Anaesth Crit Care Pain Med 43 (2024) 101390

Contents lists available at ScienceDirect

Anaesthesia Critical Care & Pain Medicine

journal homepage: www.elsevier.com

Original Article

Validation of a natural language processing algorithm using national reporting data to improve identification of anesthesia-related ADVerse evENTs: The "ADVENTURE" study

Paul M Mertes [a,e,1], Claire Morgand [b,1], Paul Barach [c,d], Geoffrey Jurkolow [e,*], Karen E. Assmann [b], Edouard Dufetelle [f], Vincent Susplugas [f], Bilal Alauddin [f], Patrick Georges Yavordios [e], Jean Tourres [e], Jean-Marc Dumeix [e], Xavier Capdevila [g,h]

[a] Department of Anesthesia and Intensive Care, Hôpitaux Universitaires de Strasbourg, Nouvel Hôpital Civil, EA 3072, FMTS de Strasbourg, Strasbourg, France
[b] Evaluation Department and Tools for Quality and Safety of Care, French national authority for health (Haute Autorité de Santé - EvOQSS), Saint Denis, France
[c] Thomas Jefferson School of Medicine, Philadelphia, USA
[d] Sigmund Freud University, Vienna, Austria
[e] CFAR - Collège Français des Anesthésistes-Réanimateurs, 75016 Paris, France
[f] Collective Thinking, 23 rue Yves Toudic, 75010 Paris, France
[g] Department of Anesthesiology and Critical Care Medicine, Lapeyronie University Hospital, 34295 Montpellier Cedex 5, France
[h] Inserm Unit 1298 Montpellier NeuroSciences Institute, Montpellier University, 34295 Montpellier Cedex 5, France

A R T I C L E I N F O

Article history:
Available online 6 May 2024

Keywords:
Adverse events
Patient safety
Natural language processing
Artificial intelligence
Quality improvement

A B S T R A C T

Background: Reporting and analysis of adverse events (AE) is associated with improved health system learning, quality outcomes, and patient safety. Manual text analysis is time-consuming, costly, and prone to human errors. We aimed to demonstrate the feasibility of novel machine learning and natural language processing (NLP) approaches for early predictions of adverse events and provide input to direct quality improvement and patient safety initiatives.
Methods: We used machine learning to analyze 9559 continuously reported AE by clinicians and healthcare systems to the French National Health accreditor (HAS) between January 1, 2009, and December 31, 2020 . We validated the labeling of 135,000 unique de-identified AE reports and determined the associations between different system's root causes and patient consequences. The model was validated by independent expert anesthesiologists.
Results: The machine learning (ML) and Artificial Intelligence (AI) model trained on 9559 AE datasets accurately categorized 8800 (88%) of reported AE. The three most frequent AE types were "difficult orotracheal intubation" (16.9% of AE reports), "medication error" (10.5%), and "post-induction hypotension" (6.9%). The accuracy of the AI model reached 70.9% sensitivity, 96.6% specificity for "difficult intubation", 43.2% sensitivity, and 98.9% specificity for "medication error."
Conclusions: This unsupervised ML method provides an accurate, automated, AI-supported search algorithm that ranks and helps to understand complex risk patterns and has greater speed, precision, and clarity when compared to manual human data extraction. Machine learning and Natural language processing (NLP) models can effectively be used to process natural language AE reports and augment expert clinician input. This model can support clinical applications and methodological standards and used to better inform and enhance decision-making for improved risk management and patient safety.
Trial Registration: The study was approved by the ethics committee of the French Society of Anesthesiology (IRB 00010254-2020-20) and the CNIL (CNIL: 118 58 95) and the study was registered with ClinicalTrials.gov (NCT: NCT05185479).

* Corresponding author.
E-mail address: contact@cfar.org (G. Jurkolow).
[1] Both first and second authors contributed equally to the manuscript.

https://doi.org/10.1016/j.accpm.2024.101390
2352-5568/

1. Introduction

The 1999 landmark report, *To Err is Human* [1], prompted a deep reflection on the state of patient care, raising awareness about the importance of a culture of safety and learning by analyzing reported adverse events, now considered a central tenet of an effective safety management system [2]. Adverse events related to unsafe care represent one of the top ten causes of death and disability worldwide, with at least a third of AE deemed to be preventable [3]. Incident reporting can contribute to patient safety but has only marginally been demonstrated to prevent patient harm [4]. Difficulties in processing incident reports, inadequate engagement of practitioners, lack of a safety culture, limited funding and institutional support for incident reporting systems, and use of complex health information technologies, remain major limitations in scaling up reporting systems for clinician engagement and effective learning from adverse incidents [5].

The field of anesthesiology has pioneered and is seen as a world leader in incident reporting classification and analyses [6–10] and in the development of methodological tools for improving the safety of care [11–13]. The first studies focused on structured data that are more easily exploitable than free text data analysis. The studies were carried out manually or semi-automatically and were time-consuming and included unrepeatable strategies [13–15]. Many national adverse event reporting systems have been implemented but with only limited demonstratable safety improvemen such as a reduction in medication errors. The classification of incident types is chosen from a structured list of options which may worsen data validity if the format does not make sense to the incident reporters [16], if structured choices do not allow to adequately summarize the incident, or if important contextual information is missing [17]. Multi-dimensional incident reports are impractical to fully explore using traditional methods by clinicians or national accreditors, and represent opportunities for automated, machine learning–based approaches [18–20].

Machine-learning (ML) techniques, particularly Natural language processing (NLP), excel in the analyses of complex signals in data-rich environments and have recently received major global interest [21–23]. Several studies have demonstrated the power of ML tools combined with big data in the analysis of medical free texts [24–38] with strategies that provide solutions to predict patient harm ranging from the indexing of texts [38] to providing apriori clinical predictions through robust data classifications [39,40].

The goal of this exploratory study was to demonstrate the feasibility of ML and NLP, approaches in the analysis of clinical adverse events reported during the perioperative period to advance patient safety. We ascertained how a "naïve" ML model can learn based on analyzing Adverse Event (AE) reports. We developed, trained, and validated our model using a large, contemporary cohort from the French Nationally Accreditation System (HAS) that mandates all healthcare providers and healthcare systems in France to report AE. We sought to identify perioperative care variation events whose correlated issues may be interacting to lead to adverse events and combined events that could predispose to patient harm and validate these associations using expert clinicians' input. We evaluated the performance of the best-performing models for further model predictions with a focus on learning how to best prevent patient harm.

2. Methods

2.1. Study design

The study's focus was on the adverse events reported by anesthesiologists. The French HAS launched the healthcare medical accreditation system in 2007, and mandates the reporting of all perioperative adverse events associated with anesthesia and surgery [41,42]. The investigation and analysis of incident reports is done manually [43,44]. Proprietary incident reporting systems record a combination of structured data entry fields and free text responses [41,42,45]. Free text responses provide the ability to describe the incident, but the completeness and accuracy of reporting limit the data validity and external generalizability of the findings. We followed the guidance of the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting checklist [46].

We describe the different stages of the study and provide an overview of which methods were used and what functions they provide as follows:

1 Selection of adverse events: All adverse events reported by anesthesiologists between January 1, 2009, and June 1, 2020, were included in the corpus (see part 2.2).
2 Data preparation: A classic approach of preprocessing was used: text was tokenized to allow the development of NLP models (see part 2.3).
3 Model development: An LDA algorithm was used to develop a model for automatically reporting AE labels (see parts 2.4 and 2.5).
4 Statistical analyses and clinical relevance: A 2D principal component analysis was used to assess the model performance and a K-means cluster analysis was used to identify deep associations between groups of AEs (see part 2.6).

