Doctoral theses at NTNU, 2023:410

Melissa Y. Yan

Capturing Signs and Events Related to Catheters in Clinical Text

Semantic Annotation and Knowledge Representation

Doctoral thesis

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of Philosophiae Doctor
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

NTNU
Norwegian University of Science and Technology

Melissa Y. Yan

Capturing Signs and Events Related to Catheters in Clinical Text

Semantic Annotation and Knowledge Representation

Thesis for the Degree of Philosophiae Doctor

Trondheim, December 2023

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

NTNU

Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Information Technology and Electrical Engineering
Department of Computer Science

ISBN 978-82-326-7518-0 (printed ver.)
ISBN 978-82-326-7517-3 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

Doctoral theses at NTNU, 2023:410

Printed by NTNU Grafisk senter

Abstract

Annotated clinical corpora are necessary to extract information from clinical text for answering clinical questions. However, publicly available annotated clinical corpora are limited because of privacy issues, ethical concerns, and resource requirements for curating annotated corpora. When available, they are annotated for specific purposes and might lack the annotations required to answer clinical research questions. These challenges open up the opportunity to identify considerations and develop an annotated clinical corpus to answer a clinical research question.

To narrow the scope of capturing clinical concepts and knowledge within clinical text, this work focuses on a use case. The use case is capturing signs and events related to catheters from clinical adverse event notes for reducing sepsis and infection rates. This work addresses four research questions:

RQ1: What methods utilize clinical text to reduce sepsis and infections?

RQ2: What characteristics of catheter-related signs and events can be captured from clinical text?

RQ3: How can an incremental annotation-based method be developed to extract information about catheter-related signs and events from clinical text?

RQ4: How can clinical knowledge about catheter-related signs and events be captured?

Addressing the research questions resulted in four publications. The contributions are identifying research gaps, developing an annotated corpus, and developing a corresponding knowledge model. Results from this work are being extended in ongoing research.

Preface

This work is submitted to the Norwegian University of Science and Technology (NTNU) for partial fulfillment of the requirements for the degree of *Philosophiae Doctor*. Doctoral work was performed at the Department of Computer Science, NTNU, Trondheim, Norway under the supervision of Associate Professor Øystein Nytrø and the co-supervision of Associate Professor Lise Tuset Gustad, Professor Jan Kristian Damås, Professor Erik Solligård, and Professor Pieter Jelle Toussaint.

Acknowledgements

Funding was provided by the Computational Sepsis Mining and Modelling (CoSeM) project through the NTNU Health Strategic Area. Special thanks to my supervisors, committee members, collaborators, annotators, colleagues, friends, and family.

Melissa Y. Yan
Trondheim, December 2023

List of Papers

Paper A - Current Methods

Melissa Y. Yan, Lise Tuset Gustad, Øystein Nytrø. "Sepsis prediction, early detection, and identification using clinical text for machine learning: a systematic review". In: *Journal of the American Medical Informatics Association*. Vol. 29, Issue. 3 (2022), pp. 559–575. DOI: 10.1093/jamia/ocab236.

Paper B - Dataset Potential

Melissa Y. Yan, Lise Husby Høvik, André Pedersen, Lise Tuset Gustad, and Øystein Nytrø. "Preliminary Processing and Analysis of an Adverse Event Dataset for Detecting Sepsis-Related Events". In: *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. (2021), pp. 1605–1610. DOI: 10.1109/BIBM52615.2021.9669410.

Paper C - Annotation Method

Melissa Y. Yan, Lise Tuset Gustad, Lise Husby Høvik, and Øystein Nytrø. "Method for Designing Semantic Annotation of Sepsis Signs in Clinical Text". In: *Proceedings of the 5th Clinical Natural Language Processing Workshop (ClinicalNLP@ACL)*. (2023), pp. 236–246. DOI: 10.18653/v1/2023.clinicalnlp-1.29.

Paper D - Terminology and Ontology

Melissa Y. Yan, Lise Tuset Gustad, Lise Husby Høvik, and Øystein Nytrø. "Terminology and ontology development for semantic annotation: A use case on sepsis and adverse events". In: *Semantic Web*. Vol. 14, No. 5 (2023), pp. 811–871. DOI: 10.3233/SW-223226.

Other Paper

Melissa Y. Yan, Lise Husby Høvik, Lise Tuset Gustad, and Øystein Nytrø. "Understanding and Reasoning About Early Signs of Sepsis: From Annotation Guideline to Ontology". In: *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. (2021), pp. 1906–1911. DOI: 10.1109/BIBM52615.2021.9669311.

Contents

List of Figures

List of Tables

Acronyms

AE adverse event. i, x, 11

BSI bloodstream infection. 9, 24, 25

CoSeM Computational Sepsis Mining and Modelling. iii

EHR electronic health record. 3, 10, 15, 24, 25

IV intravenous. x, 3, 4, 9, 23

NSD Norwegian Centre for Research Data. x, 9, 11

NTNU Norwegian University of Science and Technology. iii

PIVC peripheral intravenous catheter. x, 3–5, 9, 10, 23–25

REK Regional Committees for Medical and Health Research Ethics. x, 9, 11

RQ research question. x, xiii, 5, 6, 15

SPARQL SPARQL Protocol and RDF Query Language. 19, 24

Part I

Overview

Chapter 1
Introduction

The electronic health record (EHR) is a repository of patient data generated over time by healthcare services when delivering care [17, 23]. Information in the EHR is often documented in a structured form or in free-text form. The structured form includes age, demographics, vital signs, laboratory data, and medical diagnosis codes. In contrast, the free-text form includes clinical text such as progress notes, chief complaints, discharge summaries, and adverse event reports. Clinical text is used for communication and in clinical decision-making to plan treatments, document care provided, and assess patient outcomes. Clinicians best express medical knowledge using natural language [25]; writing clinical text allows clinicians to convey and express more information with each other than structured coded data [7, 38]. Many machine learning and natural language processing studies have found clinical text is a valuable data source [15, 37, 41, 27]. However, extracting information from clinical text to answer clinical research questions can be challenging.

A semantically annotated clinical corpus is needed to accurately extract information from clinical text for answering clinical research questions [27, 36, 30]. Unfortunately, available annotated clinical corpora are limited due to privacy and ethical concerns [32, 29, 13]. Using available biomedical corpora is inadequate, as biomedical text differs from clinical text, which is usually brief and contains misspellings, grammatical errors, and abbreviations [27, 33]. If publicly available annotated clinical corpora lack the annotations needed to answer the clinical research question, curating an annotated clinical corpus may be necessary. To capture and represent concepts documented in clinical text, many studies focus on semantic annotation using annotation guidelines [46, 40, 39, 47, 13, 34] or ontologies [32, 4, 5, 45]. Annotation guidelines instruct annotators on how concepts should be annotated. Ontologies represent and model domain knowledge into concepts for reasoning. Both annotation guidelines and ontologies are developed for specific purposes and are driven by intentional design decisions. Using annotation guidelines or ontologies, annotators can annotate clinical text to capture and represent concepts that can answer clinical research questions.

This work focuses on curating an annotated clinical corpus and developing the corresponding clinical knowledge model for a use case on catheter-related signs and events. A catheter is a medical device that can be inserted into the body. Catheters have many different types, and they serve different purposes. For example, urinary catheters help drain urine, central venous catheters can provide long-term medication, and peripheral intravenous catheters (PIVCs) administer intravenous (IV) fluid and medications. Identifying documented catheter-related signs and

events can improve patient well-being by decreasing catheter-related incidents like infections.

1.1 Catheter Use Case

A catheter is a medical device that can be inserted into the body. When a patient receives a catheter, there are observable signs that the patient has a catheter. Additionally, there is the occurrence of an observable event where a clinician has inserted the catheter into the patient. During the time from catheter insertion until removal, clinicians will provide follow-up care to assess the patient and ensure the patient's catheter is working properly and in good condition. Observed signs and events can be documented in clinical text throughout this process. The documented clinical text can have multiple purposes, such as describing the patient's condition, detailing care provided, determining treatment plans, communicating with other clinicians, or reporting incidents that have or could have harmed a patient.

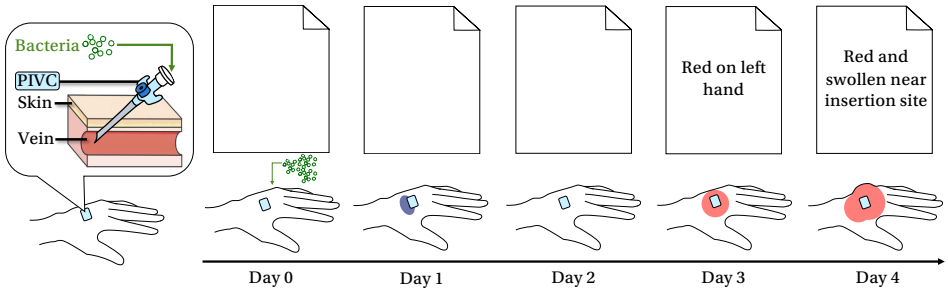


Figure 1.1: Clinical scenario of a peripheral intravenous catheter (PIVC) developing adverse signs and how it is documented. A patient admitted to the hospital receives a blue PIVC on the back of the left hand for IV medication. On day 0, the PIVC is inserted into the vein. Although bacteria have entered the body via the PIVC, it is not visible to the human eye. No PIVC information is documented on day 0. A slight purple bruise appears around the PIVC, but this can be normal after insertion. So, day 1 has no documentation about the bruise and PIVC. On day 2, the PIVC appears normal, so the clinician makes no PIVC documentation. While caring for the patient on day 3, the clinician observes redness on the left hand and documents it at the end of the work shift. The following day, the clinician observes the hand is still red and has swollen, so the adverse signs are documented as "red and swollen near insertion site".

There are differences between reality, observations, and clinical text. In reality, many events occur in the patient's body at the same time, but not all events exhibit observable signs that can be documented. For example, PIVC are a type of catheter used for administering IV fluids, IV medication, and blood transfusions. If a PIVC is well cared for and does not exhibit adverse signs, it is unlikely to be documented because the use of PIVCs among patients is common. If bacteria are introduced into the patient's body by a catheter, the bacteria are not visible to the human eye and

cannot be documented. However, the bacteria can develop into observable signs such as redness and swelling near the catheter insertion site, fever, increased heart rate, and organ failure. The clinician can document those observable signs. This example demonstrates there are differences between what occurs in reality versus what is observed and documented in clinical text. A visual example is provided in Figure 1.1.

Identifying documented catheter-related signs and events is challenging. Urinary catheter documentation lacks catheter insertion rationale [12, 16], insertion procedure [26, 10], days of catheter usage [44], and signs such as urine output, sepsis, and comfort [26]. For central venous catheters, there can be documentation errors [42] and missing documentation such as insertion location, number of lumens, and number of needle passes [8]. In clinical text, such as progress notes [48, 35, 28] and emergency department records [14], PIVC documentation often lacks the insertion site, insertion date, and follow-up site assessment care. This inadequate catheter documentation makes it challenging to identify catheter-related signs and events.

1.2 Objective and Research Questions

The objective of this work is to systematically capture catheter-related signs and events from clinical text. The objective can be divided into the following four research questions (RQs):

RQ1: What methods utilize clinical text to reduce sepsis and infections?

RQ2: What characteristics of catheter-related signs and events can be captured from clinical text?

RQ3: How can an incremental annotation-based method be developed to extract information about catheter-related signs and events from clinical text?

RQ4: How can clinical knowledge about catheter-related signs and events be captured?

1.3 Publications and Contributions

Research questions are addressed by four publications and three main areas of contribution. Additional details with specific contributions for each publication are in Chapter 3. The four publications are as follows:

Paper A - Systematic Literature Review: Sepsis prediction, early detection, and identification using clinical text for machine learning: a systematic review

Paper B - Dataset Potential: Preliminary Processing and Analysis of an Adverse Event Dataset for Detecting Sepsis-Related Events

Paper C - Annotation Method: Method for Designing Semantic Annotation of Sepsis Signs in Clinical Text

Paper D - Terminology and Ontology: Terminology and ontology development for semantic annotation: A use case on sepsis and adverse events

Contributions can be grouped into the following 3 main areas:

1. **Identify Gaps**: Identifying current research gaps in studies using clinical text for sepsis.

2. **Annotated Corpus**: Systematically developing an annotated clinical corpus and sharing the experience.

3. **Knowledge Model**: Developing an ontology corresponding to an annotated clinical corpus.

Relationships between the publications, research questions (RQs), and main contribution areas are provided in Table 1.1 and shown in Figure 1.2.

