interval of each encounter was included in the follow-up period, sparing the need for the more complex correlation adjustment (e.g., frailty analysis).

Data was collected only from the beginning of each follow-up period until either a RR event or right-censoring. To filter outliers, late RR events, defined as those occurring beyond the 99th percentile of the time from admission to RR event

among the CONCERN study population (1,282 hours), were excluded. Thus, right-censoring occurred at the earliest of the end of the follow-up period (time interval) and 1,282 hours since admission.

Note Collection

We included the following types of nursing notes in our study: *progress notes*, *consults*, *procedures*, *discharge summaries*, *assessment and plan note*, *nursing note*, *code documentation*, *significant event*, *transfer/sign off note*, *nursing summary*, and *family meeting*. Among those note types, we only obtained the notes documented by registered nurses. *RR documentation* notes were excluded from this analysis, despite being in the scope of the full CONCERN study, to prevent leakage of outcome information to the features.

Outcome Calculation

Incidence of RR events was captured from nursing flowsheets. Following the general scope of project, events occurring within 24 hours of admission were ignored, as well as events occurring past the 1,282 hours censoring time point. As multiple RR events can occur in a single encounter, only the first RR event was used.
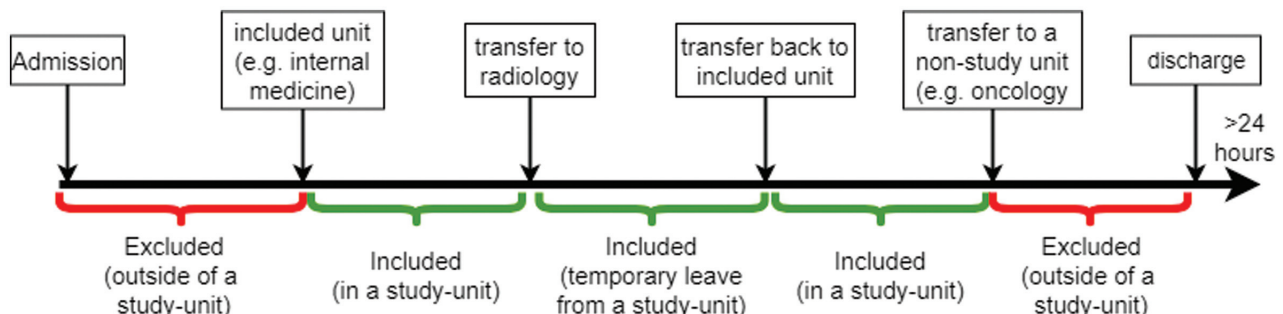

Fig. 2 Partitioning of an encounter to intervals and their inclusion status. A patient may be initially admitted to a nonstudy unit. The follow-up period begins once the patient is transferred to a study unit. Temporary transfers to radiology, procedures, operating room, and outpatient units do not change the inclusion/exclusion status. Once the patient is transferred to a nonstudy unit, follow-up ceases.

Topic Modeling

Note Preparation
The notes were tokenized and sentence segmented using the Stanford CoreNLP tokenizer.[16] Dates and numbers were collapsed to placeholder tokens. Headers and other text sequences automatically injected into the text by the electronic health record were removed, based on a manual review of the 1,000 most frequent n-grams of length 1 to 4 in the notes.

Topic Model Training
Topic models were trained on all of the notes using the Gensim implementation of the LDA algorithm.[17] The default hyperparameters were used except 50 passes and 5 iterations (►Table 1).

Topic Stabilization
Since LDA is a random process, the generated topics may change from invocation to invocation. We therefore reused a method developed by Shao et al to capture stable topics that remain similar between LDA runs.[15] The process is depicted in ►Fig. 3: for each target number of topics n, 3 topic models were trained using the same configuration (step 1), yielding 3 topic sets, each consisting of n LDA topics. For each of the $n \times n \times n$ LDA topic-triplets in the Cartesian product of the three sets, we calculated the similarity of the triplet's topics in terms of their pairwise cosine similarity (step 2). Following the original method, we retained only triplets whose average cosine similarity exceeded 0.7 (in a –1 to 1 scale), removing topics that varied between invocations and thus

are more likely to represent noise (step 3). Each retained triplet was consolidated to a stable topic by averaging its components. The density of a stable topic in a document was calculated as the average of its individual components' densities in that document.

Topic Number Selection
The number of topics is not learned by the algorithm. Rather, it is prespecified and is the most important configuration variable. Therefore, we searched for the optimal number of topics by training topic models to generate differing number of topics (50 to 250 in intervals of 50) and comparing the goodness-of-fit of a survival model based on the discovered topics, in line with the study's goal of identifying risk factors for RR event. Since the topics were learned using only the word co-occurrence information, without any information about the outcome, the full set of notes was used to fit the Cox model (without splitting the data to training and testing sets). The range of possible target numbers was set to 50 to 250 in intervals of 50. The lower limit was selected based on our previous work in which clinicians enumerated the clinical entities related to clinical deterioration amounting to 120 entities. The upper limit was selected based on two factors. First, the number of predictors in Cox model is practically limited by the number of observed events, requiring approximately 10 events per each predictor and thus limiting the number of predictors in the current study to approximately 100.[18,19] Second, the SMEs' capacity to review and name the topics placed additional limit on the target number of topics. The stable topics were applied as-is (without manual review)