2.2. Dataset

The HAS database includes 135,000 unique de-identified AE reports. Adverse events are categorized according to the clinical specialty of the report. Each AE contains structured data (sex, age, disease, comorbidity, avoid-ability, etc.) and free text data (history of the event) (see Appendix A.1, Supplementary data). Each AE is analyzed by an expert physician trained in AE analysis, in close collaboration with the reporting physician, before their validation and integration into the database (see Appendix A.2, Supplementary data).

We included all AEs consecutively reported by anesthesiologists between January 1, 2009, and June 1, 2020. Ten thousand AE reports were available at the time of data extraction. The study included the date, list of keywords, patient characteristics, summary of the AE, detailed description of the AE, type of procedure, diagnosis, duration of the procedure, severity of the AE, etc. The data could either be structured or unstructured (free text data). We used the free text data for textual analysis to assess the machine learning capability, AE summary, and detailed descriptions of the patient and what was done to them. The fields with free text data contained meaningful information about the AE in the form of unstructured free text and were considered less biased than other semi-structured data such as keyword description, for instance, which classification choice was made by the reporting physician.

2.3. Data preparation

We used the following variables to describe the population affected by the 9559 AE reported events: sex, age, BMI, ASA score, main diagnosis of treatment, and technical complexity of the intervention. We first applied the classic data preprocessing steps on free text data. The free text from the summary and detailed description with free text data were concatenated. The words were then parsed, tokenized, and lemmatized (grouping together different forms of a word). Standard stop-words were eliminated (a stop word is a common word that carries little (or perhaps no) meaningful information), and a list of n-grams (1-gram, 2-grams, and 3-grams). N-grams are sequences of words useful in many text analytic applications where sequences of words are relevant. Unigrams are single words, bigrams sequences of two words, trigrams of three words, and so on) were extracted. In order to focus on the most relevant n-grams and improve the extraction of meaningful and representative topics, an arbitrary choice was made to filter out any n-grams appearing either in less than 50 AE reports (so that the algorithm would only use words present in a sufficient number of reports) or in more than 40% of reports (thus eliminating words or expressions that were very commonly used and therefore had a lower capacity to create contrast for topic identification). Lastly, all AE reports that contained less than 25 tokens (among those remaining after the token selection process described above) were excluded from the analysis. This last step is done to ensure the presence of a sufficient number of tokens for the unsupervised topic modeling algorithm can provide relevant results. The identification of regular expressions is a central element of this study. This processing of the text makes it possible to select the less relevant verbatims, tread the text, group the closest character strings (by dealing directly with the errors in the words), and propose an automatic reading of the whole verbatim.

From the initial 9559 AE reports available at the time of database extraction, 7799 (81.6%) were retained after data pre-processing and report selection (Fig. A). All further analyses were performed on the 7799 preprocessed reports. A Latent Dirichlet Allocation (LDA) algorithm was applied to the tokenized text of the 7799 selected AE reports.

2.4. LDA models [47,48]

Latent Dirichlet Allocation (LDA) is used for topic modeling to uncover a list of abstract topics from a collection of documents. No prior knowledge is required, and every topic is discovered in an unsupervised manner. We used a classic approach: using n-gram (word cutting) makes it possible for the algorithm to recognize words (including words with mistakes, etc.). It then groups together all the Adverse Events where the recurrence of the same words is close. This grouping makes it possible to define themes (e.g. difficulty of intubation, medication errors, hemorrhages) otherwise called topics (Table 1).

Table 1
List of 45 labeled topics*.

Topic number	Label
1	Postoperative bleeding
2	Hardware/respirator problem
3	Obstetrics
4	Undefined
5	Bridging Anticoagulation
6	Post-induction collapse
7	Cardiology/Troponin elevation/Coronary artery disease
8	Patient/family conflict
9	Pain/Analgesia
10	Error in patient record/patient identity
11	Undefined
12	Side error
13	Operating table/falling or moving patient
14	Transfusion delay/blood order
15	Difficult Intubation
16	Electronic prescription software
17	Tooth dislocation/avulsion/fracture
18	Post-induction collapse
19	Undefined
20	Undefined
21	Pediatrics
22	Accidental Extubation/Post-Extubation Respiratory Distress
23	Epidural anesthesia
24	Obesity
25	Undefined
26	Communication failure/non-compliance to protocol
27	Anemia
28	Undefined
29	Severe patient/ICU transfer/sepsis
30	Medication error
31	Difficult Intubation
32	Difficult Intubation
33	Anticoagulation/Anti-aggregation
34	Difficult Intubation
35	Difficult Intubation
36	Undefined
37	Undefined
38	Operating equipment/OR/OR schedule
39	Per-operative complications (keratitis, pneumothorax, fracture, etc.)
40	Pre-anesthetic consultation/Anesthetic record
41	Biological results unavailable or not consulted
42	Medication error
43	Emergency situations/patient transfer
44	Difficult Intubation
45	Undefined

* after review by two expert anesthesiologists.

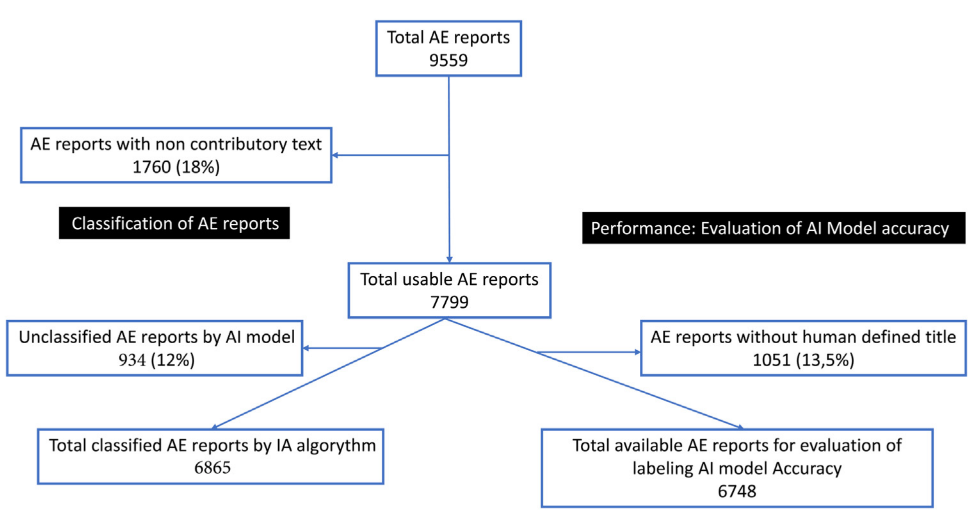

Fig. A. Selection of Adverse Events (AE) from the AE database reported by anesthesiologists between 2009 and 2020 used for the AE classification and for evaluation of the AI model accuracy.

In the LDA model, each document is produced by a mixture of topics in different proportions. Each word of the document is then produced by one of the document's topics through its per-topic word distribution. Given a set of documents and a number of topics, it is possible to infer the topic and word distributions that could have best generated it, which can be viewed as discovering the underlying topics. More complex topic modeling approaches using word embeddings were considered, namely ETM [1] and Top2Vec [2]. However, their preliminary inconclusive results led us to focus solely on LDA for the remainder of this paper [47,48].

The LDA algorithm clusters terms into a predefined number of "topics" based on the probability of those terms being used together within a document. It then predicts which topic a document will belong to, based on the terms in the document [49]. The resulting outputs include a list of topics, with each topic defined by a specific probability distribution for words or multi-word expressions. A list of the tokens that show the strongest association with a given topic can then be extracted. In this study, all topic labeling tasks were carried out by two physicians trained in the field of AE reporting and analysis (JMD, PMM).