Table 1.1: Relationship between publications, research questions (RQs) and main contribution areas.

	Research Questions				Contributions		
Papers	RQ1	RQ2	RQ3	RQ4	Identify Gaps	Annotated Corpus	Knowledge Model
Paper A - Systematic Literature Review	•				•		
Paper B - Dataset Potential		•				•	
Paper C - Annotation Method		•	•			•	
Paper D - Terminology and Ontology				•			•

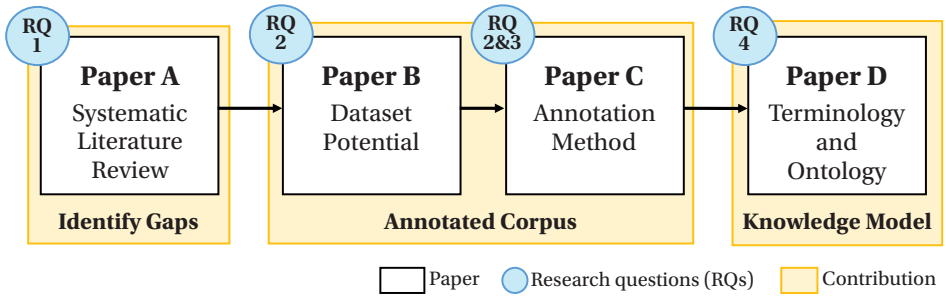


Figure 1.2: Publications, research questions (RQs), and main contribution areas.

1.4 Outline

This work is divided into two parts and outlined as follows:

- **Part I** provides an overview of the research process, results, and future work. In particular, Chapter 2 describes the research process, Chapter 3 summarizes the research and identifies specific contributions, Chapter 4 discusses the results and limitations, and Chapter 5 concludes with final remarks and future work.
- **Part II** contains the four publications included in this work.

Chapter 2
Research Process

This chapter provides insight into the clinical motivation, the ethical approval, the change in research focus, and the timeline for this work.

2.1 Clinical Motivation and Ethical Approval

The catheter use case in this work is based on a PIVCs use case provided by collaborating clinicians. All publications included within this work encompass different catheters, even though the focus has been on PIVCs. To identify PIVC-related infections, clinicians need to determine if there is a PIVC, if there is an infection, and if the two are related. Clinicians reason about different catheters and catheter-related signs and events to determine if there is a PIVC. Similarly, reasoning about different signs and events helps clinicians determine if there is an infection. Thus, clinical concepts and knowledge of signs and events for different catheters are also included.

The use case provided by clinicians Ms. Lise Husby Høvik (RN) and Dr. Lise Tuset Gustad (RN) is as follows:

> Among all catheters, PIVCs are the most regularly used invasive devices worldwide [3]. Each year, over one billion PIVCs are used [2], and approximately 80% of patients admitted to hospitals will receive at least 1 [3]. If improperly cared for, PIVCs can lead to phlebitis, infection, or sepsis. Regardless of infectious, mechanical, or chemical vein inflammation, PIVC phlebitis can be observed as pain, redness, and swelling near the PIVC insertion site [31, 20]. In cases of infectious phlebitis, bacteria could come from the skin via the insertion site, the contaminated IV solution, a contaminated catheter part, or from bacteria circulating the bloodstream attaching to the catheter [54]. That bacteria can spread through the bloodstream and become a bloodstream infection (BSI), which could lead to a life-threatening syndrome named sepsis [21]. Although routinely used, PIVCs are poorly documented in medical records [3]. This lack of documentation makes identifying and lowering catheter-related incidents like phlebitis, infection, and sepsis difficult.

The Regional Committees for Medical and Health Research Ethics (REK) has granted ethical approval to use medical data (REK approval no. 26814; 2018/1201/REKmidt). To ensure annotators are protected, the Norwegian Centre for Research Data (NSD) has granted approval to collect and process personal annotator data (NSD reference no. 142683). Furthermore, the annotators themselves have provided consent to use

their specified personal information (i.e., profession and years of experience) and their annotations.

2.2 Change in Research Focus

In research, unexpected events can lead to new opportunities. The original objective of this research was to capture sepsis signs from the medical record. This objective was divided into two goals. The primary goal was to characterize the progression of sepsis developing in patients. The secondary goal was to determine which patients have a higher risk of getting sepsis. Due to delays in obtaining medical records, adverse event reports were utilized as an alternative data source for clinical text.

An adverse event report contains a free-text adverse event note detailing a specific event where an incident has or could have harmed a patient. Adverse event notes differ from nursing or physician progress notes in the EHR. Nursing and physician progress notes are written daily by either a nurse or physician. They are for documenting the care provided and patient progress. In contrast, adverse event notes are documented separately outside the daily routine to report incidents. This can include procedural errors, hospital-acquired infections, and falls [19]. All hospital departments can report adverse events. Thus, the scope of adverse event notes is much greater than medical records.

Compared to progress notes, adverse event notes document PIVC-related adverse events more frequently. This increases the ability to identify potential signs that could develop into sepsis. Therefore, the focus was shifted from sepsis progression in medical records to catheter-related signs and events capable of leading to sepsis in adverse event notes.

2.3 Timeline

Work can be summarized into three main parts of the project timeline. The three parts are the preliminary work, the semantic annotation design process, and the annotation and guideline development. This resulted in four papers for publication, which are included. An overview of the publication timeline can be seen in Figure 2.1.

In the project timeline, preliminary work focused on ethical approval, obtaining data, setting up a server for annotations, and recruiting annotators. The semantic annotation design process includes events leading up to annotation and guideline development. The majority of 2019 was spent understanding the clinical problem and context from the perspective of the clinicians. Holding weekly meetings helped in understanding the topic and importance of the clinical problem. This led to the start of the semantic annotation design process. In 2020, obtaining adverse event data and recruiting annotators enabled the semantic annotation design process to continue. This resulted in the completion of two annotation sessions and the start of the third session. Each annotation session includes evaluation and guideline

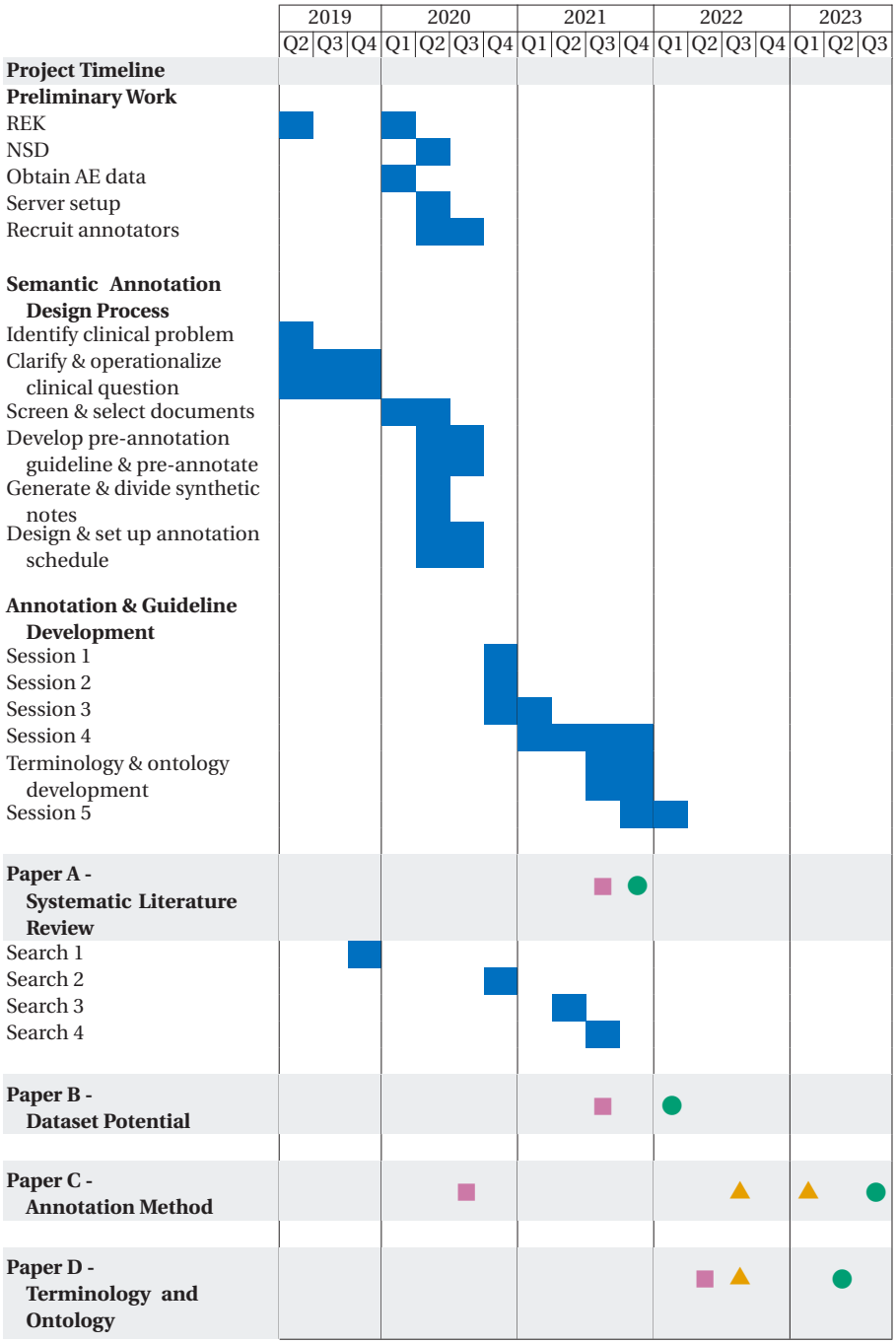


Figure 2.1: Project and publication timeline. Project-related tasks are in blue. For papers, ■ is submitted, ▲ is revised, and ● is accepted. REK: Regional Committees for Medical and Health Research Ethics; NSD: Norwegian Centre for Research Data; AE: adverse event

revision for the next session. The fourth session occurred in 2021 and overlaps with the terminology and ontology development. Annotations for this work concluded with the fifth annotation session in 2022.

Of the four publications, Paper A is a systematic literature review performed alongside the project timeline to identify research gaps within the field. Paper B focuses on the adverse event dataset and developing the annotated synthetic corpus, while Paper C focuses on detailing the semantic annotation design process. Lastly, Paper D describes the terminology and ontology development process and results.

Chapter 3
Results

This chapter provides the research questions (RQs) and main contribution areas for each paper. Additionally, specific results and contributions from each paper are detailed.

3.1 Paper A - Systematic Literature Review

Sepsis prediction, early detection, and identification using clinical text for machine learning: a systematic review

Melissa Y. Yan, Lise Tuset Gustad, Øystein Nytrø

Published in: *Journal of the American Medical Informatics Association.* Vol. 29, Issue. 3 (2022), pp. 559–575. DOI: 10.1093/jamia/ocab236.

Research Question: RQ1

Main Area of Contribution: Identify Gaps

Many studies use machine learning for prediction, early detection, and identification of sepsis. However, no literature reviews focus on utilizing clinical text for the same purpose. Paper A is a systematic literature review that identifies the methods currently utilizing clinical text to reduce sepsis through prediction, early detection, and identification. Additionally, it provides insight into using clinical data. This includes different documents in the EHR and an overview of how there is a delay between a patient's actual state and the documented data used to create models (Figure 3.1). Furthermore, it shows data selection using different windows to obtain longitudinal data (Figure 3.2). By identifying current studies, it becomes possible to understand the need and importance of utilizing text, while also identifying current research gaps. Research gaps include: (1) lack of sepsis studies using clinical text, (2) limited generalizability for hospital departments outside the intensive care units and emergency departments, (3) focus on fixed time frames within a patient's medical history, and (4) the impact of the sepsis definition used. Although the included studies are heterogeneous, combining clinical text and structured data together improves early detection and identification of sepsis and infections. Predicting sepsis 48 - 12 hours before onset appears to rely more on clinical text than structured data; this is promising for studies interested in using a patient's complete EHR to identify infection signs leading up to sepsis before a patient is in critical condition.

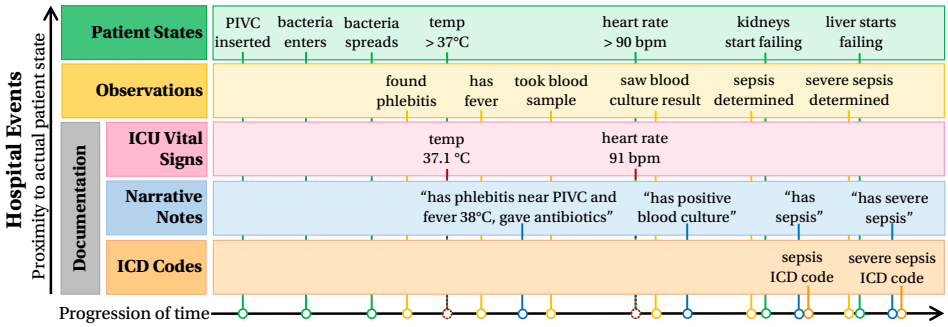


Figure 3.1: Delays between a patient's actual state, clinician observations, and documentation. Figure adapted from [51].