Table 1 Gensim's default hyperparameters used for topic model training

| Hyperparameter	Description	Default value
chunksize	Number of documents to be used in each training chunk	2,000
Passes	Number of passes through the corpus during training	1
update_every	Number of documents to be iterated through for each update. Set to 0 for batch learning, > 1 for online iterative learning	1
α		"symmetric"
eta		None
decay	The percentage of the previous lambda value that is forgotten when each new document is examined	0.5
offset	How much we will slow down the first steps the first few iterations	1
eval_every	Number of updates before evaluating log perplexity	10
Iterations	Maximum number of iterations through the corpus when inferring the topic distribution of a corpus	50
gamma_threshold	Minimum change in the value of the gamma parameters to continue iterating	0.001
minimum_probability	Threshold to filter out topics with a probability lower than it	0.01
random_state		None
minimum_phi_value	Lower bound on the term probabilities	0.01
per_word_topics	If True, computes a list of topics, sorted in descending order of most likely topics for each word, along with their phi values multiplied by the feature length	False
dtype	Data type to use during calculations inside model	numpy.float32
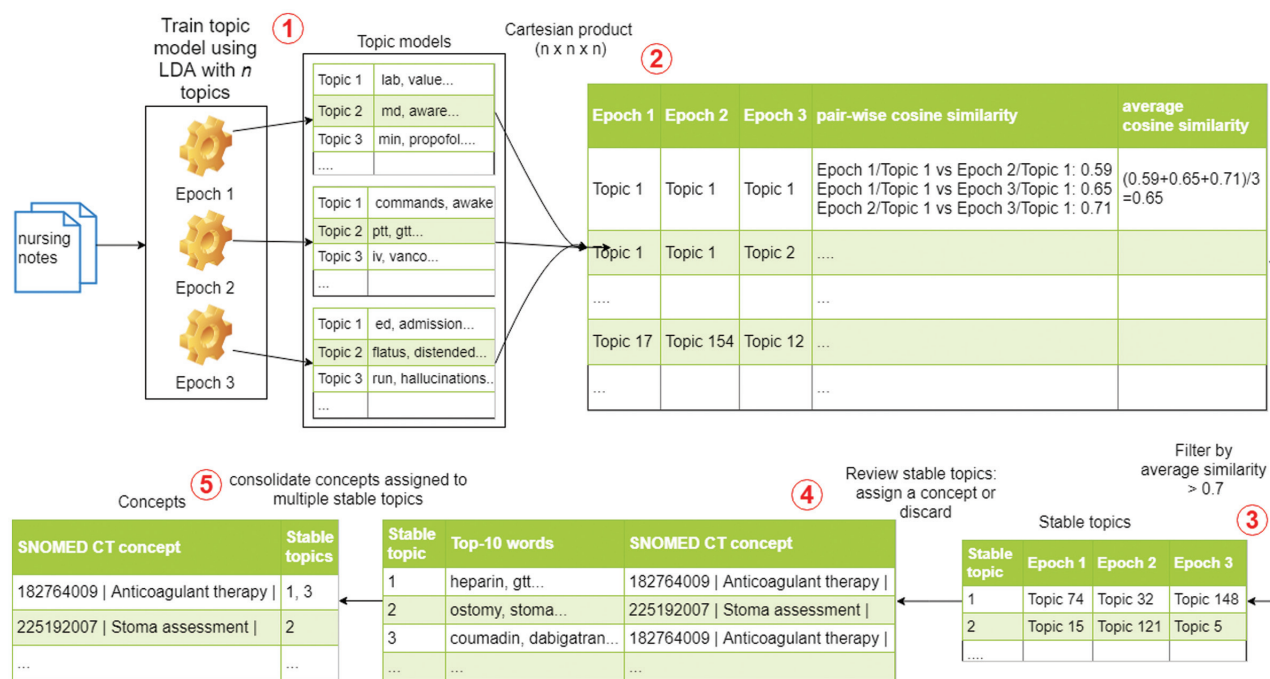
Fig. 3 Topic stabilization, naming, and consolidation process. See text for description of the numbered steps. First, three distinct topic models (epochs) are trained on the same corpus using the same target number of topics (1). Random components of the latent Dirichlet algorithm cause the generated topic to differ between the three epochs. The Cartesian product (taking all possible combination of the topics from the three epochs) is calculated, yielding n^3 triplets. The pairwise cosine similarity between the topics is calculated, yielding three values for each triplet (2). These values are averaged, and only those triplets which surpass the predefined cutoff are retained (3), to capture the *stable topics*—topics that remain similar between epochs. The stable topics are reviewed by two nursing experts and are either assigned a title from among Systematized Nomenclature of Medicine–Clinical Terms (SNOMED CT) concepts or discarded (4). The retained stable topics are consolidated (if multiple topics were assigned the same SNOMED CT concept) to generate the final list of concepts that will be used to represent the documents' content in the survival analysis.

to the full corpus. The weight of each stable topic in each note, along with the patient's age at admission, sex, and calendar hour of the note's entry were used as covariates for an extended Cox model. The different topic numbers were compared by the concordance index of their respective extended Cox models. Briefly, the concordance is defined as $P(x_i > x_j | y_i > y_j)$, the probability that the model's prediction goes in the same direction as the actual data. A pair of observations (i, j) is considered concordant if the prediction (x) and the outcome data (y) go in the same direction, that is, $(y_i > y_j$ and $x_i > x_j)$ and vice versa. The concordance index is the fraction of concordant pairs. Since in survival analysis higher risk translates to earlier event time, the definition of a concordant pair is flipped: a pair of observations is concordant if the observation with the higher risk (as estimated by the model) experiences the event *earlier* (has a shorter survival time), that is, $(y_i < y_j$ and $x_i > x_j)$. The concordance index is an extension of the area under the receiver-operating curve measure and it reflects the model's ability to rank (discriminate between) the observations according to their true risk/class. A concordance index value of 0.5 represents a model that is no better than a random guess and a value of 1 represents a model that can perfectly rank observations.