2.5. Definition of the number of topics

We used an unsupervised method, and therefore when the ratio of the number of topics to the number of analyzed documents was high, there was a risk of obtaining non-meaningful topics either because noise is added by fortuitous word associations (words statistically found together in the same reports on average, but with no interpretable meaning), or because the list of tokens associated with a topic is highly specific but not meaningful (for instance a list of measurement units). In this study, a variable number of topics (between 30 and 70) was tested. A review of the different topic lists produced allowed us to determine the final number of topics by selecting the solutions that provided the best ratio of information to noise.

2.6. Statistical analysis and clinical relevance

A two-dimensional principal component analysis (PCA) was conducted to assess the model's performance based on their token distributions to help pool together topics with similar meanings. The number of topics was increased to obtain a more refined level of information.

The LDA algorithm permitted the implementation of an automatic AE report labeling procedure. For each AE report, the LDA model provided a relative weight between 0 and 1 associated with each topic (with all weights summing up to one for each AE report). Higher weights for a specific topic reflect the higher importance of that topic within a given AE report. For each AE report, the topic with the highest weight was considered to be the predominant topic. In order to provide a general idea of the level of accuracy of the AI model, our specific focus was on two types of AE categories in the database that represent the larger proportion of the reported AEs. The two categories are "difficult intubation" and "medication error" and in both, there is little room for interpretation or bias. For these two categories, the titles of the AE reports were screened for evidence of either one or the other, and the dominant topics identified by the AI model were checked against the labeling process. The labeling performance of the artificial intelligence model was analyzed by calculating the sensitivity and specificity measures. The reference ("true") label was defined as the manual classification provided by expert anesthesiologists, all board certified, and in practice for at least 12 years (PMM, JT, JMD, PGY, XC, GJ). The review and validation of classifications and topics were initially carried out independently by two clinician experts (GJ and PMM), and then discussed together for final validation. Disagreements were discussed until a full consensus was reached.

2.6.1. K-means clustering

In addition to the identification of the "predominant topic" for each AE report, the document-topic distributions obtained by the LDA algorithm were also used as input for K-means clustering. This clustering procedure provids an exclusive classification of AE reports in addition to the multi-topic labeling. The advantage of a "predominant topic" over a simple attribution is that important and meaningful nuances of the AE description related to the presence of multiple and strongly associated topics are not neglected but taken into account when clustering the AE reports.

2.6.2. A t-SNE (t-distributed stochastic neighbor embedding) 2-D map

A t-SNE (t-distributed stochastic neighbor embedding) 2-D map to visualize all AE reports was used to demonstrate the cluster display visual associations among clusters (neighboring clusters on the map are more closely related than distant clusters, *i.e.*, when there is a level of correlation in their underlying topics) [49]. Several associations were reviewed by three expert anesthesiologists (hybrid analysis) in order to ensure they align with real-world clinical practice (JMD, PMM, JT).

The categorical data were expressed in absolute values (n) or frequencies (%). Descriptive statistics were conducted with STATA 12 SE (StataCorp. 2011. Stata Statistical Software: Release 12. College Station, TX: StataCorp LP) and NLP analysis with Python (spacy 2.3.2 fr-core-news-md, nltk, genism 3.8.3, sklearn 0.0, pyLDAvis 2.1.2, numpy, pandas, embedding, fasttext 0.9.2, seaborn, matplotlib, bokeh 2.1.1).

3. Results

3.1. Model development and validation

Fig. A displays the performance of the model derived from the 9559 AE reports. The principal characteristics of patients in the reported AE are presented in Appendix B, Supplementary data. The final number of topics used as an input parameter for the LDA model was 45 (for results see Appendix C, Supplementary data). The list of 45 topics reviewed by the expert physicians is provided in Table 1.

The two-dimensional results of the principal component analysis (PCA) for the 45 topics identified based on their token contents are presented in Fig. B. Out of 45 topics inferred by the LDA model, 36 (80%) have an interpretable meaning. Some topics identified as distinct concepts by the LDA model appeared to share the same meaning and were labeled identically. These topics are close to each other or even overlap in the PCA representation. The classification is based on the defined dominant topic (the topics that have the strongest recurrence of terms). For the undefined topics in which several topics are illustrated we cluster them together as we didnt find one dominant topic.

We chose to focus on the following three topic pools because historically, they represent the most frequent types of reported AE in anesthesia AE databases including: "difficult intubation", "medication error", and "post-induction hypotension" (Fig. B). The first 3 predominant topics identified in the AE reports by the AI model were "difficult intubation" (16.9% of AE reports), "medication error" (10.5%), and "post-induction hypotension" (6.9%) Fig. C. In 12.0% of the AE reports (934/7799) the predominant topic labeled by the AI model was undefined (Fig. A).

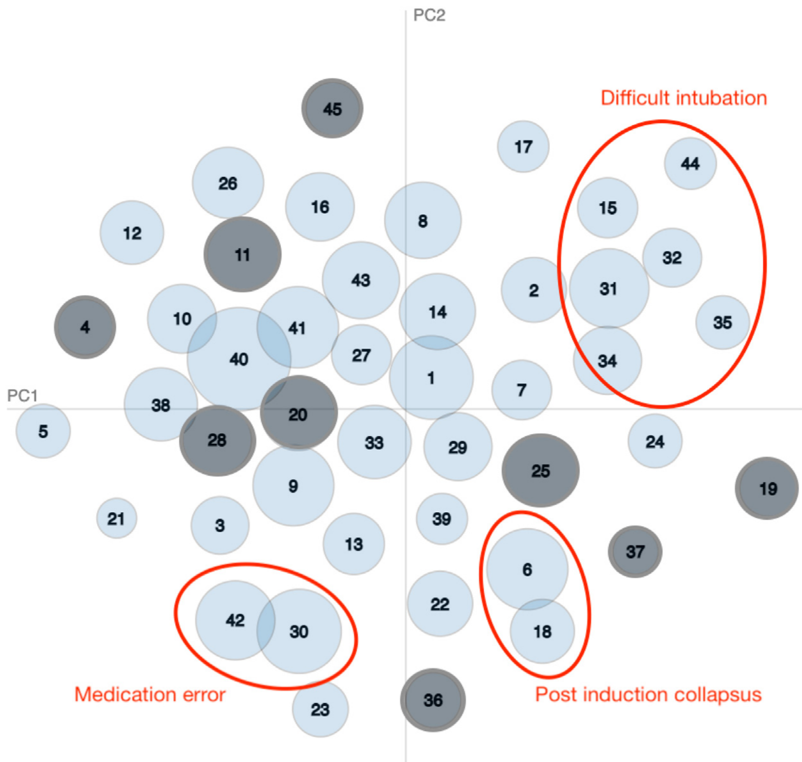

Fig. B. Principal component analysis (PCA) in 2 dimensions (PC1 and PC2) of the 45 identified topics based on their token contents. As an illustration, the groups of topics related to difficult intubation, medication error, or post-induction hypotension are circled in red. The topics considered as undefined are shown in grey.

Further results on the performance of the model make it possible to highlight the deep associations between the topics within the AE reports as displayed in Tables 2a, 2b, and 2c. The tables list the five topics most associated with the frequent AEs that impact clinical care: difficult intubation, medication error, and post-induction hypotension, respectively.