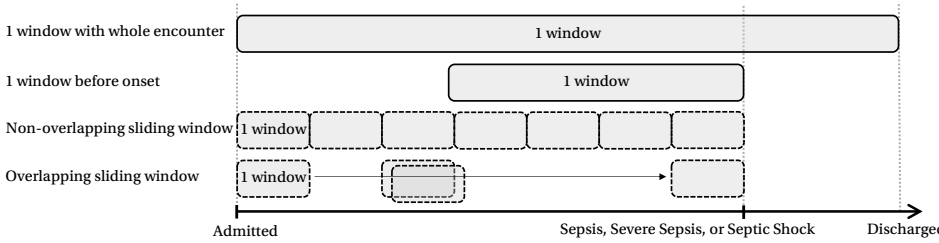


Figure 3.2: How and where to obtain longitudinal clinical data for models using different windows. Figure adapted from [51].

3.2 Paper B - Dataset Potential

Preliminary Processing and Analysis of an Adverse Event Dataset for Detecting Sepsis-Related Events

Melissa Y. Yan, Lise Husby Høvik, André Pedersen, Lise Tuset Gustad, and Øystein Nytrø.

Published in: *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. (2021), pp. 1605–1610. DOI: 10.1109/BIBM52615.2021.9669410.

Research Question: RQ2

Main Area of Contribution: Annotated Corpus

Understanding the dataset and its potential guides research. This paper gives a perspective of adverse events and current available databases. It introduces a Norwegian clinical adverse event dataset and the developed annotated synthetic version of the adverse event dataset. Additionally, it provides insight into the health care policies and purpose of Norwegian clinical adverse events. Further, preliminary results demonstrate the research potential of the dataset.

3.3 Paper C - Annotation Method

Method for Designing Semantic Annotation of Sepsis Signs in Clinical Text

Melissa Y. Yan, Lise Tuset Gustad, Lise Husby Høvik, and Øystein Nytrø.

Published in: *Proceedings of the 5th Clinical Natural Language Processing Workshop (ClinicalNLP@ACL).* (2023), pp. 236–246. DOI: 10.18653/v1/2023.clinicalnlp-1.29.

Research Question: RQ2 and RQ3

Main Area of Contribution: Annotated Corpus

The text must be annotated to capture catheter-related signs and events from clinical text needed to answer a clinical research question. Paper C describes the annotation guideline design process in detail, and illustrates the systematic and iterative annotation process taken. As shown in Figure 3.3, the semantic annotation design process consists of seven steps: (1) identify the clinical problem, (2) clarify and operationalize the clinical research question(s), (3) screen and select documents, (4) annotate, (5) generate and divide synthetic clinical text documents, (6) design and set up the annotation schedule, and (7) the annotation process and guideline development. Both design and annotation processes are driven by the clinical research question, which determines the corpus requirements that aid in designing the annotation guideline (see Figure 3.4). The annotation guideline is then applied to clinical text to produce an annotated corpus that is evaluated by the clinical research question. Further, the experiences and challenges are described to help researchers interested in annotating a corpus for their own research.

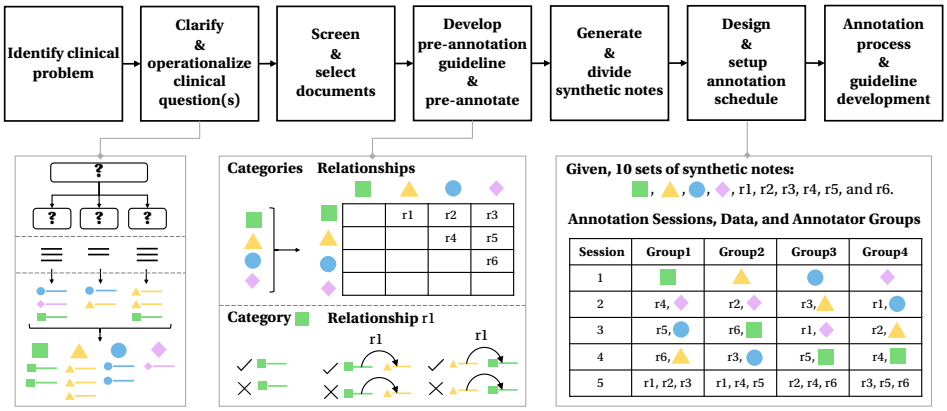


Figure 3.3: Annotation design process. Figure from [49].

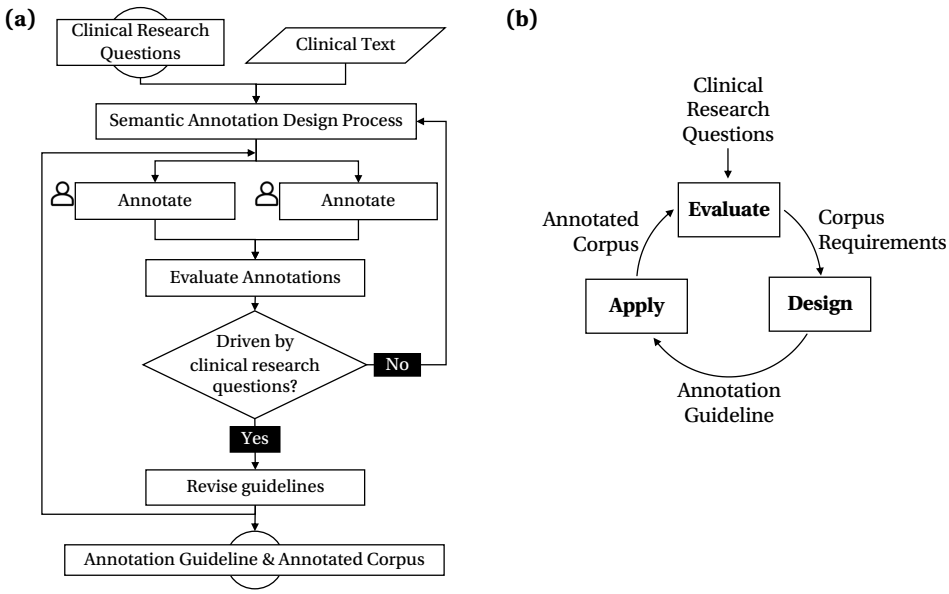


Figure 3.4: The annotation process and how it is driven by clinical research questions. **(a)** The semantic annotation design and annotation processes. **(b)** Clinical research questions drive the design and annotation processes. There is a dependency between the clinical research questions, corpus requirements, annotation guideline, and annotated corpus. Clinical research questions are evaluated to form the corpus requirements, which are used to design the annotation guideline. Then, the annotation guideline is applied to develop an annotated corpus that is evaluated by the clinical research questions again. Figure **(a)** is a revised version from [52] (© 2021, IEEE.) and **(b)** is adapted from [49].

3.4 Paper D - Terminology and Ontology

Terminology and ontology development for semantic annotation: A use case on sepsis and adverse events

Melissa Y. Yan, Lise Tuset Gustad, Lise Husby Høvik, and Øystein Nytrø.

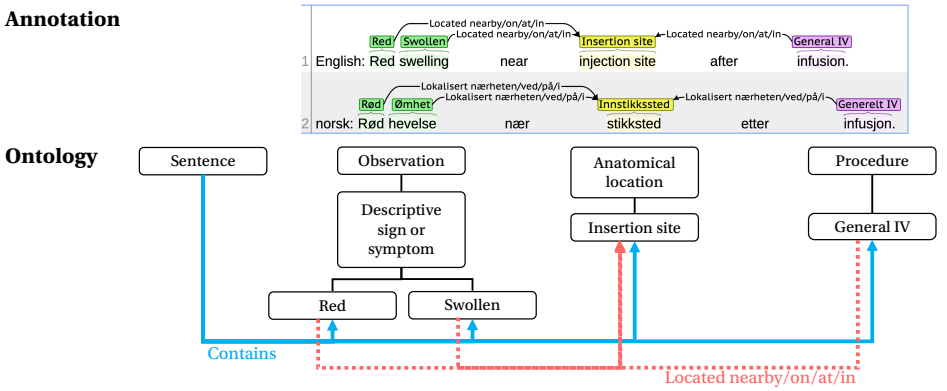
Published in: *Semantic Web*. Vol. 14, No. 5 (2023), pp. 811–871. DOI: 10.3233/SW-223226.

Research Question: RQ4

Main Area of Contribution: Knowledge Model

Building upon annotations from Paper B and Paper C, Paper D develops a terminology and ontology. The terminology indexes an annotated corpus, and the corresponding

ontology captures clinical knowledge about catheter infection indications. Terms in the terminology and ontology were simple and more general to align with terms used in clinical documentation instead of clinical guidelines. Additionally, the development process is compared against other ontology development methods. To be more accessible and understandable, all queries used in the evaluation are included as natural language for clinicians and as SPARQL Protocol and RDF Query Language (SPARQL) queries for computer scientists. Further, the terminology and ontology are released in English and Norwegian so that others can also identify and reason about catheter-related infections in a clinical adverse event corpus. Figure 3.5 shows different layers of clinical information and how annotations can be linked to an ontology for reasoning using indications.



Indications

Infusion phlebitis Indication:
Early stage of infusion phlebitis is indicated by an insertion site or infusion with 2 of the following signs:
(i) pain or tenderness,
(ii) red,
(iii) swollen or edema, or
(iv) warm.

Catheter Indication:
Any intravenous (IV) usage or infusion indicates some type of catheter is used.

Figure 3.5: Layers of clinical information and reasoning. Documented signs and events captured in annotations are linked to an ontology. The ontology represents clinical knowledge and can reason about the presence of infusion phlebitis and a catheter using indications as rules. Figure adapted from [50].

Chapter 4
Discussion

There are four perspectives when capturing clinical concepts and knowledge within clinical text for answering clinical research questions. First, an awareness of methods utilizing clinical text. Second, an understanding of documented characteristics that can be extracted. Third, the implementation of an annotation method for extracting information. Finally, the acquisition of knowledge for utilizing annotated content to answer questions. The process of developing the annotation guideline, annotated corpus, and corresponding terminology and ontology enables this integration of clinical knowledge and clinician feedback.

4.1 Research Question Findings

RQ1: What methods utilize clinical text to reduce sepsis and infections?

Methods utilizing clinical text to reduce sepsis and infection focus mainly on identification or early detection. Those methods most frequently performed word tokenization, removed tokens to improve representation, used a term frequency-inverse document frequency representation, and utilized gradient boosted trees [51]. For evaluation, the area under the receiver operating characteristic curve was reported for most.

Paper A found two studies focused on identifying infection using clinical text. However, Paper A heavily focuses on sepsis and could miss infection studies. This is because "infection" and possible sources of sepsis infection were not included in the sepsis-related search terms (i.e., "sepsis," "septic shock," and "systemic inflammatory response syndrome") [51]. Possible terms for sepsis infection sources include healthcare-acquired infection, bloodstream infection, catheter-associated infection, catheter adverse events, pneumonia, and postoperative surgical complications. A more comprehensive overview of the methods should include different sources of infection.

RQ2: What characteristics of catheter-related signs and events can be captured from clinical text?

Documented catheter-related signs and events within clinical text can either be explicit or implicit. As mentioned in Paper B and Paper C, certain catheter types are distinctly documented, whereas other catheters can be distinguished based on anatomical insertion sites or procedures [53, 49]. To capture characteristics of signs

and events related to catheters and infections in clinical text, it is essential to identify signs and symptoms, anatomical locations, medical devices, and procedures. The terms for these characteristics were kept simple because clinical documentation contains common clinical knowledge written for other clinicians and is more general than clinical guidelines (e.g., insert the central venous catheter into the "chest" versus the "jugular vein until the superior vena cava"). Additionally, relationships between identified signs must be identified to reason about the presence of catheters and catheter-related events. Therefore, catheter-related characteristics, events, and clinical knowledge must also be captured from clinical text.

RQ3: How can an incremental annotation-based method be developed to extract information about catheter-related signs and events from clinical text?