Topic Naming
Two nursing SMEs separately reviewed the words from each stable topic (see ►**Fig. 3**, step 4) and assigned a concept from

Systematized Nomenclature of Medicine–Clinical Terms (SNOMED CT) as the stable topic's name or discarded the stable topic if it was not clinically relevant (step 5). Each of these steps was performed based on the stable topic's top 10 words and top-10 weighted documents. Any disagreement between the two SMEs was resolved by consensus.

Calculation of Concept Weights per Document
The weight of each stable topic in each document was calculated by averaging the weight of its components (the LDA topics from the three topic models). The weight of a concept in a document was calculated using the value of the stable topic to which it was assigned. When multiple stable topics were assigned to the same concept, the weights of the stable topics were aggregated by summation. Thus, throughout the process multiple LDA topics are aggregated to a single stable topic (topic stabilization step), and multiple stable topics are aggregated to a single SNOMED CT concept (topic naming step). Eventually, each document was represented by the weight of each of the SNOMED CT concepts.

Survival Analysis
Extended Cox model with time-dependent covariates was used to estimate the association between each concept and the hazard of RR event.[20] Correlation between the concepts was estimated using variation inflation factor (VIF), to guide their inclusion together in a single model versus separate

model for each concept. Since vital signs are part of existing RR activation criteria, they were incorporated into the model as well. Heart rate, blood pressure (separating regular vs. arterial and systolic vs. diastolic to different variables), respiratory rate, oxygen saturation, and temperature values were collected from flowsheets. The most recent measurement within the 8 hours preceding the note was used. The topic weights and vital sign values were normalized to mean 0 and variance 1. Potential confounders were added to the model including both time-dependent (calendar hour of the note's entry) and time-independent (age at admission and sex) ones.

To investigate the effect of topic weights, three Cox models were built: vital signs alone, concept weights alone, and concept weights plus vital signs. The models were compared using their concordance index. For the concepts, a 0.05 significance level was prespecified and Bonferroni correction (dividing the prespecified significance level by the number of comparisons) was used to account for the multiple comparisons.

Results

The study cohort included 45,299 patients (23,110 women [51.0%] and 22,152 men [48.9%]; mean [standard deviation, SD] age, 62.1 [17.4] years) with a total of 61,740 hospital encounters and 1,067 RR events. RR events occurred at a median of 82 hours after the admission. The follow-up time (i.e., the time from entering a study unit to the earliest of exit from study unit, discharge, or RR event) averaged 140 (SD: 164) hours with a median of 82 and interquartile range of 44 to 167 hours. ►Table 1 lists the types and the numbers of notes included in this study.

Topic Model Search and Naming

Out of the five topic models built, the $n = 250$ topics model achieved the highest concordance index of 0.714. The concordance index showed a monotonically increasing relationship with the number of topics (Spearman's r: 0.900, p-value: 0.037).

The selected model ($n = 250$) yielded 95 stable topics (out of 15,625,000 possible triplets). The SMEs assigned 80 of them to 72 distinct concepts, discarding 15 (15.7%) of them. The stable topics and their assigned concepts are listed in ►Table 2.

Table 2 The number of notes by note type

Note type	Count
Progress notes	727,705
Nursing summary	36,913
Nursing note	5,411
Procedures	3,099
Significant event	1,739
Transfer/Sign off note	1,672
Code documentation	272
Family meeting	38
Total	776,849

Survival Analysis

VIF was low for all concepts (1.00–1.39) indicating the absence of substantial correlation between them. The models using vital signs alone and concept weights alone achieved a concordance index of 0.657 and 0.694, respectively, while the model combining both vital signs and concept weights achieved a higher concordance index of 0.720. The covariates whose Schoenfeld residuals were significantly associated with time are listed in ►Table 3. The hazard ratio (HR), 95% confidence interval (95% CI), and statistical significance at the two levels (raw and corrected) for all covariates in the final model are presented in ►Fig. 4. Using the raw significance level (0.05), 30 concepts were found statistically significant, dropping to 11 after the application of Bonferroni correction. The significantly hazard-increasing covariates represent various themes including patient factors (age, HR = 1.492, 95% CI = 1.385–1.607), vital signs (respiratory rate, HR = 1.086, 95% CI = 1.068–1.105), and clinical attention (physical examination, HR = 1.074, 95% CI = 1.033–1.117; informing doctor, HR = 1.054, 95% CI = 1.028–1.081; and assessment of eating and drinking behavior, HR = 1.053, 95% CI = 1.025–1.082). The most protective covariates include concepts representing clinical improvement (weaning from mechanically assisted ventilation, HR = 0.838, 95% CI = 0.766–0.917 and ambulating patient, HR = 0.594, 95% CI = 0.486–0.726).