3.2. Model performance

The evaluation of the AI model's accuracy is shown in Fig. B for labeling and assigning the predominant topic of AE reports, for example, in the cases of "difficult intubation" and "medication error". The manual expert classification is based on the titles used

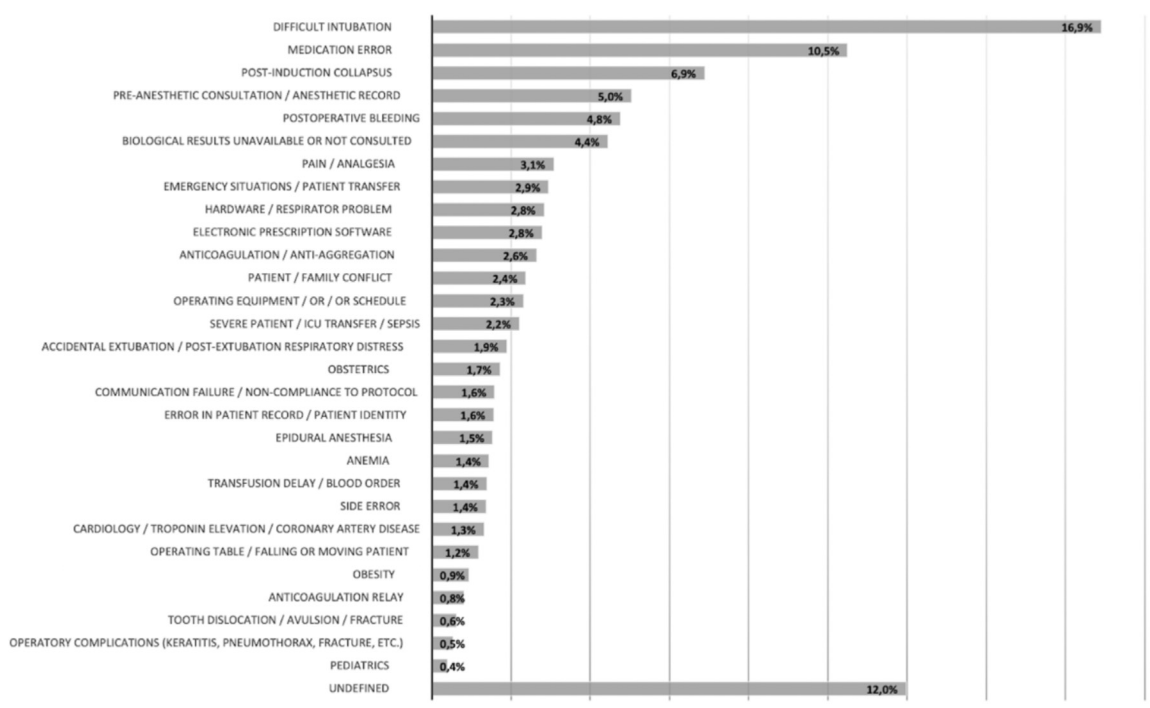

Fig. C. Classification of AE reports according to their predominant topic. Results are expressed as a percentage.

Table 2a
Top 5 topics with difficult intubation as a predominant topic in AE reports (n = 1318).

Associated topic	No of reports	% of reports
Tooth dislocation/avulsion/fracture	72	5.5%
Accidental Extubation/Post-Extubation Respiratory Distress	67	5.1%
Hardware/respirator problem	64	4.9%
Obesity	56	4.2%
Pre-anesthetic consultation/Anesthetic record	53	4.0%

Table 2b
Top 5 topics with medication error as a predominant topic in AE reports (n = 816).

Associated topic	No of reports	% of reports
Post-induction collapse	120	14.7%
Epidural anesthesia	72	8.8%
Operating equipment/OR/OR schedule	56	6.9%
Obstetrics	55	6.7%
Pain/Analgesia	49	6.0%

Table 2c
Top 5 topics with post-induction collapse as a predominant topic in AE reports (n = 533).

Associated topic	No of reports	% of reports
Medication error	71	13.3%
Difficult Intubation	48	9.0%
Operating table/falling or moving patient	45	8.4%
Pain/Analgesia	29	5.4%
Cardiology/Troponin elevation/coronary artery disease	25	4.7%

Table 3
List of 20 cluster labels.

Cluster number	Cluster label
0	Difficult intubation
1	Respirator issue
2	Post-operative hemorrhage
3	Severe complication
4	Conflict with patient or family
5	Analgesia
6	Pre-anesthetic consultation
7	Medication error & epidural anesthesia
8	Difficult intubation
9	Medication error & epidural anesthesia
10	Delayed transfusion
11	Post induction collapse
12	Prescription software
13	Identity/record error
14	Difficult intubation
15	Bridging Anticoagulation
16	Unavailable or unchecked lab results
17	Laterality error
18	Emergency situation
19	Undefined

in the AE reports as a reference to identify the two types of AE categories in the database. The AE report titles were available for 6748 out of 7799 reports (Fig. A). The valuation could only be carried out on the sample of 6748 AE reports. For "difficult intubation", a total of 89 different possible titles were identified, and 1109 out of 6748 AE had a title close to "difficult intubation". For "medication error", 61 different possible titles were identified, and 1328 out of 6748 AE had a title close to "medication error". For "difficult intubation", the AI model achieved 70.9% sensitivity and 96.6% specificity for labeling the topic and assigning it as the predominant topic of the AE report. For medication error, the sensitivity was 43.2% and specificity was 98.9%.

3.3. Clinical relevance of the observed results

A K-means clustering algorithm was applied to the 7799 AE reports based on their topic distributions as assigned by the LDA model. The optimal number of clusters was found to be 20 according to the silhouette score. The list of clusters and their tokens content with associated term frequency-inverse document frequency weights (TF-ITF) is provided in Appendix D, Supplementary data. The cluster contents were reviewed and labeled by two expert physicians (JT, GJ). The list of cluster labels is provided in Table 3.

Fig. D shows the t-SNE 2-D map visualization created from the clustering algorithm. Each dot on the map represents an individual AE report (n = 7799). The AE reports clustered together have the same color code. The geographical proximity of clusters (as well as overlapping areas) points to deep associations between the groups of AEs (and their underlying topics). For instance, four specific inter-topic associations are highlighted in Fig. D as follows:

i) on the upper left: proximity between delayed blood transfusion, lack of lab results, anticoagulation delay, and post-op hemorrhage;
ii) on the lower left: between anesthesiology consultation, identity or medical record confusion, and laterality error in the operating room;
iii) on the upper right: between prescription software, medication error, and post-induction hypotension; and,
iv) on the lower right: proximity between conflict with a patient or family (sometimes between caregivers) and emergency situations (life-threatening).

To illustrate the value and impact of using a hybrid analysis, expert clinicians distinguished between 2 situations, contextual and temporal, for a wrong-side surgery laterality error when reviewing cluster 17 (Table 3). Indeed, part of the TOKENS of cluster 17: "cataract", "check list", and "verification" led the human experts to analyze the description of the AEs of this cluster, which concerns cataract surgery. This effort highlights a key finding in that the WHO surgery checklist, performed before surgical incision and whose completion is mandatory in France, was an efficient barrier for detecting temporality side errors *before incision* and therefore before any *harm* was caused to the patient.

Conversely, another part of the TOKENS of cluster 17: "regional anesthesia" "side error" and "block performance" led the experts to review in detail the text of the AEs of this cluster which concerns the performance of regional anesthesia blocks on the wrong side of the patient. In this case, the detection of the laterality error unfortunately often takes place *after* performing the regional anesthesia, which was performed before the surgeon entered the operating room to conduct the checklist , implying that the WHO Surgical Safety checklist was not implemented correctly. This provides important insights into how best to improve the effective implementation of surgical checklists.