Annotating clinical text requires clinical knowledge and an understanding of the context to extract data, information, and knowledge properly. This context includes the relationships between medical facts and the thought processes involved in investigating, diagnosing, and treating medical conditions [1]. The context and undocumented information in the clinical text will affect the interpretation [6]. For example, intensive care unit progress notes do not document normal heart rates because only an abnormal heart rate requires additional monitoring. Different types of clinical text have different purposes. For example, nursing progress notes are written by nurses usually at the end of their shift to summarize the care provided and observed signs so that the next nurse taking over can continue patient care [9, 43, 22]. When discharging a patient from the hospital, discharge summaries summarize the patient's stay, follow-up treatment plans, prescriptions, and referrals [18]. In comparison, adverse events are reported by all departments outside normal work when an incident or potential mishap occurs. When annotating clinical text, it is important to consider what information can be annotated and in what context to extract information properly. Understanding how, why, and who has generated the clinical text can provide insight into what type of information can be annotated.

The development of an incremental annotation-based method is described in Paper C. Systematically annotating clinical text for catheter-related signs and events starts at the semantic annotation design process [49]. The design process involves understanding the clinical problem and determining what is documented so that categories or entities for annotation can be formulated to capture concepts. Categories are utilized to develop an annotation guideline that annotators use. During annotation, annotators use the annotation guideline and leave comments about issues related to the guideline. Combining annotation results and annotator comments makes it possible to make revision guidelines to reduce ambiguity and increase inter-annotator agreement. Additionally, annotations are evaluated to determine if annotated concepts are relevant for answering the clinical research question of interest. Then, this process is repeated iteratively and forms the incremental annotation-based method to extract information, signs, and events related to catheters.

This method for designing and annotating clinical text for a specific clinical use case can be beneficial for researchers needing to annotate a corpus. However, there are some limitations. First, the experiences are based on a specific clinical case and focus on the qualitative aspects. Details of certain parts of the design and annotation process will likely need to be adjusted based on resources available to other researchers. This can include the data selected for annotation, the number of annotators available, and the annotators' level of expertise. For instance, the use case in the design process is based on using 8 annotators to annotate 100 synthetic AE notes over 5 sessions. Second, expertise and additional time are required to generate synthetic notes for annotation. Finally, future work is still needed to replicate the described design and annotation process on other forms of clinical text and problems.

The developed incremental annotation-based method is for a specific use case and needs to be replicated on different types of clinical text and problems [49]. Depending on the resources available, the design and annotation processes will need to be adjusted. When making adjustments, considerations can include the clinical research question that should be addressed, the type of clinical text available, the ability to generate synthetic clinical text, the text selected for annotation, the number of annotators, the annotators' clinical experience, evaluation metrics for inter-annotator agreement, and the feasibility of the project timeline.

RQ4: How can clinical knowledge about catheter-related signs and events be captured?

Capturing signs and events alone is not enough. Clinicians recognize different combinations of signs or events within specific situations as indicators of unnamed catheters or infections. As previously mentioned, relationships between identified signs and events are also necessary to reason about the presence of catheters and catheter-related events. Thus, an ontology that represents clinical knowledge and corresponds with annotated clinical text is needed to reason and identify indications of catheters and infections.

The ontology represents documented clinical knowledge used by clinicians to reason about the presence of catheters in clinical adverse event reports. The list of indications in Appendix C.3 of [50] and their usage through competency questions can be found in Appendix C.4 of [50]. However, in certain situations, infusion phlebitis is not a catheter-related infection and complication but an expected side effect. For example, using the Cordarone heart medication containing Amiodarone to treat irregular heartbeat can result in severe phlebitis [11]. Use of certain antibiotics can also lead to PIVC-related phlebitis [24]. Expanding the ontology to include more specific IV medications and their expected side effects can assist in finding more catheters.

The ontology was developed with clinicians and for clinicians. The competency questions are written in natural language for the clinicians and have corresponding

SPARQL queries. Thus, with some training, clinicians should be able to maintain the ontology themselves.

4.2 Future Work

To reduce infections, hospitals need continuous and precise infection monitoring at the system level. Continuous infection surveillance strengthens the feedback loops of hospital management and clinicians. It enables rapid evaluation of clinical practice and interventions to reduce infections. Thus, it is essential that hospitals need to learn from both adverse and successful events to target infections, especially catheter and PIVC-related BSIs. Therefore, the primary focus of future work will be implementing a pipeline to systematically learn better infection prevention from adverse and successful events by automating BSI risk detection.

A potential pipeline would include the following seven steps: (1) data management and collection, (2) annotation, (3) data preprocessing, (4) machine learning model 1, (5) machine learning model 2, (6) a reasoning system, and (7) comparison of adverse and successful events. Data management and collection of EHR and adverse event data will be used for annotation and data preprocessing. Machine learning model 1 will then classify documents into venous catheter-related, infection-related, both, and none. Afterward, machine learning model 2 will perform named entity recognition and relation extraction to automatically predict labels for signs, precursors, causes, preceding events, and relationships between labels in documents. Finally, labels will be used by the reasoning system to infer PIVC-related BSIs, so it will be possible to compare adverse PIVC-related BSI events and successful events without infection. Thus, this pipeline can be divided into three projects.

The first project is annotating medical records from the EHR system. Medical records can be annotated for catheter-related signs and events by utilizing the annotation guideline generated from annotating adverse event notes. Additionally, the annotation guideline, terminology, and ontology can be expanded to include additional annotation labels or clinical knowledge. Replicating the semantic annotation process would further reinforce and validate the method described in Paper C. Further annotations may be required based on downstream analyses. An annotated medical corpus for catheter-related infections would make it possible to identify catheter-related infections and study the progression of infection developing in patients.

Using both the annotated adverse event corpus and medical record corpus, the second project focuses on classifiers and the reasoning system. The primary objective is to create classifiers that can identify catheter-related signs and events by using the annotated adverse event corpus and medical record corpus. Then, the secondary objective is to combine the classifiers with clinical knowledge in the ontology to infer catheter presence, signs, and events. In addition to identifying catheter-related signs and events, another possibility is tracking the progression of catheter-related

phlebitis, infection, or sepsis in medical records. Depending on the data, it could also be possible to create early detection or prediction models for the risk of developing catheter-related phlebitis, infection, or sepsis.

Finally, the third project is developing a tool for the pipeline to compare adverse and successful events. This tool can be a visual dashboard for monitoring, analyzing, and visualizing adverse events to assist the hospital in prioritizing tasks for patient safety improvement of catheters. Thus, future work will mainly focus on and expand upon those three projects to complete the pipeline to develop an infection surveillance tool and improve patient safety at the system level.

Identifying PIVC-related BSIs within the EHR gives a perspective of its prevalence. Additionally, it provides an opportunity to contextually analyze signs, precursors, common clinical causes, and events leading to a PIVC-related BSI at the bedside. Inversely, identifying patterns, procedures, and situations with few BSI events enables learning from successful procedures. Thus, clinicians will be more aware of PIVC-related BSIs successful and adverse events. Raising awareness of BSI events among clinicians and hospital management is possible through learning from both successful events without infection and adverse PIVC-related events. Such awareness can improve patient care and outcomes, reduce clinician distress, and minimize organizational problems. Therefore, it is promising to build upon the annotation method, annotated adverse event corpus, terminology, and ontology developed from this research for future work.

Chapter 5
Conclusion

Focused on capturing catheter-related signs and events from clinical text, this work is comprised of four publications. Those publications describe (1) a systematic literature review about using clinical text for sepsis identification and early detection, (2) the annotation design process, (3) the development of an annotated clinical corpus and its corresponding annotation guideline, and (4) the process of developing a terminology to represent annotations and an ontology to capture clinical knowledge. This has resulted in a method to annotate clinical text, an annotated corpus, an annotation guideline, a terminology, and an ontology for catheter-related signs and events.

Based on the results of this work, preparation is underway for three projects. The first project is annotating medical records using the annotation guideline from this work. Using the annotated adverse event corpus, the second project focuses on classifiers to identify catheter-related signs and events. The final project's goal is to develop a visual dashboard for monitoring, analyzing, and visualizing catheter-related adverse events from the annotated corpus to assist the hospital in prioritizing tasks for patient safety improvement. Thus, future work on these projects will work towards improving patient safety by reducing catheter-related adverse events.

Bibliography

[1] Houtan Aghili, Richard A. Mushlin, Rose M. Williams, and Jeffrey S. Rose. Progress notes model. *Proceedings: a conference of the American Medical Informatics Association. AMIA Fall Symposium*, pages 12–16, 1997.

[2] Evan Alexandrou, Gillian Ray-Barruel, Peter J. Carr, Steven Frost, Sheila Inwood, Niall Higgins, Frances Lin, Laura Alberto, Leonard Mermel, and Claire M. Rickard. International prevalence of the use of peripheral intravenous catheters. *Journal of Hospital Medicine*, 10(8):530–533, August 2015.

[3] Evan Alexandrou, Gillian Ray-Barruel, Peter J. Carr, Steven A. Frost, Sheila Inwood, Niall Higgins, Frances Lin, Laura Alberto, Leonard Mermel, and Claire M. Rickard. Use of short peripheral intravenous catheters: Characteristics, management, and outcomes worldwide. *Journal of Hospital Medicine*, 13(5), 2018.

[4] Michael Bada, Miriam Eckert, Martha Palmer, and Lawrence Hunter. An overview of the CRAFT concept annotation guidelines. In *Proceedings of the Fourth Linguistic Annotation Workshop, LAW 2010, Uppsala, Sweden, July 15-16, 2010*, pages 207–211. Association for Computational Linguistics, 2010.

[5] Michael Bada, Nicole A. Vasilevsky, William A. Baumgartner Jr., Melissa A. Haendel, and Lawrence E. Hunter. Gold-standard ontology-based anatomical annotation in the CRAFT corpus. *Database*, 2017:bax087, 2017.

[6] Marc Berg and Els Goorman. The contextual nature of medical information. *International Journal of Medical Informatics*, 56(1-3):51–60, 1999.

[7] Marc Berg, Chris Langenberg, Ignas V.D. Berg, and Jan Kwakkernaat. Considerations for sociotechnical design: experiences with an electronic patient record in a clinical context. *International Journal of Medical Informatics*, 52(1-3):243–251, 1998.

[8] Malin Björkander, Peter Bentzer, Ulf Schött, Marcus E. Broman, and Thomas Kander. Mechanical complications of central venous catheter insertions: A retrospective multicenter study of incidence and risks. *Acta Anaesthesiologica Scandinavica*, 63(1):61–68, July 2018.

[9] Wendy Blair and Barbara Smith. Nursing documentation: Frameworks and barriers. *Contemporary nurse*, 41(2):160–168, 2012.

[10] Sandy L. Bozarth, Supriya Mannepalli, Wanda G. Hester, Sabrina D. Broome, and Stephen S. Pirkle. Effect of Physician Documentation on Catheter Associate

Urinary Tract Infection Rates at a Community Hospital. *American Journal of Infection Control*, 48(8, Supplement):S23, August 2020.

[11] Gunhild Brørs, Kari Hanne Gjeilo, Tonje Lund, Karin Skevik, Elizabeth Aa, Lise Husby Høvik, Torhild Skarsvaag, and Ole Christian Mjølstad. Amiodarone-induced phlebitis: incidence and adherence to a clinical practice guideline. *European Journal of Cardiovascular Nursing*, page zvad003, 2023.

[12] Melanie De Jaeger. Exploring urinary catheters: The perspectives of patients and nurses. *British Journal of Nursing*, 20(7):400–408, 2011.

[13] Louise Deleger, Q. Qi Li, Todd Lingren, Megan Kaiser, Katalin Molnar, Laura Stoutenborough, Michal Kouril, Keith Marsolo, and Imre Solti. Building gold standard corpora for medical natural language processing tasks. *AMIA Annual Symposium proceedings. AMIA Symposium.*, 2012:144–153, November 2012.

[14] Mohamad G Fakih, Karen Jones, Janice E Rey, Robert Takla, Susanna Szpunar, Karrie Brown, Arlene Boelstler, and Louis Saravolatz. Peripheral venous catheter care in the emergency department: Education and feedback lead to marked improvements. *American Journal of Infection Control*, 41(6):531–536, June 2013.

[15] Elizabeth Ford, John A. Carroll, Helen E. Smith, Donia Scott, and Jackie Cassell. Extracting information from the text of electronic medical records to improve case detection: a systematic review. *Journal of the American Medical Informatics Association*, 23(5):1007–1015, 2016.

[16] Radha Ramana Murthy Gokula, John A. Hickner, and Mindy A. Smith. Inappropriate use of urinary catheters in elderly patients at a midwestern community teaching hospital. *American Journal of Infection Control*, 32(4):196–199, June 2004.