Discussion

The topics revealed by the unsupervised method corresponded well (80 out of 95 topics retained) to real-world clinical concepts. Moreover, these topics enhanced the ability to predict RR event hazard over vital signs alone. Some of the significantly associated concepts match the expected results, for example, the protective effect of "ambulating patient," while others are novel findings and may guide further investigation, including validation of the concept's manifestation in the notes and validation studies. Notably, several concepts describing increased medical attention, including "close observation," "physical examination," "informing doctor," and "assessment of eating and drinking behavior," were found to increase the hazard of RR event. These significant associations may represent early and subtle cues about impending clinical deterioration expressed by the nursing staff before the patient's condition crosses the threshold for RR team activation. From a clinical perspective our results in ►Fig. 4 are encouraging. For example, several concepts that increased risk for RR correspond with traditional criteria for initiating a RR (i.e., respiratory rate, general health deterioration, cardiac arrhythmia). Also, several concepts that indicated decreased risk of RR, align with preventative clinical actions (i.e., prevention of deep vein thrombosis, neurological mental status, diuretic therapy). To further develop these signals, we propose future work which would validate the discovered risk/protective factors by manual chart review and testing on data from other organizations, as well as automated term recognition methods to capture complex clinical entities from the text. Nonsignificant associations between plausible concepts and RR event hazard could result from various reasons.

Table 3 The top-10 words and the concept assigned to each stable topic

Concept	SNOMED CT ID	Top-10 words
Abdominal pain	21522001	perforated, pain, diverticulitis, appendectomy, iv, appendicitis, rlq, qtc, sotalol, cipro
Admission assessment	406152008	admission, nursing, note, arrived, arrival, stretcher, ed, floor, pain, pacu
Advanced directive status	310301000	living, hill, wingate, golden, chestnut, wks, str, salem, weston, assisted
Alcohol dependence	66590003	ativan, etoh, withdrawal, ciwa, anxiety, prn, po, tremors, shift, mg
Ambulating patient	62013009	pain, steady, gait, flatus, oob, bs, voiding, abd, soft, bm
Ambulating patient	62013009	ad, lib, pain, up, oob, denies, well, room, steady, vss
Antibiotic therapy	281789004	vanco, iv, trough, dose, po, picc, shift, id, due, vanc
Anticoagulant therapy	182764009	heparin, ptt, gtt, units, kg, time_placeholder, next, due, therapeutic, coumadin
Arteriovenous fistula	439470001	hd, dialysis, fistula, esrd, avf, arm, removed, renal, mwf, bruit
Assessment of eating and drinking behavior	710848001	crackers, toast, juice, rechecked, pain, st, aware, md, degree, orange
Assessment of pain control	370778008	pain, well, controlled, diet, monitor, tolerating, managed, voiding, vss, oob
At risk for aspiration	371736008	aspiration, pills, liquids, crushed, diet, applesauce, slp, dysphagia, meds, whole
Backache	161891005	pain, lumbar, spine, back, spinal, fusion, cervical, mg, posterior, date_placeholder
Bladder retention of urine	130951007	cath, bladder, straight, void, cc, urinary, scan, retention, scanned, time_placeholder
Blood pressure alteration	129899009	line, pa, milrinone, goal, stable, failure, tbb, mg, remains, bleeding
Cardiac arrhythmia	698247007	run, beat, nsvt, runs, asymptomatic, shift, am, completed, mg, started
Case management	386230005	management, case, completed, screening, initial, follow, assessment, high, available, risk
Close observation	225415001	psych, time_placeholder, sitter, shift, times, bed, leave, agitated, section, safety
Cutaneous hypersensitivity	21626009	rash, benadryl, itching, sarna, lotion, iv, pain, noted, itchiness, prn
Diabetic care management	385806006	fs, hs, ac, insulin, lispro, scale, lantus, sliding, coverage, ss
Discharge planning	371754007	discharge, facility, care, ambulance, paperwork, time, snf, must, transfer, via
Discharge planning	371754007	rehab, bed, cm, ready, spaulding, follow, team, discharge, facility, snf
Discharge planning	371754007	discharge, understanding, reviewed, instructions, home, verbalized, all, questions, up, written
Discomfort	247347003	complaint, fib, pain, chief, roll, tach, iv, tib, time, lsctab
Disease of liver	235856003	lactulose, shift, liver, cirrhosis, following, hepatic, encephalopathy, bms, po, sbp
Diuretic therapy	722048006	mg, lasix, iv, ivp, po, am, afib, sob, started, dose
Emotional support	133921002	support, emotional, provided, pain, monitor, safety, maintained, see, care, bedside
Evaluation of response to administration of fluids and electrolytes	372068006	pac, sodium, ph, blood, iv, shift, bicarb, urine, accessed, cal
Evaluation of tubes and drains	711139001	drain, ir, pain, drainage, iv, flushed, drains, cc, fluid, abdominal
Evaluation of tubes and drains	711139001	chest, ct, pain, tube, leak, suction, right, noted, left, site
Examination of limb	302773001	groin, pulses, left, hematoma, artery, repair, pain, site, sbp, bilateral
Falls education	390997009	call, bed, bell, reach, alarm, light, denies, within, safety, pain
Finding of pattern of pain	301369003	anti, pain, spasms, pigtail, β, xa, spasm, cxray, shift, am
Fluid balance regulation	276026009	ml, ns, calcium, volume, fluid, bolus, exchange, time, albumin, total
Fracture care	385691007	fall, fracture, fx, left, pain, mg, right, orif, tylenol, hip
General health deterioration	285384003	change, reassess, team, ccrn, prn, subject, lead, aware, remains, medical
Handoff communication	432138007	summary, illness, action, verbal, confirm, awareness, concerns, handoff, severity, issues
Hemodialysis observable	4.81E+11	date_placeholder, hd, laboratory, date, results, component, value, post, negative, weight
High risk of bleeding	711536002	hct, gi, egd, bleeding, cbc, prbc, stool, unit, shift, hgb
History taking	84100007	history, YEAR, htn, who, chronic, presents, disease, past, pmh, year
Home health aide service assessment	385780008	independent, subscriber, assistance, address, home, services, name, functional, prior, primary
Indigestion	162031009	maalox, pain, indigestion, stomach, reflux, heartburn, iron, simethicone, omeprazole, md
Infection control procedure	77248004	doxycycline, washed, pain, tick, iv, bite, lyme, lvp, analgesia, cat
Informing doctor	304562007	responding, clinician, bipap, paged, iv, md, access, rn, aware, notified