4. Discussion

Machine learning, Natural language processing (NLP) tools can provide insights into adverse events and support improved standards of implementation. These tools allow for the successful classification of 88% of AEs as compared to the manual extraction without the use of keywords, and demonstrate the feasibility of

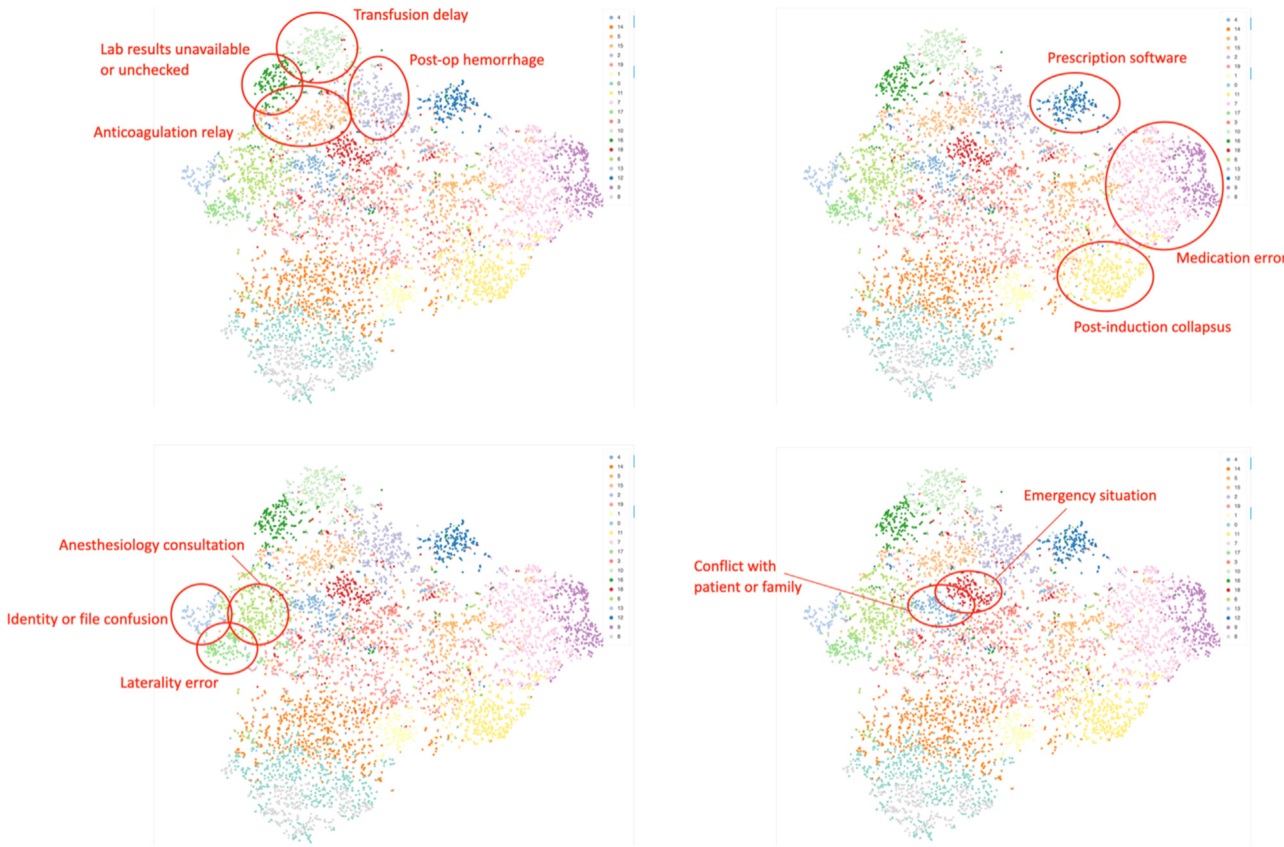

Fig. D. A t-SNE 2-D map is used to visualizae high dimensional data where each point represents an individual AE report (n = 7799). AE reports that clustered together have the same color code.

NLP approaches in the analysis of multi-dimensional clinical AEs. The protocol identified the most frequent categories of events: unpredicted difficult intubation [50], medication error [51,52], and post-induction hypotension [53–55]. These results support the need to modify how clinician report, collect, and analyze adverse events. A simple free-text declaration becomes possible without the use of predefined keywords, which allows for easier AE reporting by practitioners. Searching for AE categories chosen at the time of data extraction allowed for easier and more efficient analysis of the database, even if the search and analysis of this type of events had not been planned, greatly enlarging the possibilities for data analyses and care improvement.

4.1. Strengths of machine learning to enhance AE reports analysis

Our novel study approach has several strengths. First, our approach allows for a less reactive and global quantitative and qualitative review of the causes of, and barriers to, reporting perioperative adverse events. It also allows for deeper, system-wide insights into the multidisciplinary causes of adverse events. This approach supports recommendations regarding the typologies of AE or associations of system's root causes that often lead to common AEs and patient harm.

Reporting of AEs using free text provides a possibility for a deeper understanding of why AE continues to occur in spite of many policy and educational efforts and enriches the information contained in each AE report. The ML allows for a better characterization of the frequency of AE occurrences using various clusters. ML can help guide future preventative actions aimed at improving the quality and safety targeting of events according to their frequency. Similarly, this approach makes it possible to highlight and amplify the emergence of weak signals and new risky situations, which are impossible to identify when a large number of AE reports are read by different experts. Part of the increased frequency observed in the occurrence of these AE is linked to the unique structure of our database and the French HAS national accreditation program. During the first years of implementation of this program, the adverse events reported were predefined (including difficult intubation and medication errors). This reflects the desire for critical patient events considered important by the French College of Anesthesiologists and Intensivists (FCAI). However, in 2011, the FCAI required that the reporting be extended to all types of AEs. This may explain why the occurrence of post-induction hypotension rapidly appeared among the most frequently reported AEs, addressing a major concern that has emerged in the anesthetic literature over the past decade [52–54]. Moreover, although the ML-based predictive models we report on only allow for a ranking of the main topics for each type of event reported, the identification of a recurrence of themes has help guide the management of specific patient risks. Second, the use of machine learning tools for this type of data analysis allows to relieve the time demands of expert anesthesiologists which is a key constraint in r the HAS perioperative AE reporting and feedback program.

Third, expert AE analysis is often focused on the analysis of the risks identified by the tool and the association of root causes. The AEs reported in our database are classified using a standard manner using traditional keywords. Each report is co-designed in a joint analysis by an expert clinician and by the anesthesiologist reporting the event to improve the quality, candor and frequency of voluntary reports. The classification using NLP makes it possible to refocus the time and work of experts, freeing them to focus on the system's causes of events and potential correlated risks that may be interacting and lead to adverse events. Third, the number of experts does not increase the number of reported events. Extending the reporting to all AEs encountered by anesthesiologists has significantly increased the workload for experts reviewing AEs. In many cases, the keywords used to facilitate the categorization of events have not been predefined. In addition, the number of event categories initially selected does not address all types of reported AE. The gradual enrichment in the types of events reported has also led to a loss of information because their frequency could only be analyzed once the keyword or event category had been implemented. For example, the emergence of an increasing number of adverse drug events related to computer malfunctions has become a significant issue in recent years [56,57]. Unfortunately, this theme [58–60] was not initially individualized. Fourth, the possible links observed between the different AEs seem clinically consistent and represent a new application in assessing the impacts of corrective measures applied to prevent AE and a source for quality improvement initiatives.