[17] Kristiina Häyrinen, Kaija Saranto, and Pirkko Nykänen. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. *International Journal of Medical Informatics*, 77(5):291–304, 2008.

[18] Ragnhild Hellesø. Information handling in the nursing discharge note. *Journal of Clinical Nursing*, 15(1):11–21, 2006.

[19] Peter D. Hibbert, Charlotte J. Molloy, Tamara D. Hooper, Louise K. Wiles, William B. Runciman, Peter Lachman, Stephen E. Muething, and Jeffrey Braithwaite. The application of the global trigger tool: a systematic review. *International Journal for Quality in Health Care*, 28(6):640–649, December 2016.

[20] Ray Higginson and Andrew Parry. Phlebitis: treatment, care and prevention. *Nursing Times*, 107(36):18–21, 2011.

[21] Luis E. Huerta and Todd W. Rice. Pathologic Difference between Sepsis and Bloodstream Infections. *The Journal of Applied Laboratory Medicine*, 3(4):654–663, 10 2019.

[22] M. Johnson, P. Sanchez, H. Suominen, J. Basilakis, L. Dawson, B. Kelly, and L. Hanlen. Comparing nursing handover and documentation: forming one set of patient information. *International Nursing Review*, 61(1):73–81, 2014.

[23] Ellen Kim, Samuel M. Rubinstein, Kevin T. Nead, Andrzej P. Wojcieszynski, Peter E. Gabriel, and Jeremy L. Warner. The Evolving Use of Electronic Health Records (EHR) for Research. *Seminars in Radiation Oncology*, 29(4):354–361, 2019. Big Data in Radiation Oncology.

[24] Peter Lanbeck, Inga Odenholt, and Otto Paulsen. Antibiotics differ in their tendency to cause infusion phlebitis: A prospective observational study. *Scandinavian Journal of Infectious Diseases*, 34(7):512–519, 2002.

[25] Christian Lovis, Robert H. Baud, and Pierre Planche. Power of expression in the electronic patient record: Structured data or narrative text? *International Journal of Medical Informatics*, 58:101–110, 2000.

[26] C. McAteer, R. Sullivan, C. McRory, O. O'Domhnaill, D. E. Murphy, M. A. Rehman, T. Muller, M. Gallagher, J. Miranda, V. Parihar, and K. Mulpeter. 95 DOCUMENTATION OF URINARY CATHETER INDICATION AND INSERTION PROCEDURE FOR MEDICAL INPATIENTS AT A UNIVERSITY TEACHING HOSPITAL. *Age and Ageing*, 50(Supplement_3):ii9–ii41, November 2021.

[27] S. M. Meystre, G. K. Savova, K. C. Kipper-Schuler, and J. F. Hurdle. Extracting information from textual documents in the electronic health record: a review of recent research. *Yearbook of Medical Informatics*, pages 128–44, 2008.

[28] Karen A. New, Joan Webster, Nicole M. Marsh, and Barbara Hewer. Intravascular device use, management, documentation and complications: A point prevalence survey. *Australian Health Review*, 38(3):345–349, June 2014.

[29] Maite Oronoz, Koldo Gojenola, Alicia Pérez, Arantza Díaz de Ilarraza, and Arantza Casillas. On the creation of a clinical gold standard corpus in Spanish: Mining adverse drug reactions. *Journal of Biomedical Informatics*, 56:318–32, 2015.

[30] Pinal Patel, Disha Davey, Vishal Panchal, and Parth Pathak. Annotation of a Large Clinical Entity Corpus. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, pages 2033–2042. Association for Computational Linguistics, 2018.

[31] Gillian Ray-Barruel, Denise F. Polit, Jenny E. Murfield, and Claire M. Rickard. Infusion phlebitis assessment measures: A systematic review. *Journal of Evaluation in Clinical Practice*, 20(2):191–202, April 2014.

[32] Angus Roberts, Robert Gaizauskas, Mark Hepple, George Demetriou, Yikun Guo, Ian Roberts, and Andrea Setzer. Building a semantically annotated corpus of clinical texts. *Journal of Biomedical Informatics*, 42 5:950–66, 2009.

[33] Thomas Brox Røst, Ola Huseth, Øystein Nytrø, and Anders Grimsmo. Lessons from Developing an Annotated Corpus of Patient Histories. *Journal of Computing Science and Engineering*, 2(2):162–179, 2008.

[34] Thomas Brox Røst, Christine Raaen Tvedt, Haldor Husby, Ingrid Andås Berg, and Øystein Nytrø. Capturing Central Venous Catheterization Events in Health Record Texts. In *IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2018, Madrid, Spain, December 3-6, 2018*, pages 488–495. IEEE Computer Society, 2018.

[35] Emily Russell, Raymond Javan Chan, Nicole Marsh, and Karen New. A point prevalence study of cancer nursing practices for managing intravascular devices in an Australian tertiary cancer center. *European Journal of Oncology Nursing*, 18(3):231–235, June 2014.

[36] Brett R. South, Shuying Shen, Makoto Jones, Jennifer Garvin, Matthew H. Samore, Wendy W. Chapman, and Adi V. Gundlapalli. Developing a manually annotated clinical document corpus to identify phenotypic information for inflammatory bowel disease. *BMC Bioinformatics*, 10(Suppl 9):S12 – S12, 2009.

[37] Irena Spasic and Goran Nenadic. Clinical Text Data in Machine Learning: Systematic Review. *JMIR Medical Informatics*, 8(3):e17984, 2020.

[38] H. David Stein, Prakash Nadkarni, Joseph Erdos, and Perry L. Miller. Exploring the Degree of Concordance of Coded and Textual Data in Answering Clinical Queries from a Clinical Data Repository. *Journal of the American Medical Informatics Association*, 7(1):42–54, January 2000.

[39] Amber Stubbs and Özlem Uzuner. Annotating longitudinal clinical narratives for de-identification: The 2014 i2b2/UTHealth corpus. *Journal of Biomedical Informatics*, 58(Suppl):S20–S29, December 2015.

[40] Weiyi Sun, Anna Rumshisky, and Ozlem Uzuner. Annotating temporal information in clinical narratives. *Journal of Biomedical Informatics*, 46(Suppl):S5–S12, December 2013.

[41] Maryam Tayefi, Phuong Ngo, Taridzo Chomutare, Hercules Dalianis, Elisa Salvi, Andrius Budrionis, and Fred Godtliebsen. Challenges and opportunities beyond structured data in analysis of electronic health records. *Wiley Interdisciplinary Reviews: Computational Statistics*, 13(6):e1549, 2021.

[42] Sheri Chernetsky Tejedor, Gina Garrett, Jesse T. Jacob, Ellen Meyer, Mary Dent Reyes, Chad Robichaux, and James P. Steinberg. Electronic Documentation of Central Venous Catheter—Days: Validation Is Essential. *Infection Control & Hospital Epidemiology*, 34:899–907, 2013.

[43] Marion Tower, Wendy Chaboyer, Quentine Green, Kirsten Dyer, and Marianne Wallis. Registered nurses' decision-making regarding documentation in patients' progress notes. *Journal of Clinical Nursing*, 21(19-20):2917–2929, 2012.

[44] Barbara W. Trautner, Jan E. Patterson, Nancy J. Petersen, Sylvia Hysong, Deborah Horwitz, G. John Chen, Patti Grota, and Aanand D. Naik. Quality Gaps in Documenting Urinary Catheter Use and Infectious Outcomes. *Infection Control & Hospital Epidemiology*, 34(8):793–799, August 2013.

[45] Ravensara S. Travillian, Tomasz Adamusiak, Tony Burdett, Michael Gruenberger, John Hancock, Ann-Marie Mallon, James Malone, Paul Schofield, and Helen Parkinson. Anatomy ontologies and potential users: Bridging the gap. *Journal of Biomedical Semantics*, 2(Suppl 4):S3, August 2011.

[46] Özlem Uzuner, Imre Solti, Fei Xia, and Eithon Cadag. Community annotation experiment for ground truth generation for the i2b2 medication challenge. *Journal of the American Medical Informatics Association*, 17(5):519–523, September 2010.

[47] Fei Xia and Meliha Yetisgen-Yildiz. Clinical Corpus Annotation: Challenges and Strategies. In *Proceedings of the Third Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM'2012) in conjunction with the International Conference on Language Resources and Evaluation (LREC 2012)*, pages 32–39, 2012.

[48] Lokesh Yagnik, Angela Graves, and Ken Thong. Plastic in patient study: Prospective audit of adherence to peripheral intravenous cannula monitoring and documentation guidelines, with the aim of reducing future rates of intravenous cannula-related complications. *American Journal of Infection Control*, 45(1):34–38, January 2017.

[49] Melissa Y. Yan, Lise Tuset Gustad, Lise Husby Høvik, and Øystein Nytrø. Method for designing semantic annotation of sepsis signs in clinical text. In *Proceedings of the 5th Clinical Natural Language Processing Workshop, ClinicalNLP@ACL 2023, Toronto, Canada, July 14, 2023*, pages 236–246. Association for Computational Linguistics, 2023.

[50] Melissa Y. Yan, Lise Tuset Gustad, Lise Husby Høvik, and Øystein Nytrø. Terminology and Ontology Development for Semantic Annotation: A Use Case on Sepsis and Adverse Events. *Semantic Web*, 14(5):811–871, May 2023.

[51] Melissa Y. Yan, Lise Tuset Gustad, and Øystein Nytrø. Sepsis prediction, early detection, and identification using clinical text for machine learning: a systematic review. *Journal of the American Medical Informatics Association*, 29(3):559–575, January 2022.

[52] Melissa Y. Yan, Lise Husby Høvik, Lise Tuset Gustad, and Øystein Nytrø. Understanding and Reasoning About Early Signs of Sepsis: From Annotation Guideline to Ontology. In *IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2021, Houston, TX, USA, December 9-12, 2021*, pages 1906–1911. IEEE, 2021.

[53] Melissa Y. Yan, Lise Husby Høvik, André Pedersen, Lise Tuset Gustad, and Øystein Nytrø. Preliminary Processing and Analysis of an Adverse Event Dataset for Detecting Sepsis-Related Events. In *IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2021, Houston, TX, USA, December 9-12, 2021*, pages 1605–1610. IEEE, 2021.

[54] Li Zhang, Siyu Cao, Nicole Marsh, Gillian Ray-Barruel, Julie Flynn, Emily Larsen, and Claire M. Rickard. Infection risks associated with peripheral vascular catheters. *Journal of Infection Prevention*, 17(5):207–213, September 2016.

Part II

Publications

A

Paper A

Sepsis prediction, early detection, and identification using clinical text for machine learning: a systematic review

Melissa Y. Yan, Lise Tuset Gustad, and Øystein Nytrø

Published in:
Journal of the American Medical Informatics Association, March 2022, volume 29, issue 3, pp. 559–575. DOI: 10.1093/jamia/ocab236.

Journal of the American Medical Informatics Association, 29(3), 2022, 559–575
https://doi.org/10.1093/jamia/ocab236
Advance Access Publication Date: 13 December 2021
Review

Review

Sepsis prediction, early detection, and identification using clinical text for machine learning: a systematic review

Melissa Y. Yan [1], Lise Tuset Gustad [2,3], and Øystein Nytrø [1]

[1]Department of Computer Science, Faculty of Information Technology and Electrical Engineering, Norwegian University of Science and Technology, Trondheim, Norway, [2]Department of Circulation and Medical Imaging, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway, and [3]Department of Medicine, Levanger Hospital, Clinic of Medicine and Rehabilitation, Nord-Trøndelag Hospital Trust, Levanger, Norway

Corresponding Author: Melissa Y. Yan, MBI, Department of Computer Science, Faculty of Information Technology and Electrical Engineering, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway; melissa.yan@ntnu.no

Received 4 July 2021; Revised 11 September 2021; Editorial Decision 29 September 2021; Accepted 11 October 2021

ABSTRACT

Objective: To determine the effects of using unstructured clinical text in machine learning (ML) for prediction, early detection, and identification of sepsis.
Materials and methods: PubMed, Scopus, ACM DL, dblp, and IEEE Xplore databases were searched. Articles utilizing clinical text for ML or natural language processing (NLP) to detect, identify, recognize, diagnose, or predict the onset, development, progress, or prognosis of systemic inflammatory response syndrome, sepsis, severe sepsis, or septic shock were included. Sepsis definition, dataset, types of data, ML models, NLP techniques, and evaluation metrics were extracted.
Results: The clinical text used in models include narrative notes written by nurses, physicians, and specialists in varying situations. This is often combined with common structured data such as demographics, vital signs, laboratory data, and medications. Area under the receiver operating characteristic curve (AUC) comparison of ML methods showed that utilizing both text and structured data predicts sepsis earlier and more accurately than structured data alone. No meta-analysis was performed because of incomparable measurements among the 9 included studies.
Discussion: Studies focused on sepsis identification or early detection before onset; no studies used patient histories beyond the current episode of care to predict sepsis. Sepsis definition affects reporting methods, outcomes, and results. Many methods rely on continuous vital sign measurements in intensive care, making them not easily transferable to general ward units.
Conclusions: Approaches were heterogeneous, but studies showed that utilizing both unstructured text and structured data in ML can improve identification and early detection of sepsis.