Table 3 (*Continued*)

Concept	SNOMED CT ID	Top-10 words
Insertion of catheter into blood vessel	429446009	cm, date_placeholder, picc, time_placeholder, procedure, lumen, dressing, insertion, placement, catheter
Language barrier	422693009	speaking, spanish, english, able, needs, only, ipop, primarily, make, known
Left ventricular assist device present	723438005	lvad, vad, flows, stable, alarms, date_placeholder, vt, dressing, changed, mg
Legal guardian	58626002	guardian, guardianship, court, wheelchair, hearing, eye, blind, bound, ck, baseline
Measuring output from thoracic drain	72162008	effusion, pleural, chest, pericardial, drained, cxr, pigtail, pain, ct, thoracentesis
Monitoring pain	710995003	cont, monitor, pain, conts, iv, vss, po, amb, denies, effect
Nasogastric tube maintenance	52260009	cc, ngt, npo, output, pain, iv, lws, brown, draining, abd
Nausea care management	408882007	nausea, zofran, pain, effect, iv, vomiting, good, mg, po, emesis
Neurological assessment	225398001	neuro, speech, left, facial, weakness, commands, right, strengths, perrl, follows
Neurological mental status determination	392257007	status, mental, hospitalization, condition, adult, during, altered, infection, progressing, respiratory
Nursing care coordination	385777007	home, discharge, cm, vna, services, care, met, referral, follow, team
Nursing evaluation of patient and report	19681004	did, him, about, stated, would, does, when, states, said, asked
Observational assessment	310813001	pain, mg, iv, md, anxious, portacath, humalog, wbcs, shift, aware
Oxygen therapy	57485005	oxygen, liters, nc, up, sats, pick, sat, nasal, air, breath
Pacemaker care assessment	410096008	ppm, ep, pacemaker, pacer, placement, site, degree, block, stable, device
Pain control	225782006	hip, mg, knee, pain, csm, tylenol, oxycodone, prn, pp, ice
Pain control	225782006	pain, mg, back, patch, po, prn, dilaudid, oxycodone, tylenol, colace
Pain control	225782006	pca, pain, dilaudid, limit, mg, iv, outlined, lockout, dose, use
Pain control	225782006	pain, spasms, spec, tylenol, valium, mg, motrin, back, ivig, iv
Palliative care	103735009	family, care, morphine, hospice, palliative, comfort, cmo, comfortable, meeting, bedside
Patient discharge	58000006	dc, private, np, mgh, am, epic, home, ip, dr, thurs
Physical examination	5880005	sounds, denies, clear, pain, soft, lung, noted, abdomen, sob, bs
Physical examination	5880005	spo, temp, wt, oral, lb, nc, bp, shift, pulse, pain
Pressure ulcer care	225357008	skin, applied, mepilex, coccyx, area, cream, barrier, buttocks, noted, red
Prevention of deep vein thrombosis	439993001	pfo, filter, ivs, ivc, all, dvt, pe, iv, shift, procedures
Procedure aiding diagnosis	165167006	scan, pet, biopsy, mass, ct, pain, npo, onc, bx, throat
Procedure on cardiovascular system	118672003	completed, shift, cad, cabg, ccl, cath, transferred, ef, osh, htn
Respiratory assessment	422834003	cough, productive, nebs, sats, sputum, ra, nc, ls, sob, prn
Secondary malignant neoplastic disease	128462008	radiation, oncology, pain, ca, chemo, metastatic, mets, xrt, potential, cancer
Seizure precautions	64461008	seizure, activity, noted, eeg, neuro, keppra, shift, precautions, seizures, event
Smoking cessation therapy	710081004	smoker, smoking, patch, nicotine, former, quit, pain, cessation, security, floor
Stoma assessment	225192007	ostomy, stoma, ileostomy, date_placeholder, intact, wound, pouch, colostomy, output, appliance
Tracheostomy care	385858000	trach, secretions, tf, tube, via, peg, place, shift, thick, suctioning
Urinary catheter	20568009	urine, foley, clots, pain, stent, urology, pink, colored, hematuria, catheter
Weaning from mechanically assisted ventilation	243174005	remains, off, mcg, propofol, weaned, map, goal, min, gtt, levo
Wound care	225358003	wound, vac, dressing, changed, drainage, suction, foot, mmhg, change, intact

Abbreviation: SNOMED CT: Systematized Nomenclature of Medicine–Clinical Terms.
Note: Multiple stable topics could be assigned to the same concept. Such stable topics are presented sequentially.

Particularly, negation, experiencer, and other context modifiers of a mention (e.g., "*no* pain or swelling") divide the various mentions of a word to different meanings. However, such differences are not captured by bag-of-words methods like topic modeling which consider only the presence or absence of the word, highlighting these methods' limitation.

This study suffers from several limitations. First, while the addition of concepts to the survival model achieved a higher predictive ability, the enhanced model does not achieve perfect accuracy (concordance index of 0.720 in a scale of 0.5–1.0). During the target topic number search, the predictive ability increased monotonically with the number of topics (Spearman's *r*: 0.900, *p*-value: 0.037), suggesting a benefit from further increase of the target number of topics. As explained above, the target number of topics was limited in this work by the sample and the available human effort.