4.2. Clinical relevance of machine learning and natural language processing

ML and NLP point to pervasive communication problems within a clinical team or between teams, as well as the quality of relationships between caregivers, to be the determining factors in the occurrence and prevention of AEs [58–63]. Although the analyses of links between clusters based on the analysis of t-SNE should be viewed with caution, some links appear to be highly relevant. For example, the proximity between the different clusters suggests potential links between team conflicts (particularly with patients or their families, but also between caregivers) and emergency situations (Fig. D). These observations indicate that a combination of different human factors likely contributes to or increases the incidence of ineffective communications causing or contributing to the AE occurrence [64]. This result raises the possibility of apparently unrelated corrective measures which, through their combination, may reduce the frequency of AE. This is consistent with the growing awareness that instead of applying corrective measures to one factor they could be applied to a bundle of needed factors to better eliminate or mitigate the risks of complex patterns and underlying system vulnerabilities that lead to AE [65].

Research in the field of risk prevention has continued to grow for several years [66], but evaluating the impact of these novel methods remains incomplete [67]. It is essential to limit the use of time-consuming, inefficient, and unrepeatable methods in favor of effective methods [68–70]. ML offers a promising solution in this regard. This is illustrated by our analysis of the clusters of events related to wrong side/site laterality error surgery prevention with respect to potential AEs detected before they reach the patient [71]. Measures designed to reduce wrong-site surgery [71], including the WHO Safe Surgery checklist, were shown to be efficient measures [72]. On the contrary, medication error-related AEs are usually detected after the patient was harmed, suggesting that despite many efforts to reduce medication harm occurrence, medication related AEs remain stubbornly frequent [51,73,74].

4.3. Limitations of the study

Our study has several limitations. First, the French HAS reporting system is not exhaustive and is not intended to be epidemiologically comprehensive [41,42]. The reported events probably do not address all the risky perioperative situations that patients experience. We know that adverse events as a metric can be problematic due to the heterogenicity in events reported that have different root causes and might require different interventions [75]. Nevertheless, the advantage of the ML-based, NLP approach using identification of non-detectable weak signals is unique when compared to a cumbersome, manual analysis by experts. Second, there is no gold standard reference to evaluate the performance of the AI model for labeling AE reports due to the heterogeneity of the titles and keywords used in the AE reports database. Since the AE database was created over time, there were no strict guidelines or constraints in describing the AE within the structure-free text data, or when there were, they evolved over time. Third, the low sensitivities for medication errors suggest that AI-based ML needs further refinement, as a large number of false negatives could be missed. Fourth, it is also possible that reported events did not actually occur, which could skew the observed frequencies. Nevertheless, this risk is a marginal risk, and an external control of AE reports by peer clinicians and the French HAS staff, help ensure the overall accuracy of the reports [42]. Fifth, with respect to the ML approach used in this study, we note that this is an unsupervised language model. It was not possible to classify all reported AEs, but this performance is the one usually applied and accepted for this type of research [76]. The purpose of this study was to assess the feasibility and relevance of such tools on unannounced medical data without a prediction objective. The purpose of the applied models was merely to classify the AE according to identified main themes. Sixth, several human factors and systems' themes can coexist in a specific AE report but the classification into "dominant topics" highlights only the topics showing the strongest association with a given AE. In the 12% unclassified AE, several themes were present, but without the emergence of one main theme, the tool is limited in integrating them into a cluster. This advocates for a supervised approach to AE analysis and is therefore directly dependent on the quality of the AEs reported by clinicians. The quality control was carried out on only three themes (the results for two of them were described in this article) and deserves to be extended to all identified AE themes. Seventh, with regard to the performance of the protocol and the differences in observed sensitivities between the "difficult intubation" and the "medication error" themes, several explanations can also be provided: i) the difference in semantics and synonymy (phrase that means exactly or nearly the same as another word) between the two themes: intuitively, it is clear that the vocabulary used to evoke drug errors is richer than for difficult intubation, and as the tool is not assisted at all, it must learn about word associations and their recurrence in order to classify medication errors, an "effort" that it does not have to provide for intubation difficulties; ii) the human brain comprehends the associations of words and expressions more effectively in relation to a drug error and even if this precise term does not appear in a clear statement. An expert is able to produce this title without difficulty, whereas the applied algorithm was less successful in retrieving complex word associations. There is a greater risk of observing noise, and classification difficulties when synonymy is important and reasoning is more complex. Finally, external validation from other countries and varied datasets, with differently run and overseen healthcare systems is lacking, thus our results require further validation.

5. Conclusions

We developed and validated an automated ML and NLP search algorithm protocol using a large volume of documents sourced from a nationwide reporting system. This novel approach can effectively analyze and interpret adverse incident identification and propose a novel classification of perioperative adverse events. We found the performance of the tool variable but it can effectively guide complementary analyses by expert clinicians. This work needs to be supplemented with supervised models to allow for its refinement. The good performance of our proof-of-concept models highlights how ML-surfaced relationships between adverse events could help guide futureefforts directed at preventive safety actions by taking into account the varied, contextual contributing factors. Machine learning may be leveraged to classify adverse events and advance risk management to help providers save time while delivering safer patient care. Our study should inform health policy in support of well-designed and overseen ML data analytic tools.

Ethical statement

The French law on biomedical researchs (Article L.1121-1 and Articles R.1121-1 and R.1121-2 of the public health code) does not apply to this retrospective and observational study, however, this one does not present a particular ethical problem.

We call nevertheless your attention the fact that, in this context, because of the recording of various data and information, it is up to you to inquire about the obligations connected to the statements with the CNIL (NATIONAL COMMISSION FOR INFORMATION TECHNOLOGY AND CIVIL LIBERTIES).

Funding statement

Support was provided solely from the Collège Français des Anesthésistes-Réanimateurs (CFAR).

Conflicts of interest

P M Mertes (PMM): has no conflict of interest to declare;
C Morgand (CM): has no conflict of interest to declare;
P Barach (PB): has no conflict of interest to declare;
G Jurkolow (GJ): has no conflict of interest to declare;
K Assmann (KA): has no conflict of interest to declare;
E Dufetelle (ED): works at Collective Thinking; Collective Thinking was hired by CFAR to carry out AI development and data analysis for the study; was not involved in data interpretation, manuscript preparation or final conclusions.
V Susplugas (VS): works at Collective Thinking; Collective Thinking was hired by CFAR to carry out AI development and data analysis for the study; was not involved in data interpretation, manuscript preparation or final conclusions.
B Alauddin (BA): works at Collective Thinking; Collective Thinking was hired by CFAR to carry out AI development and data analysis for the study; was not involved in data interpretation, manuscript preparation or final conclusions.
PG Yavordios (PGY): has no conflict of interest to declare;
J Tourres (JT): has no conflict of interest to declare;
JM Dumeix (JMD): has no conflict of interest to declare;
X Capdevila (XC): has no conflict of interest to declare.

Acknowledgments

CFAR: Collège Français des Anesthésistes-Réanimateurs: funded the project, collected and provided the content of the database.

HAS: Haute Autorité de Santé: collected, stored and provided the content of the database.

Collective Thinking: developed AI algorithms, processed all available data, and carried out AI model training and data analysis.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.accpm.2024.101390.