Key words: sepsis, natural language processing, machine learning, electronic health records, systematic review

INTRODUCTION

Sepsis is a life-threatening illness caused by the body's immune response to an infection that leads to multi-organ failure.[1] Annually, there are 31.5 million sepsis cases, 19.4 million severe sepsis cases, and 5.3 million sepsis deaths estimated in high-income countries.[2] Studies have shown that early identification of sepsis following rapid initiation of antibiotic treatment improves patient outcomes,[3] and 6 h of treatment delay is shown to increase the mortality risk by 7.6%.[4] Unfortunately, sepsis is commonly misdiagnosed and mistreated because deterioration with organ failure is also common in

other diseases.[5–8] The heterogeneity in infection source, immune responses, and pathophysiological changes make identification and therefore sepsis treatment difficult. Additionally, the diversity in age, gender, and comorbidities affect the symptoms and outcome of septic patients.[7]

Machine learning (ML) has been employed to improve sepsis outcomes through early detection. ML can utilize structured and unstructured data from electronic health records (EHRs).[9–14] Structured clinical data come in a fixed format, such as age, vital signs, and laboratory data, which make data preprocessing easier. In contrast, clinical notes are in unstructured free-text form, such as progress notes, nursing notes, chief complaints, or discharge summaries. Clinical notes contain abbreviations, grammatical errors, and misspellings. Using clinical text is a complex, time-consuming process because it requires using natural language processing (NLP) to extract features that transform text into a machine-understandable representation.[15–22] This usually requires assistance from clinical experts to convert text into machine-interpretable representations that capture clinical knowledge for specific clinical domains. The effort required to utilize unstructured clinical text can deter researchers; however, unstructured clinical text contains valuable information.[16,22–25] Multiple studies and a review[25] have shown that using unstructured clinical text has increased model performance to detect or predict colorectal surgical complications,[26] postoperative acute respiratory failure,[27] breast cancer,[28] pancreatic cancer,[29] fatty liver disease,[30] pneumonia,[31] inflammatory bowel disease,[32,33] rheumatoid arthritis,[34–36] multiple sclerosis,[37] and acute respiratory infection.[38,39]

Prior reviews related to sepsis detection and prediction include: sepsis detection using Systemic Inflammatory Response Syndrome (SIRS) screening tools,[40] sepsis detection using SIRS and organ dysfunction criteria with EHR vital signs and laboratory data,[41] clinical perspectives on the use of ML for early detection of sepsis in daily practice,[14] ML for diagnosis and early detection of sepsis patients,[9–13] infectious disease clinical decision support,[42] and healthcare-associated infections mentioning sepsis.[43–45] However, to the best of our knowledge, no reviews focus on the effect of utilizing unstructured clinical text for sepsis prediction, early detection, or identification; this makes it challenging to assess and utilize text in future ML and NLP sepsis research.

OBJECTIVE

The review aims to gain an overview of studies utilizing clinical text in ML for sepsis prediction, early detection, or identification.

MATERIALS AND METHODS

This systematic review follows the Preferred Reporting Items for Systematic review and Meta-Analyses guidelines.[46]

Search strategy

Relevant articles were identified from 2 clinical databases (PubMed and Scopus) and 3 computer science databases (ACM DL, dblp, and IEEE Xplore) using defined search terms. The 3 sets of search terms included: (1) "sepsis," "septic shock," or "systemic inflammatory response syndrome"; (2) "natural language processing," "machine learning," "artificial intelligence," "unstructured data," "unstructured text," "clinical note," "clinical notes," "clinical text," "free-text," "free text," "record text," "narrative," or "narratives"; and (3) detect, identify, recognize, diagnosis, predict, prognosis, progress, develop, or onset. Searches on clinical databases were performed using all 3 sets of search terms and excluded animal-related terms. Whereas searches on computer science databases only used the first set of search terms. No additional search restrictions, such as date, language, and publication status, were included. Additional articles were identified from relevant review articles or backward reference and forward citation searches of eligible articles. Complete search strategies are in Supplementary Table S1.

The search was initially conducted using only computer science databases on December 10, 2019 and was updated to include clinical databases on December 14, 2020. The first search found that 4 of 454 articles met inclusion criteria,[47–50] and the second search uncovered 2 more articles that met inclusion criteria (6 of 1335 articles).[51,52] Those 2 searches did not contain the search terms: "systemic inflammatory response syndrome," "artificial intelligence," identify, recognize, diagnosis, prognosis, progress, develop, and onset. Hence, a search on May 15, 2021, including those terms, found 2 additional articles.[53,54] To ensure inclusion of other relevant articles, a broader search was conducted on September 3, 2021 to include the following terms: "unstructured data," "unstructured text," "clinical note," "clinical notes," "clinical text," "free-text," "free text," "record text," "narrative," or "narratives." This resulted in 1 additional article.[55]

Study selection

Titles, abstracts, and keywords were screened using Zotero v5.0.96.3 (Corporation for Digital Scholarship, Vienna, VA) and Paperpile (Paperpile LLC, Cambridge, MA). Screening removed duplicates and articles that did not contain the following terms: (1) text, (2) notes, or (3) unstructured. Full-text articles were evaluated to determine if the study used unstructured clinical text for the identification, early detection, or prediction of sepsis onset in ML. Thus, selected articles had to rely on methods that automatically improve based on what they learn and not rely solely on human-curated rules. Additionally, articles solely focusing on predicting sepsis mortality were excluded as these articles are based on already established sepsis cases. Reviews, abstract-only articles, and presentations were removed. Additionally, a backward and forward search was performed on eligible full-text articles.

Data extraction

One author independently extracted data, which a second author verified. Any discrepancies were resolved either through discussion with the third author by assessing and comparing data to evidence from the studies or by directly communicating with authors from included articles. The following information was extracted: (1) general study information including authors and publication year, (2) data source, (3) sample size, (4) clinical setting, (5) sepsis infection definition, (6) task and objective, (7) characteristics of structured and unstructured data, (8) underlying ML and NLP techniques, and (9) evaluation metrics.

RESULTS

Selection process

The initial search identified 2268 articles from 5 databases and 5 additional articles[56–60] from 2 relevant review articles (Figure 1).[43,44] From the 1817 unique articles, 1620 articles were excluded based on eligibility criteria described in the methods. After assessing the

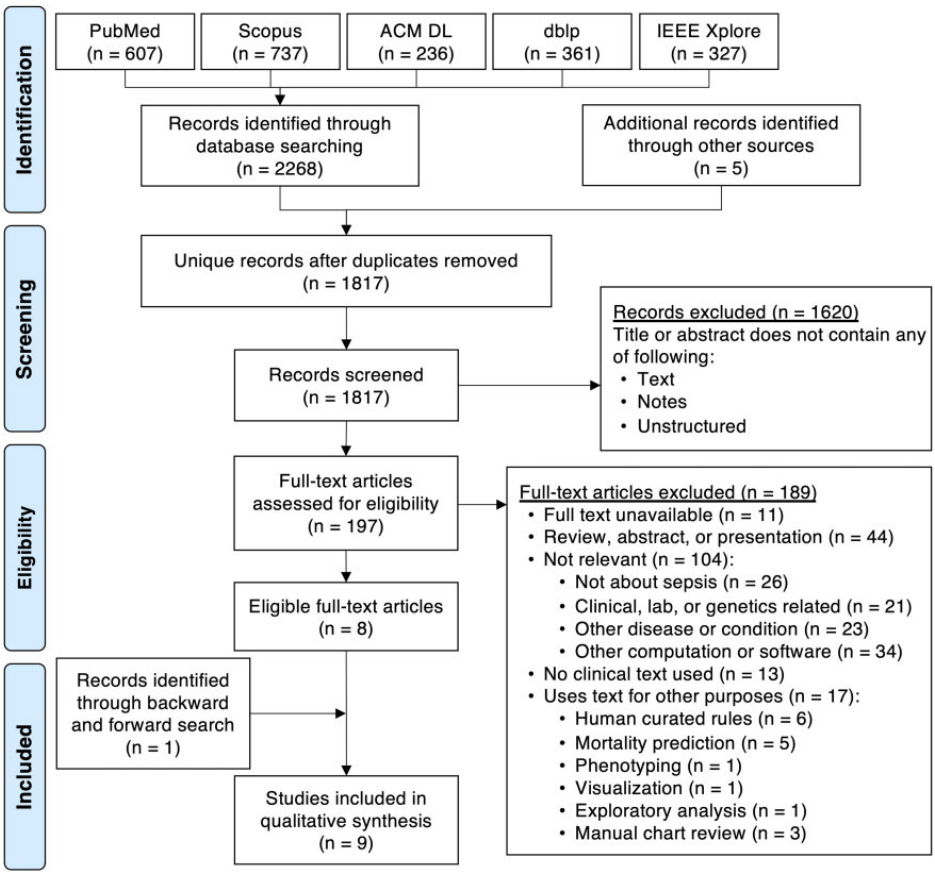


Figure 1. PRISMA (Preferred Reporting Items for Systemic reviews and Meta-Analyses) flowchart for study selection.

remaining 197 articles, most studies (189 of 197, ie, 96%) were excluded because they had not used or attempted to use unstructured clinical text in their ML models to identify, detect, or predict sepsis onset. For instance, there were sepsis-related studies that used text but for other purposes such as mortality prediction,[61–65] phenotyping,[66] visualization,[67] exploratory data analysis,[68] and manual chart review.[69–71] Additionally, 6 articles about infection detection,[60] central venous catheter adverse events,[58] postoperative sepsis adverse events,[72–74] and septic shock identification[75] were excluded because they used manually human-curated rules instead of ML methods that automatically learn from data. The remaining 8 eligible articles were used to perform backward and forward searches,[47–50,52–55] which led to the inclusion of 1 additional article.[51] This resulted in 9 articles for synthesis.

Study characteristics

Of the 9 identified articles, 2 studies aimed at identifying infection,[47,48] 6 studies focused on early detection of sepsis,[51,53,55] severe sepsis,[49] or septic shock,[50,54] and 1 study considered both identification and early detection for a combination of sepsis, severe sepsis, and septic shock.[52] Most studies focused on intensive care unit (ICU)[48,50,52–55] or emergency department (ED)[47,51] data; only 1 used inpatient care data.[49] Four studies utilized data from hospitals,[47,49,51,52] 1 utilized MIMIC-II[54] and 4 utilized MIMIC-III.[48,50,53,55] MIMIC-II and MIMIC-III are publicly available ICU datasets created from Boston's Beth Israel Deaconess Medical Center; MIMIC-II contains data from 2001–2007[76] and MIMIC-III contains data from 2001–2012.[77] Eight studies used data from the United States[47–51,53–55] and 1 study used data from Singapore.[52] Sample sizes varied greatly in terms of the number of patients or notes used. To select patient cohorts or notes associated with sepsis, 3 studies used International Statistical Classification of Diseases and Related Health Problems (ICD) codes,[47,49,52] 5 applied sepsis definition criteria,[49–51,53,55] 1 utilized descriptions of antibiotics usage,[48] and another[54] applied criteria from Henry et al[78] that include ICD codes, sepsis criteria, and notes mentioning sepsis or septic shock. Table 1 summarizes the study characteristics and additional details are in Supplementary Table S2 (for Culliton et al,[49] the 8 structured variables for the Modified Baystate clinical definition of severe sepsis and 29 structured variables used in models were provided through personal communications with the corresponding author of Culliton et al,[49] Steve Gallant, on June 4, 2021).