Ef.	Covariate type	Covariate	log(hazard ratio) with 95% confidence interval	Hazard ratio (95% confidence interval)	P value	Corrected P value
↑	confounder	Age		1.49 (1.38 - 1.61)	<.001	<.001
↑	vital sign	Respiratory Rate		1.09 (1.07 - 1.1)	<.001	<.001
↑	concept	Oxygen therapy		1.08 (1.06 - 1.11)	<.001	<.001
↑	concept	Seizure precautions		1.08 (1.04 - 1.12)	<.001	<.001
↑	concept	Physical examination		1.07 (1.03 - 1.12)	<.001	<.001
↑	concept	Admission assessment		1.07 (1.02 - 1.12)	<.001	<.001
↑	concept	Arteriovenous fistula		1.07 (1.03 - 1.11)	<.001	<.001
↑	concept	History taking		1.06 (1.01 - 1.11)	0.013	0.934
↑	concept	High risk of bleeding		1.05 (1.01 - 1.1)	0.017	>.99
↑	concept	Informing doctor		1.05 (1.03 - 1.08)	<.001	<.001
↑	concept	Assessment of eating and drinking behavior		1.05 (1.02 - 1.08)	<.001	<.001
↑	concept	General health deterioration		1.04 (1.01 - 1.08)	<.001	<.001
↑	concept	Nursing evaluation of patient and report		1.04 (1 - 1.09)	0.078	>.99
↑	concept	Fluid balance regulation		1.03 (1 - 1.07)	0.041	>.99
↑	concept	Hemodialysis observable		1.03 (0.98 - 1.09)	0.21	>.99
↑	concept	Bladder retention of urine		1.03 (1 - 1.07)	0.047	>.99
↑	concept	Pressure ulcer care		1.03 (0.98 - 1.08)	0.27	>.99
↑	concept	Cardiac arrhythmia		1.03 (0.99 - 1.07)	0.14	>.99
↑	concept	Advanced directive status		1.03 (1 - 1.06)	0.042	>.99
↑	concept	Discomfort		1.03 (1 - 1.05)	0.059	>.99
↑	concept	Examination of limb		1.02 (0.96 - 1.08)	0.47	>.99
↑	concept	Monitoring pain		1.02 (0.97 - 1.07)	0.48	>.99
↑	concept	Observational assessment		1.02 (0.98 - 1.06)	0.38	>.99
↑	concept	Evaluation of response to administration of fluids and electrol		1.02 (0.97 - 1.07)	0.43	>.99
↑	concept	Handoff communication		1.01 (0.96 - 1.07)	0.62	>.99
↑	concept	Falls education		1.01 (0.96 - 1.06)	0.61	>.99
↑	concept	Procedure aiding diagnosis		1.01 (0.97 - 1.06)	0.64	>.99
↑	concept	Nasogastric tube maintenance		1.01 (0.96 - 1.06)	0.68	>.99
↑	concept	Indigestion		1.01 (0.98 - 1.05)	0.59	>.99
↑	concept	Nausea care management		1.01 (0.96 - 1.06)	0.82	>.99
=	concept	Emotional support		1 (0.95 - 1.06)	0.95	>.99
=	vital sign	Blood Pressure diastolic		1 (0.91 - 1.1)	0.97	>.99
=	time interaction	Neurological assessment		1 (1 - 1)	<.001	<.001
=	time interaction	Arterial Blood Pressure systolic		1 (1 - 1)	0.61	>.99
=	time interaction	Heart Rate		1 (1 - 1)	0.16	>.99
=	time interaction	Arterial Blood Pressure diastolic		1 (1 - 1)	0.52	>.99
=	time interaction	Left ventricular assist device present		1 (1 - 1)	0.13	>.99
=	time interaction	Close observation		1 (1 - 1)	0.13	>.99

Fig. 4 Extended Cox model results. The vertical bar in the plot represents 0, the value of no-effect (in logarithmic scale), corresponding to a hazard ratio of 1. Ef., effect on the hazard rate; "↓," decreasing hazard; "↑," increasing hazard; "=," no effect on hazard.