References

[1] Institute of Medicine, Kohn LT, Corrigan JM, Donaldson MS. To err is human. Washington DC: National Academy of Sciences; 2000.
[2] Wood KE, Nash DB. Mandatory state-based error-reporting systems: current and future prospects. Am J Med Qual 2005;20:297–303.
[3] Bates DW, Singh H. Two decades since to err is human: an assessment of progress and emerging priorities in patient safety. Health Aff (Millwood) 2018;37:1736–43.
[4] Arnal-Velasco D, Barach P. Anaesthesia and perioperative incident reporting systems: opportunities and challenges. Best Pract Res Clin Anaesthesiol 2021;35:93–103.
[5] Mitchell I, Schuster A, Smith K, Pronovost P, Wu A. Patient safety incident reporting: a qualitative study of thoughts and perceptions of experts 15 years after 'To Err is Human'. BMJ Qual Saf 2016;25:92–9.
[6] Tuppin P, Rivière S, Rigault A, Tala S, Drouin J, Pestel L, et al. Prevalence and economic burden of cardiovascular diseases in France in 2013 according to the national health insurance scheme database. Arch Cardiovasc Dis 2016;109:399–411.
[7] Bannay A, Chaignot C, Blotière PO, Basson M, Weill A, Ricordeau P, et al. The best use of the Charlson comorbidity index with electronic health care database to predict mortality. Med Care 2016;54:188–94.
[8] Grodner C, Sbidian E, Weill A, Mezzarobba M. Epidemiologic study in a real-world analysis of patients with treatment for psoriasis in the French national health insurance database. J Eur Acad Dermatol Venereol 2021;35:411–6.
[9] Mercereau L, Todd N, Rey G, Valleron AJ. Comparison of the temperature-mortality relationship in foreign born and native born died in France between 2000 and 2009. Int J Biometeorol 2017;61:1873–84.
[10] Constantinou P, Pelletier-Fleury N, Olié V, Gastaldi-Ménager C, JuillÈre Y, Tuppin P. Patient stratification for risk of readmission due to heart failure by using nationwide administrative data. J Card Fail 2021;27:266–76.
[11] Tuppin P, Samson S, Fagot-Campagna A, Lukacs B, Alla F, CNAMTS scientific board members. et al. Prostate cancer outcomes in France: treatments, adverse effects and two-year mortality. BMC Urol 2014;14:48.
[12] Tuppin P, Samson S, Fagot-Campagna A, Woimant F. Care pathways and healthcare use of stroke survivors six months after admission to an acute-care hospital in France in 2012. Rev Neurol (Paris) 2016;172:295–306.
[13] Cabarrot P, Coniel M, Haniquaut F, Fourali R, Morgand C, May-Michelangeli L, et al. La crise Covid a-t-elle submergé les barrières de sécurité du système de santé? Analyse des déclarations d'événements indésirables en lien avec la Covid-19 déclarés dans la base de l'accréditation des médecins et revue critique de la littérature. Risques Qualité 2020;17:195–205.
[14] Jougla E, Rossollin F, Niyonsenga A, Chappert JL, Johansson LA, Pavillon G. *Comparability and quality improvement in European causes of death statistics. Eurostat, Project 96/S 99-5761/EN* 2001.
[15] Pavillon G, Johansson LA. Production of methods and tools for improving causes of death statistics at codification level; 2001.
[16] Barach P, Phelps G. Clinical sensemaking: a systematic approach to reduce the impact of normalised deviance in the medical profession. J R Soc Med 2013;106:387–90.
[17] Liang C, Gong Y. Automated classification of multi-labeled patient safety reports: a shift from quantity to quality measure. Stud Health Technol Inform 2017;245:1070–4.
[18] Govindan M, Van Citters AD, Nelson EC, Kelly-Cummings J, Suresh G. Automated detection of harm in healthcare with information technology: a systematic review. Qual Saf Health Care 2010;19:e11.
[19] Carrell DS, Schoen RE, Leffler DA, Morris M, Rose S, Baer A, et al. Challenges in adapting existing clinical natural language processing systems to multiple, diverse health care settings. J Am Med Inform Assoc 2017;24:986–91.
[20] Pivovarov R, Elhadad N. Automated methods for the summarization of electronic health records. J Am Med Inform Assoc 2015;22:938–47.
[21] Krumholz HM. Big data and new knowledge in medicine: the thinking, training, and tools needed for a learning health system. Health Aff (Millwood) 2014;33:1163–70.
[22] Haug CJ, Drazen JM. Artificial intelligence and machine learning in clinical medicine, 2023. N Engl J Med 2023;388:1201–8.
[23] Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. N Engl J Med 2023;388:1233–9.
[24] Rochefort CM, Verma AD, Eguale T, Lee TC, Buckeridge DL. A novel method of adverse event detection can accurately identify venous thromboembolisms (VTEs) from narrative electronic health record data. J Am Med Inform Assoc 2015;22:155–65.
[25] Tian Z, Sun S, Eguale T, Rochefort CM. Automated extraction of VTE events from narrative radiology reports in electronic health records: a validation study. Med Care 2017;55:e73–80.
[26] Gálvez JA, Pappas JM, Ahumada L, Martin JN, Simpao AF, Rehman MA, et al. The use of natural language processing on pediatric diagnostic radiology reports in the electronic health record to identify deep venous thrombosis in children. J Thromb Thrombolysis 2017;44:281–90.
[27] Yim W-w, Kwan SW, Yetisgen M. Classifying tumor event attributes in radiology reports. J Assoc Inform Sci Technol 2017;68:2662–74.
[28] Moore CR, Farrag A, Ashkin E. Using natural language processing to extract abnormal results from cancer screening reports. J Patient Saf 2017;13:138–43.
[29] Yetisgen-Yildiz M, Gunn ML, Xia F, Payne TH. Automatic identification of critical follow-up recommendation sentences in radiology reports. In:

AMIA ... Annual Symposium Proceedings. AMIA Symposium 2011; 2011. p. 1593–602.
[30] Duke JD, Friedlin J. ADESSA: a real-time decision support service for delivery of semantically coded adverse drug event data. In: AMIA ... Annual Symposium Proceedings. AMIA Symposium 2010; 2010. p. 177–81.
[31] Li Q, Kirkendall ES, Hall ES, Ni Y, Lingren T, Kaiser M, et al. Automated detection of medication administration errors in neonatal intensive care. J Biomed Inform 2015;57:124–33.
[32] Ni Y, Lingren T, Hall ES, Leonard M, Melton K, Kirkendall ES. Designing and evaluating an automated system for real-time medication administration error detection in a neonatal intensive care unit. J Am Med Inform Assoc 2018;25:555–63.
[33] Cai T, Giannopoulos AA, Yu S, Kelil T, Ripley B, Kumamaru KK, et al. Natural language processing technologies in radiology research and clinical applications. Radiographics 2016;36:176–91.
[34] Nunes AP, Yang J, Radican L, Engel SS, Kurtyka K, Tunceli K, et al. Assessing occurrence of hypoglycemia and its severity from electronic health records of patients with type 2 diabetes mellitus. Diabetes Res Clin Pract 2016;121:192–203.
[35] Toyabe S. Characteristics of inpatient falls not reported in an incident reporting system. Glob J Health Sci 2015;8:17–25.
[36] Tanushi H, Kvist M, Sparrelid E. Detection of healthcare-associated urinary tract infection in Swedish electronic health records. Stud Health Technol Inform 2014;207:330–9.
[37] Carrell DS, Halgrim S, Tran DT, Buist DS, Chubak J, Chapman WW, et al. Using natural language processing to improve efficiency of manual chart abstraction in research: the case of breast cancer recurrence. Am J Epidemiol 2014;179:749–58.
[38] Falissard L, Morgand C, Ghosn W, Imbaud C, Bounebache K, Rey G. Neural translation and automated recognition of ICD10 medical entities from natural language. MIR Med Inform 2022;10e26353. http://dx.doi.org/10.2196/26353.
[39] Jensen K, Soguero-Ruiz C, Oyvind Mikalsen K, Lindsetmo RO, Kouskoumvekaki I, Girolami M, et al. Analysis of free text in electronic health records for identification of cancer patient trajectories. Sci Rep 2017;746226.
[40] Young IJB, Luz S, Lone N. A systematic review of natural language processing for classification tasks in the field of incident reporting and adverse event analysis. Int J Med Inf 2019;132103971.
[41] Cabarrot P, Legris C, May L, Grenier C. French process of accreditation of medical teams. London: Int For QaSiHC; 2017.
[42] Haute Autorité de Santé (HAS). Rapport d'activité 2020 des organismes agréés pour l'accréditation de la qualité de la pratique professionnelle des médecins et des équipes médicales. Available from /www.has-sante.fr/upload/docs/application/pdf/2021-04/196_rapport_d_activite_2020_0a_cd_2021_03_25_vd.pdf.
[43] Taylor-Adams S, Vincent C. Systems analysis of clinical incidents. The London protocol. London: Imperial College of London; 2019.
[44] Pronovost PJ, Thompson DA, Holzmueller CG, Lubomski LH, Dorman T, Dickman F, et al. Toward learning from patient safety reporting systems. J Crit Care 2007;21:305–15.
[45] Godlee F, Cabarrot P, Desplanques A, Smith J, Degos L. Foreword. Qual Saf Health Care 2010;19:A1–2.
[46] von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ 2007;335:806–8.
[47] Dieng AB, Ruiz FJ, Blei DM. Topic modeling in embedding spaces. Trans Assoc Comput Linguist 2020;439–53.
[48] Angelov D. Top2vec: Distributed representations of topics. arXiv preprint 2020: arXiv:2008.09470.
[49] Van der Maaten L, Hinton G. Visualizing data usint t-SNE. JMLR 2008;9:2579–605.
[50] Xu Z, Ma W, Hester DL, Jiang Y. Anticipated and unanticipated difficult airway management. Curr Opin Anaesthesiol 2018;31:96–103.
[51] Nanji KC, Patel A, Shaikh S, Seger DL, Bates DW. Evaluation of perioperative medication errors and adverse drug events. Anesthesiology 2016;124:25–34.
[52] Wahr JA, Abernathy 3rd JH, Lazarra EH, Keebler JR, Wall MH, Lynch I, et al. Medication safety in the operating room: literature and expert-based recommendations. Br J Anaesth 2017;118:32–43.
[53] Kang AR, Lee J, Jung W, Lee M, Park SY, Woo J, et al. Development of a prediction model for hypotension after induction of anesthesia using machine learning. PLoS ONE 2020;15e0231172.
[54] Khan AI, Fischer M, Pedoto AC, Seier K, Tan KS, Dalbagni G, et al. The impact of fluid optimisation before induction of anaesthesia on hypotension after induction. Anaesthesia 2020;75:634–41.
[55] Südfeld S, Brechnitz S, Wagner JY, Reese PC, Pinnschmidt HO, Reuter DA, et al. Post-induction hypotension and early intraoperative hypotension associated with general anaesthesia. Br J Anaesth 2017;119:57–64.
[56] Wachter RM, Howell MD. Resolving the productivity paradox of health information technology: a time for optimism. JAMA 2018;320:25–6.
[57] Choi K, Gitelman Y, Asch DA. Subscribing to your patients - reimagining the future of electronic health records. N Engl J Med 2018;378:1960–2.
[58] Etherington N, Wu M, Cheng-Boivin O, Larrigan S, Boet S. Interprofessional communication in the operating room: a narrative review to advance research and practice. Can J Anaesth 2019;66:1251–60.
[59] Frasier LL, Pavuluri Quamme SR, Ma Y, Wiegmann D, Leverson G, DuGoff EH, et al. Familiarity and communication in the operating room. J Surg Res 2019;235:395–403.
[60] Lee CT, Doran DM. The role of interpersonal relations in healthcare team communication and patient safety: a proposed model of interpersonal process in teamwork. Can J Nurs Res 2017;49:75–93.
[61] Lingard L, Espin S, Whyte S, Regehr G, Baker GR, Reznick R, et al. Communication failures in the operating room: an observational classification of recurrent types and effects. Qual Saf Health Care 2004;13:330–4.
[62] Sexton JB, Thomas EJ, Helmreich RL. Error, stress, and teamwork in medicine and aviation: cross sectional surveys. BMJ 2000;320:745–9.
[63] Douglas RN, Stephens LS, Posner KL, Davies JM, Mincer SL, Burden AR, et al. Communication failures contributing to patient injury in anaesthesia malpractice claims. Br J Anaesth 2021;127:470–8.
[64] Kelly FE, Frerk C, Bailey CR, Cook TM, Ferguson K, Flin R, et al. Human factors in anaesthesia: a narrative review. Anaesthesia 2023;78:479–90.
[65] Cassin BR, Barach PR. Making sense of root cause analysis investigations of surgery-related adverse events. Surg Clin North Am 2012;92:101–15.
[66] Wensing M, Bosch M, Grol R. Developing and selecting interventions for translating knowledge to action. CMAJ 2010;182:E85–8.
[67] Kellogg KM, Hettinger Z, Shah M, Wears RL, Sellers CR, Squires M, et al. Our current approach to root cause analysis: is it contributing to our failure to improve patient safety? BMJ Qual Saf 2017;26:381–7.
[68] Smith A, Alderson P. Guidelines in anaesthesia: support or constraint? Br J Anaesth 2012;109:1–4.
[69] Crosby E. Review article: the role of practice guidelines and evidence-based medicine in perioperative patient safety. Can J Anaesth 2013;60:143–51.
[70] Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. Lancet 2003;362:1225–30.
[71] Seiden SC, Barach P. Wrong-side/wrong-site, wrong-procedure, and wrong-patient adverse events: are they preventable? Arch Surg 2006;141:931–9.
[72] Algie CM, Mahar RK, Wasiak J, Batty L, Gruen RL, Mahar PD. Interventions for reducing wrong-site surgery and invasive clinical procedures. Cochrane Database Syst Rev 2015;2015Cd009404.
[73] Bratch R, Pandit JJ. An integrative review of method types used in the study of medication error during anaesthesia: implications for estimating incidence. Br J Anaesth 2021;127:458–69.
[74] Sanduende-Otero Y, Villalón-Coca J, Romero-García E, Díaz-Cambronero Ó, Barach P, Arnal-Velasco D. Patterns in medication incidents: a 10-yr experience of a cross-national anaesthesia incident reporting system. Br J Anaesth 2020;124:197–205.
[75] Thomas EJ, Petersen LA. Measuring errors and adverse events in health care. J Gen Intern Med 2003;18:61–7.
[76] Falissard L, Morgand C, Roussel S, Imbaud C, Ghosn W, Bounebache K, et al. A deep artificial neural network-based model for prediction of underlying cause of death from death certificates: algorithm development and validation. JMIR Med Inform 2020;8e17125.

Glossary

AE: Adverse Event
AI: Artificial Intelligence
ASA: American Society of Anesthesiology
BMI: Body Mass Index
CFAR: Collège Français d'Anesthésie Réanimation (French College of Anesthesiologists and Intensivists)
CMR: Computerized Medical Record
HAS: Haute Autorité de Santé (French National Health accreditor)
LDA: Latent Dirichlet Allocation
ML: Machine Learning
OR: Operating Room
PAC: Pre-Anesthesia Consultation
PCA: Principal Component Analysis
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
t-SNE: t-distributed Stochastic Neighbor Embedding
TF-ITF: term frequency-inverse document frequency weights
WHO: World Health Organization