Clinical text used in models

The 9 studies utilized narrative notes written by nurses,[47–50,53–55] physicians,[49–53,55] or specialists[49–51,54,55] to document symptoms,

Table 1. Study characteristics

Study (year)	Clinical setting and data source	Sample size[a]	Cohort criteria infection definition	Task and objective
Horng et al.[47] (2017)	• ED • Beth Israel Deaconess (Boston, MA, United States) • Dec 17, 2008—Feb 17, 2013	230 936 patient visits • Infection: 32 103 P; 14% • No infection: 198 833 P; 86% Train : 147 799 P; 64% Validation: 46 187 P; 20% Test: 36 950 P; 16%	Angus Sepsis ICD-9-CM abstraction criteria[79]	Identify patients with suspected infection to demonstrate benefits of using clinical text with structured data for detecting ED patients with suspected infection.
Apostolova and Velez[48] (2017)	• ICU • MIMIC-III • 2001–2012	634 369 nursing notes • Infection presence: 186 158 N; 29% • Possible infection: 3262 N; 1% • No infection: 448 211 N; 70% Train: 70% Test: 30%	Notes describing patient taking or being prescribed antibiotics for treating infection	Identify notes with suspected or presence of infection to develop a system for detecting infection signs and symptoms in free-text nursing notes.
Culliton et al.[49] (2017)	• Inpatient care • Baystate hospitals (Springfield, MA, United States) • 2012–2016	203 000 adult inpatient admission encounters • Used 68 482 E • Severe sepsis: 1427 E; 2.1% 3-fold cross validation: only text data Model construction: 2012–2015 data Test set: 2016 data: • Used 13 603 E • Severe sepsis: 425 P; 3.1%	Modified Baystate clinical definition of severe sepsis (8 structured variables) and severe sepsis ICD codes	Predict severe sepsis 4, 8, and 24 h before the earliest time structured variables meet the severe sepsis definition to compare accuracy of predicting patients that will meet the clinical definition of sepsis when using unstructured data only, structured data only, or both types.
Delahanty et al.[51] (2019)	• ED • Tenet Healthcare Hospitals (Nashville, TN, United States) • January 1, 2016—October 31, 2017	2 759 529 patient encounters • Sepsis: 54 661 E; 2% • No Sepsis: 2 704 868 E; 98% Train: 1 839 503 E; 66.7% • Sepsis: 36 458 E; 2% • No sepsis: 1 803 045 E; 98% Test: 920 026 E; 33.3% • Sepsis: 18 203 E; 2% • No sepsis: 901 823 E; 98%	Rhee's modified Sepsis-3 definition[80]	Predict sepsis risk in patients 1, 3, 6, 12, and 24 h after the first vital sign or laboratory result is recorded in the EHR to develop a new sepsis screening tool comparable to benchmark screening tools.
Liu et al.[50] (2019)	• ICU • MIMIC-III • 2001–2012	38 645 adult patients Train: 70% P Test: 30% P Applied model to: 15 930 P with suspected infection and at least 1 physiological EHR data	Sepsis-3 definition[1]	Predict septic shock in sepsis patients before the earliest time septic shock criteria are met to demonstrate an approach using NLP features for septic shock prediction.
Amrollahi et al.[53] (2020)	• ICU • MIMIC-III • 2001–2012	40 175 adult patients • Sepsis: 2805 P; ~7% Train 80% P Test 20% P	Sepsis-3 definition[1]	Predict sepsis onset hours in advance using a deep learning approach to show a pre-trained neural language representation model can improve early sepsis detection.

(continued)

Table 1. continued

Study (year)	Clinical setting and data source	Sample size[a]	Cohort criteria infection definition	Task and objective
Hammoud et al.[54] (2020)	• ICU • MIMIC-II • 2001–2007	17 763 patients • Sepsis: 6097 P • Severe sepsis: 3962 P • Septic shock : 1469 P 5-fold cross validation	Sepsis definition based on what Henry et al[78] used	Predict early septic shock in ICU patients using a model that can be optimized based on user preference or performance metrics.
Goh et al.[52] (2021)	• ICU • Singapore government-based hospital (Singapore, Singapore) • Apr 2, 2015—Dec 31, 2017	5317 patients (114 602 notes) Train and validation: 3722 P (80 162 N) • Sepsis: 6.45% • No sepsis: 93.55% Test: 1595 P (34 440 N) • Sepsis: 5.45% • No sepsis: 94.55%	ICU admission with an ICD-10 code for sepsis, severe sepsis, or sepsis shock	Identify if a patient has sepsis at consultation time or predict sepsis 4, 6, 12, 24, and 48 h after consultation to develop an algorithm that uses structured and unstructured data to diagnose and predict sepsis.
Qin et al.[55] (2021)	• ICU • MIMIC-III • 2001–2012	49 168 patients Train: 33 434 P • Sepsis: 1353 P • No Sepsis: 32 081 P Validation: 8358 P • Sepsis: 338 P • No Sepsis: 8020 P Test: 7376 P • Sepsis: 229 P • No Sepsis: 7077 P	PhysioNet Challenge restrictive Sepsis-3 definition[81]	Predict if a patient will develop sepsis to explore how numerical and textual features can be used to build a predictive model for early sepsis prediction.

ED: emergency department; ICU: intensive care unit; ICD: International Classification of Diseases; ICD-9 CM: ICD Clinical Modification, 9th revision; ICD-10: ICD 10th revision; MIMIC-II: Multiparameter Intelligent Monitoring in Intensive Care II database; MIMIC-III: Medical Information Mart for Intensive Care dataset.

[a]Sample size unit abbreviations: P: patients; N: notes; E: encounters.

signs, diagnoses, treatment plans, care provided, laboratory test results, or reports. EHRs contain various types of clinical notes. A note covers an implicit time period or activity and describes events, hypotheses, interventions, and observations within the health care provider's responsibilities. The note's form depends on its function: an order, a plan, a prescription, an investigation or analysis report, a narrative or log of events, information for the next shifts, or a requirement for legal, medical, or administrative purposes. An episode of care begins when a patient is admitted to the hospital and ends when the patient is discharged. Throughout a patient's hospital stay, documentation can include chief complaints, history-and-physical notes, progress notes, reports, descriptions of various laboratory tests, procedures, or treatments, and a discharge summary. Chief complaints are the symptoms or complaints provided by a patient for why they are seeking care.[82] History-and-physical notes can include history about the current illness, medical history, social history, family history, a physical examination, a chief complaint, probable diagnosis, and a treatment plan.[83] Progress notes document care provided and a description of the patient's condition to convey events to other clinicians.[84] Free-text reports can include interpretations of echocardiograms, electrocardiograms (ECGs), or imaging results such as X-rays, computerized tomography scans, magnetic resonance imaging scans, and ultrasounds. At discharge, the health care personnel write a discharge summary note comprised of patient details, hospital admittance reason, diagnosis, conditions, history, progress, interventions, prescribed medications, and follow-up plans.[85–87] The discharge summary letter is a formal document used to transfer patient care to another provider for further treatment and follow-up care.[88–90]

Studies have shown that nursing documentation differs from physician documentation.[91,92] Nurses document more about a patient's functional abilities than physicians,[91] and the information from notes used and the frequency of viewing and documenting differs between health care personnel.[92] Additionally, documentation varies between hospitals,[93,94] hospitals have different resources and practices,[95–97] and communicative behavior differs among professions in different wards.[98] Hence, the type of notes used, who wrote the notes, and purpose of the note will play a role in how the documentation is interpreted.[99]

Table 2 provides information regarding documentation types, author of the note, time content of the data, time latency between documentation and availability in records, and the documentation frequency. In Figure 2, the relationship between hospital events and longitudinal data used to train models is shown. As sepsis develops in a patient over time, it shows there are typically delays between a patient's actual state, clinical observations, and recorded documentation, such as ICU vital signs, narrative notes, and ICD codes.

The included studies utilized the following types of notes: 6 studies used unstructured nursing-related documentation,[47,48,50,53–55] 4 used physician notes,[50,52,53,55] 3 used radiology reports,[50,54,55] 3 used respiratory therapist progress notes,[50,54,55] 2 used ED chief complaints,[47,51] 2 used ECG interpretations,[50,54] 2 used pharmacy reports,[50,54] 2 used consultation notes,[50,52] 1 used discharge summaries,[50] 1 included mostly progress notes and history-and-physical notes,[49] and 3 used additional unspecified notes.[49,50,54] Not all notes used are listed. Liu et al[50] used all MIMIC-III notes to build a vocabulary of unique words, and discharge summaries were likely not used in predictions because they are unlikely to occur before

Table 2. Clinical documentation from electronic health records

Documentation types	Author	Description	Temporal perspective	Record latency[a]	Frequency
Chief complaints	• Physician • Nurse • Specialist	Symptoms or complaints provided by a patient at start of care for why they are seeking care.	Current	Seconds to days	One per episode
History-and-physical notes	• Physician • Nurse	Past medical history, family history, developmental history of present illness, problems about present illness, past medications or immunizations, allergies, or habits.	Retrospective	Immediately	One per episode
Progress notes	• Physician • Nurse • Specialist (eg, respiratory therapist)	Observations of patient status and care provided to document progress and response to treatment plans. For physician, it includes determining diagnosis, prescriptions, and laboratory orders.	• Retrospective • Prospective	4–8 h	One per shift
Reports	Specialist	Radiologist results and cardiology results.	Retrospective	Days	One to many per episode
Discharge summary notes	Health care personnel	Episode of care summary and follow-up plans.	• Retrospective • Prospective	At discharge or days after	One per episode
Discharge summary letter	Physician	Formal required letter containing follow-up treatment plans.	• Retrospective • Prospective	Days to months after episode	One per episode
Laboratory results	Laboratory technician	Laboratory test analysis results from provided samples (eg, blood, urine, skin, and device) based on the physician's order.	Retrospective	Days	One to many per episode
ICD codes	• Physician • Professional ICD coder • ICD data aggregator organization	Diagnosis classification for billing.	Retrospective	Days to months	One per episode
Administrative	• Administration	Patient information such as name, age, gender, address, contact information, and occupation.	• Retrospective • Current	Immediately	One per episode

[a]Record latency is defined as time between measurement/observation and the availability of the results in electronic health records.

observations. Additionally, Hammoud et al[54] used all MIMIC-II notes except discharge summaries.

These 9 studies utilized clinical notes differently. For the unit of analysis, 6 studies used a single note,[47,48,50,52–54] 1 used a set of many notes from a patient encounter,[49] 1 used a set of many notes within a specific hour of consideration,[55] and 1 used keywords from notes.[51] To identify infection signs, Horng et al[47] and Apostolova and Velez[48] processed individual notes. While Goh et al[52] used notes at each patient consultation instance to identify sepsis patients. For early detection, 5 studies defined onset time as the earliest time when definition criteria are met[49,50,53–55] and 1 defined sepsis onset time as ICU ward admission time.[52] Studies for early detection used varying windows with different durations. A window decides how and where to obtain longitudinal data, and duration is the length of

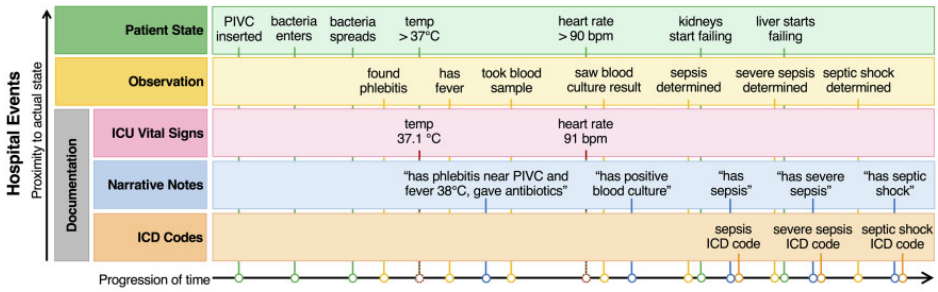


Figure 2. Overview of data from a patient timeline used to create models. The proximity of events toward a patient's actual state and the actual documentation recorded in the electronic health records typically has delays. Green represents patient states as sepsis develops in a patient. Yellow are observations made by clinicians. Documentation includes ICU vital signs[a] in pink, narrative notes in blue, and ICD codes in orange. ICU vital sign[a] documentation can be instantaneous, narrative notes can be written after observations are made, and ICD codes are typically registered after a patient is discharged. PIVC: peripheral intravenous catheter. [a]Vital signs include temperature, pulse, blood pressure, respiratory rate, oxygen saturation, and level of consciousness and awareness.