Ef.	Covariate type	Covariate	log(hazard ratio) with 95% confidence interval	Hazard ratio (95% confidence interval)	P value	Corrected P value
=	time interaction	Tracheostomy care		1 (1 - 1)	<.001	<.001
=	time interaction	Wound care		1 (1 - 1)	0.061	>.99
=	time interaction	Legal guardian		1 (1 - 1)	0.074	>.99
=	time interaction	Cutaneous hypersensitivity		1 (1 - 1)	0.026	>.99
=	concept	Backache		1 (0.94 - 1.06)	0.95	>.99
=	concept	Blood pressure alteration		1 (0.95 - 1.05)	0.9	>.99
=	concept	Finding of pattern of pain		1 (0.93 - 1.07)	0.91	>.99
↓	vital sign	O2 Saturation		0.99 (0.93 - 1.06)	0.86	>.99
↓	confounder	Calendar hour		0.99 (0.93 - 1.05)	0.75	>.99
↓	concept	Secondary malignant neoplastic disease		0.99 (0.92 - 1.07)	0.76	>.99
↓	concept	Prevention of deep vein thrombosis		0.99 (0.92 - 1.06)	0.71	>.99
↓	concept	Measuring output from thoracic drain		0.98 (0.93 - 1.04)	0.54	>.99
↓	concept	Left ventricular assist device present		0.98 (0.89 - 1.08)	0.72	>.99
↓	concept	Neurological mental status determination		0.98 (0.94 - 1.02)	0.37	>.99
↓	concept	Diuretic therapy		0.97 (0.93 - 1.02)	0.31	>.99
↓	vital sign	Temperature		0.97 (0.92 - 1.04)	0.4	>.99
↓	concept	At risk for aspiration		0.97 (0.92 - 1.03)	0.37	>.99
↓	concept	Respiratory assessment		0.97 (0.93 - 1.02)	0.26	>.99
↓	concept	Language barrier		0.97 (0.91 - 1.03)	0.31	>.99
↓	concept	Abdominal pain		0.97 (0.91 - 1.03)	0.29	>.99
↓	concept	Anticoagulant therapy		0.97 (0.91 - 1.02)	0.22	>.99
↓	concept	Home health aide service assessment		0.96 (0.89 - 1.03)	0.28	>.99
↓	concept	Urinary catheter		0.96 (0.9 - 1.02)	0.18	>.99
↓	concept	Insertion of catheter into blood vessel		0.96 (0.89 - 1.04)	0.31	>.99
↓	concept	Fracture care		0.95 (0.89 - 1.03)	0.21	>.99
↓	concept	Evaluation of tubes and drains		0.95 (0.88 - 1.02)	0.15	>.99
↓	vital sign	Blood Pressure systolic		0.95 (0.86 - 1.04)	0.27	>.99
↓	confounder	Female sex		0.95 (0.83 - 1.08)	0.4	>.99
↓	concept	Pacemaker care assessment		0.94 (0.87 - 1.02)	0.15	>.99
↓	concept	Palliative care		0.94 (0.89 - 1)	0.04	>.99
↓	concept	Disease of liver		0.94 (0.87 - 1.01)	0.097	>.99
↓	concept	Alcohol dependence		0.93 (0.88 - 1)	0.037	>.99
↓	concept	Case management		0.93 (0.87 - 1)	0.041	>.99
↓	concept	Discharge planning		0.92 (0.84 - 1.02)	0.12	>.99
↓	concept	Patient discharge		0.91 (0.82 - 1.01)	0.091	>.99
↓	concept	Nursing care coordination		0.91 (0.84 - 1)	0.043	>.99
↓	concept	Pain control		0.91 (0.83 - 1)	0.048	>.99
↓	concept	Infection control procedure		0.89 (0.81 - 0.98)	0.019	>.99
↓	concept	Assessment of pain control		0.88 (0.81 - 0.95)	<.01	<.001

Fig. 4 (Continued)
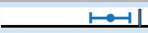
Ef.	Covariate type	Covariate	log(hazard ratio) with 95% confidence interval	Hazard ratio (95% confidence interval)	P value	Corrected P value
↓	concept	Antibiotic therapy		0.87 (0.79 - 0.95)	<.01	<.001
↓	concept	Diabetic care management		0.85 (0.77 - 0.95)	<.01	<.001
↓	concept	Smoking cessation therapy		0.85 (0.75 - 0.97)	0.017	>.99
↓	concept	Neurological assessment		0.85 (0.76 - 0.94)	<.01	<.001
↓	concept	Stoma assessment		0.84 (0.71 - 0.99)	0.042	>.99
↓	concept	Weaning from mechanically assisted ventilation		0.84 (0.77 - 0.92)	<.01	<.001
↓	concept	Procedure on cardiovascular system		0.8 (0.72 - 0.9)	<.01	<.001
↓	concept	Ambulating patient		0.59 (0.49 - 0.73)	<.01	<.001

Fig. 4 (*Continued*)

Table 4 Covariates violating the proportionality assumption from the vital signs-only and the concepts-only extended Cox models

Model	Covariate	Correlation coefficient	*p*-Value
Vital signs-only	Heart rate	−0.13694	3.09E-07
	Arterial blood pressure systolic	0.145736	0.000214
	Arterial blood pressure diastolic	−0.11664	0.010001
Concepts-only	Close observation	−0.06205	0.036023
	Cutaneous hypersensitivity	−0.05408	0.024085
	Legal guardian	−0.09688	0.043812
	Tracheostomy care	−0.15304	2.65E-05
	Wound care	−0.08908	0.014788

The current findings may spur collection of a larger sample and additional human effort to allow a higher target number of topics to enhance the resulting survival model's capacity.

Second, while the topics' manifestation in the notes was inspected in the top-10 notes for each topic, the accuracy by which the assigned concepts' weights describe the note's content was not formally evaluated. In addition to requiring intensive SME effort, such evaluation faces inherent challenges stemming for the continuous nature of topic weights compared with the dichotomous manner by which human annotators typically grade a topic occurrence. Finally, this study included data from a single organization, precluding estimation of the findings' generalizability to other organizations which might differ in documentation habits and patient population.

Conclusion

The present study demonstrates the ability of unsupervised machine learning to automatically extract interpretable and informative textual features from free-text without manual feature engineering, facilitating large-scale and exploratory studies of clinical outcomes from unstructured data.

Clinical Relevance Statement

Unsupervised machine learning can be used to discover nursing topics associated with a clinical outcome's risk with reduced manual effort.

Multiple Choice Questions

1. What manual effort is needed from the subject matter experts (SMEs) to use topic modeling?
   a. The SMEs have to define each topic and its root word, so the topic modeling algorithm can expand these roots to other words.
   b. SMEs have to define the number of top words in each topic, to guide the algorithm about the distribution of words in the topics.
   c. SMEs have to manually tag notes with annotations of topics, so the topic modeling algorithm can learn how topics look and find new ones.
   d. SMEs have to define only the number of topics. Optionally, they can review the topics to assign concepts to each of them.

   Correct Answer: The correct answer is option d. As an unsupervised machine learning method, latent Dirichlet allocation-based topic modeling generates the topics (word distribution) from observed corpus and does not require a root for each topic (a). Labeled examples are needed for supervised machine learning algorithms (c). The target number of topics needs to be defined (d) rather than the number of top words in each topic (b).

2. When learning a topic model, how is the target number of topics determined?
   a. The target number of topics is determined automatically by the latent Dirichlet allocation (LDA) algorithm, since it is an unsupervised machine learning algorithm.
   b. The target number is determined manually and requires careful selection.
   c. The largest target allowed by the computer hardware should be used.
   d. The number of the topics depends on the number of distinct words in the corpus.

Correct Answer: The correct answer is option b. The target number of topics is a configuration parameter and is not determined by either the topic modeling algorithm (a) or the vocabulary size (d). Rather, it is set manually and requires careful selection, either based on prior knowledge or based on a relevant benchmark (b). Even when the hardware is capable of handling a higher number of topics, other considerations (e.g., ratio of variables to number of examples) might dictate a lower number (c).

Protection of Human and Animal Subjects

The study was approved by the institutional review board of Partners HealthCare System.

Funding

Conflict of Interest

None declared.

References

1 Winters BD, Weaver SJ, Pfoh ER, Yang T, Pham JC, Dy SM. Rapid-response systems as a patient safety strategy: a systematic review. Ann Intern Med 2013;158(5 Pt 2):417–425

2 Solomon RS, Corwin GS, Barclay DC, Quddusi SF, Dannenberg MD. Effectiveness of rapid response teams on rates of in-hospital cardiopulmonary arrest and mortality: a systematic review and meta-analysis. J Hosp Med 2016;11(06):438–445

3 Huh JW, Lim C-M, Koh Y, et al. Activation of a medical emergency team using an electronic medical recording-based screening system*. Crit Care Med 2014;42(04):801–808

4 Kollef MH, Chen Y, Heard K, et al. A randomized trial of real-time automated clinical deterioration alerts sent to a rapid response team. J Hosp Med 2014;9(07):424–429

5 Collins SA, Cato K, Albers D, et al. Relationship between nursing documentation and patients' mortality. Am J Crit Care 2013;22 (04):306–313

6 Bellomo R, Goldsmith D, Uchino S, et al. Prospective controlled trial of effect of medical emergency team on postoperative morbidity and mortality rates. Crit Care Med 2004;32(04):916–921

7 Hillman K, Chen J, Cretikos M, et al; MERIT study investigators. Introduction of the medical emergency team (MET) system: a cluster-randomised controlled trial. Lancet 2005;365(9477):2091–2097

8 Parr MJ, Hadfield JH, Flabouris A, Bishop G, Hillman K. The Medical Emergency Team: 12 month analysis of reasons for activation, immediate outcome and not-for-resuscitation orders. Resuscitation 2001;50(01):39–44

9 Hodgetts TJ, Kenward G, Vlackonikolis I, et al. Incidence, location and reasons for avoidable in-hospital cardiac arrest in a district general hospital. Resuscitation 2002;54(02):115–123

10 Rubin TN, Chambers A, Smyth P, Steyvers M. Statistical topic models for multi-label document classification. Mach Learn 2012;88(01):157–208

11 Wang L, Sha L, Lakin JR, et al. Development and validation of a deep learning algorithm for mortality prediction in selecting patients with dementia for earlier palliative care interventions. JAMA Netw Open 2019;2(07):e196972

12 Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. J Mach Learn Res 2003;3(Jan):993–1022

13 Perotte A, Ranganath R, Hirsch JS, Blei D, Elhadad N. Risk prediction for chronic kidney disease progression using heterogeneous electronic health record data and time series analysis. J Am Med Inform Assoc 2015;22(04):872–880

14 Disease Trajectories and End-of-Life Care for Dementias: Latent Topic Modeling and Trend Analysis Using Clinical Notes. ResearchGate. Available at: https://www.researchgate.net/publication/328899169_Disease_Trajectories_and_End-of-Life_Care_for_Dementias_Latent_Topic_Modeling_and_Trend_Analysis_Using_Clinical_Notes. Accessed December 18, 2018

15 Shao Y, Mohanty AF, Ahmed A, et al. Identification and use of frailty indicators from text to examine associations with clinical outcomes among patients with heart failure. AMIA Annu Symp Proc 2017;2016:1110–1118

16 Manning CD, Surdeanu M, Bauer J, Finkel J, Bethard SJ, McClosky D. The Stanford CoreNLP Natural Language Processing Toolkit. In: Association for Computational Linguistics (ACL) System Demonstrations; 2014:55–60. Available at: http://www.aclweb.org/anthology/P/P14/P14-5010. Accessed November 25, 2019

17 gensim: topic modelling for humans. Available at: https://radimrehurek.com/gensim/. Accessed March 4, 2018

18 Concato J, Peduzzi P, Holford TR, Feinstein AR. Importance of events per independent variable in proportional hazards analysis. I. Background, goals, and general strategy. J Clin Epidemiol 1995;48(12):1495–1501

19 Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. J Clin Epidemiol 1995;48(12):1503–1510

20 Andersen PK. Repeated assessment of risk factors in survival analysis. Stat Methods Med Res 1992;1(03):297–315