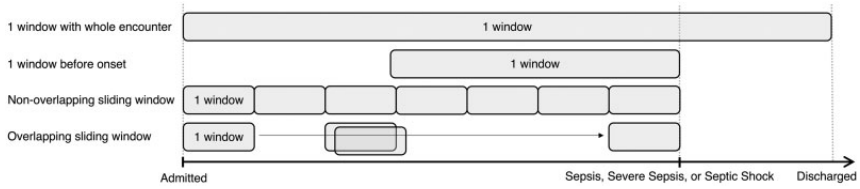


Figure 3. Different types of windows were used to obtain longitudinal data. Each gray box represents a single window, which can vary in duration (length of time) depending on the study. One window with the whole encounter means the study used a single window containing data with a duration of the whole encounter from admittance until discharge. One window before onset signifies data from a window with a duration of time before sepsis, severe sepsis, or septic shock onset. Sliding windows are consecutive windows until before sepsis, severe sepsis, or septic shock onset; this includes non-overlapping and overlapping sliding windows. Non-overlapping sliding windows indicate that data within one window of a fixed duration does not contain data in the next window. In contrast, overlapping sliding windows indicate windows of a fixed duration overlap, and data within one window will be partially in the next window.

time. As shown in Figure 3, studies can use windows differently, such as a window with the duration of the whole encounter, a window with a duration of hours before onset, non-overlapping sliding windows with a fixed duration until onset, or overlapping sliding windows with a fixed duration until onset. Culliton et al[49] used a 4-, 8-, or 24-h duration window before severe sepsis, and concatenated all text within a window. Goh et al[52] used a 4-, 6-, 12-, 24-, or 48-h duration window of before sepsis, severe sepsis, or septic shock onset. Liu et al[50] used 10 data points within a 1-h duration window spanning 2 h before septic shock, and used the most recently entered note for a data point to predict septic shock. Hammoud et al[54] binned data in 15-minute duration non-overlapping sliding windows to update septic shock predictions every 15 minutes, and used the last note within the window. Amrollahi et al[53] binned data into 1-h duration non-overlapping sliding windows to provide hourly sepsis predictions, and used sentences within a note to capture the semantic meanings. Qin et al[55] used 6-h duration overlapping sliding windows with 6 data points to predict sepsis; a data point was generated from each hour within the window and all clinical notes within the hour were concatenated in random-order. Delahanty et al[51] used a 1-, 3-, 6-, 12-, or 24-h duration window after the first vial sign or laboratory result was documented in the EHR to identify patients at risk for sepsis, and utilized keywords.

First 2 columns in Table 3 show the type of text and unit of analysis used. Additional details about variables and specific notes used are listed in Supplementary Table S3 (the types of notes and usage for Liu et al[50] was confirmed through personal communications with Ran Liu on June 2, 2021, for Hammoud et al[54] by Ibrahim Hammoud on May 29, 2021, and for Qin et al[55] by Fred Qin on September 9, 2021. Additionally, the structured variables used in models for Culliton et al[49] were provided through personal communications with Steve Gallant on June 4, 2021). In Figure 4, single notes or a set of many notes are preprocessed and represented to extract features, whereas keywords are used as is. Then structured data can be added, and the data are used to train ML models.

As shown in Figures 3 and 4 and listed in Tables 1 and 3 and Supplementary Tables S2 and S3, although all studies are related to sepsis, there are varying sample sizes, data types, inclusion criteria, and objectives. This heterogeneity makes it challenging to compare results for a meta-analysis.

Natural language processing and machine learning study outcomes

To utilize text in ML, it must be transformed into a representation understandable by computers. In order to do that, Bag-of-words (BoW),[100] n-gram, term frequency-inverse document frequency (tf-idf), and paragraph vectors (PV)[101] representations can be used. These representations can be improved using additional NLP techniques, such as stop word removal, lemmatization, and stemming. In addition, other useful features can be extracted from text using part-of-speech (POS) tagging, named entity recognition, or Latent Dirichlet Allocation (LDA) topic modeling.[102] In recent years, neural networks (NNs) have shown high predictive performance. As a result, many state-of-the-art results have been achieved using NNs to learn

Table 3. Text used in studies

Study (year)	Free-text document type	Unit of analysis	Text processing
Horng et al.[47] (2017)	• ED chief complaints • Nursing triage assessments	One note	Representation: • Bi-gram • BoW (15 240-word vocabulary) • LDA topic modeling (500 topics) Techniques: • Convert to lowercase • Remove rare tokens and punctuation • Negation
Apostolova and Velez[48] (2017)	Nursing notes	One note	Representation: • BoW • CBOW (200 vector size with window size of 7 = 441-term vocabulary of antibiotics usage and rules for negation and speculations) • tf-idf • PV (600 vector size for document-level representation) Techniques: • Convert to lowercase • Remove frequent tokens and non-alphanumeric characters • Negation
Culliton et al.[49] (2017)	Clinical notes (mostly progress notes and history-and-physical notes)	One patient encounter = many notes	Representation: • GloVe (300-dimensional vector) + summing word vectors Techniques: • Concatenated all notes for an encounter into a single text block
Delahanty et al.[51] (2019)	ED chief complaints	Keywords	Other: • Keywords extracted by experts
Liu et al.[50] (2019)	All MIMIC-III clinical notes, such as but not limited to: • Nursing notes • Physician notes	One note	Representation: • BoW (8907 unique term vocabulary and 832 predictive terms) • GloVe (300-dimensional vector for each unique term) Techniques: • Convert to lowercase • Remove rare tokens, frequent tokens, and non-alphanumeric characters
Amrollahi et al.[53] (2020)	• Nursing notes • Physician notes	One note	Representation: • tf-idf (2227 vector size features = 2187 text features + 40 structured features) • ClinicalBERT (808 vector size features = 768 text features + 40 structured features) Techniques: • Remove rare tokens, frequent tokens, stop words, dates, and special characters
Hammoud et al.[54] (2020)	All MIMIC-II notes except discharge summaries, such as but not limited to: • Nursing progress notes • Respiratory therapist progress notes	One note	Representation: • BoW • tf-idf Techniques: • Remove rare and frequent tokens

(continued)

Table 3. continued

Study (year)	Free-text document type	Unit of analysis	Text processing
Goh et al.[52] (2021)	Physician notes: • Admission notes • Progress notes • ICU consultations • Pharmacy notes • Allied health notes	One note	Representation: • tf-idf • LDA topic modeling (100 topics) Techniques: • Remove rare tokens, punctuation, and stop words • Lemmatization • POS tagging • Manual classification of topics into categories
Qin et al.[55](2021)	• Nursing notes • Physician notes • Radiology notes • Respiratory notes	Many notes	Representation: • tf-idf (1000 vector size = 1000 most common term vocabulary) • ClinicalBERT (768 vector size features[a] = either by concatenating all text first as input or using individual notes as input and concatenating output of individual notes) Techniques: • Random-order concatenation of all clinical notes within the hour of consideration.[a] • Named entity recognition

BoW: Bag-of-words; CBOW: Continuous bag-of-words; ClinicalBERT: Clinical Bidirectional Encoder Representations from Transformers; ED: emergency department; GloVe: Global Vectors for Word Representation; ICU: intensive care unit; LDA: Latent Dirichlet Allocation; POS tagging: Part-of-speech tagging; PV: paragraph vectors; tf-idf: term frequency-inverse document frequency.

[a]Representation and technique details for Qin et al[55] were provided through personal communications (with Fred Qin on September 7, 2021).

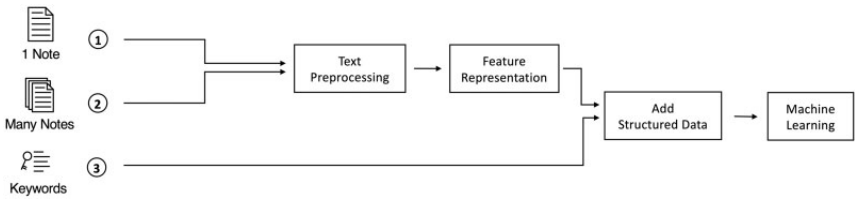


Figure 4. The unit of analysis used to train machine learning models for the included studies was either (1) a single note, (2) a set of many notes, or (3) keywords. In general, text was preprocessed and represented as features interpretable by a computer, then structured data were added, and the data were used to fit machine learning models.

a suitable representation of texts, often known as embeddings.[103] Embedding techniques include Global Vectors for Word Representation (GloVe),[104] Word2Vec as a continuous bag-of-words (CBOW) model or skip-gram model,[105] Bidirectional Encoder Representations from Transformers (BERT),[106] and ClinicalBERT.[107] The advantage of using embeddings is that it retains the sequential information lost in a BoW representation and does feature extraction automatically.[103]

Utilized text processing operations are in Table 3. One study used keyword extraction instead of text processing operations.[51] Six studies used tokenization of words for word-level representation,[47–50,52,54] 1 also tried PV for document-level representation,[48] and another used the first 40 tokens in a sentence to get sentence-level representation and averaged sentence-level representations to provide document-level representation.[53] The most common technique for improving representation was token removal, such as removing rare tokens,[47,50,52–54] frequent tokens,[48,50,53,54] punctuation or special characters,[47,48,50,52,53] and stop words.[52,53] The most frequently used representation was tf-idf,[48,52–55] followed by BoW,[47,48,50,54] LDA,[47,52] GloVe,[49,50] ClinicalBERT,[53,55] bi-gram,[47] CBOW,[48] and PV.[48] Three studies created a vocabulary of unique terms using BoW,[50] CBOW,[48] and tf-idf.[53] Apostolova and Velez[48] found that using structured data was inadequate for identifying infection in nursing notes, so they used antibiotic usage and word embeddings to create a labeled dataset of notes with infection, suspected infection, and no infection. Additionally, Horng et al[47] and Liu et al[50] listed predictive terms in their models, and Goh et al[52] provided a list of categories used to classify the top 100 terms. Examples of predictive features

are: (1) For sepsis, severe sepsis, or septic shock, Goh et al[52] classified the top 100-topics into 7 categories: clinical condition or diagnosis, communication between staff, laboratory test order or results, non-clinical condition updates, social relationship information, symptoms, and treatments or medication. (2) Liu et al's[50] most predictive NLP terms for the pre-shock versus non-shock state include "tube," "crrt," "ards," "vasopressin," "portable," "failure," "shock," "sepsis," and "dl." (3) Horng et al's[47] most predictive terms or topics for having an infection in the ED include "cellulitis," "sore_throat," "abscess," "uti," "dysuria," "pneumonia," "redness_swelling," "erythema," "swelling," "redness, celluititis, left, leg, swelling, area, rle, arm, lle, increased, erythema," "abcess, buttock, area, drainage, axilla, groin, painful, thigh, left, hx, abcesses, red, boil," and "cellulitis, abx, pt, iv, infection, po, keflex, antibiotics, leg, treated, started, yesterday." Whereas the least predictive terms or topics for not having an infection include "motor vehicle crash," "laceration," "epistaxis," "pancreatitis", "etoh"(ethanol for drunkenness), "etoh, found, vomiting, apparently, drunk, drinking, denies, friends, trauma_neg, triage," and "watching, tv, sitting, sudden_onset, movie, television, smoked, couch, pt, pot, 5pm, theater."

ML methods for detecting sepsis using clinical text included: ridge regression,[49] lasso regression,[54] logistic regression,[47,48,52] Naïve Bayes (NB),[47] support vector machines (SVMs),[47,48] K-nearest neighbors (KNNs),[48] random forest (RF),[47,52] gradient boosted trees (GBTs),[50–52,55] gated recurrent unit (GRU),[50] and long short-term memory (LSTM).[53] Although the methods are listed separately, 2 studies combined different ML methods[48,52] (see Supplementary Table S4 for details). Ridge and lasso regression are linear regression methods that constrain the model parameters. A linear regression model is represented as $\hat{y} = \beta_1 x + \beta_0$, where $\hat{y}$ is the predicted value, x is the input variable and β_1 and β_0 are model parameters. Model parameters are estimated by minimizing $\sum_{i=1}^{N} (y_i - \hat{y}_i)^2$, where y_i is the label and N is the number of training samples. In ridge and lasso regression, $\sum_{i=1}^{N} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{2} f(\beta_j)$ is minimized instead, where λ is a hyperparameter that trades-off between fitting the data and model complexity, and $f(z) = z^2$ for ridge regression or $f(z) = |z|$ for lasso regression. Logistic regression is a classification method that models $P(y|x)$, which is the probability of a class y given the feature x. The logistic regression model is defined as $f(x) = \frac{1}{1+e^{-(\beta_1 x + \beta_0)}}$. NB is a Bayesian network that eases computation by assuming all input variables are independent given the outcome.[108] SVM is an extension of a support vector classifier that separates training data points into 2 class regions using a linear decision boundary and classifies new data points based on which region they belong to. To accommodate for non-linearity in the data, SVM enlarges the feature space by applying kernels.[109] KNNs assume similar data points are close together and use similarity measures to classify new data based on "proximity" to points in the training data.[110] RF and GBT are ensemble models that use a collection of decision trees to improve the predictive performance of the models. RF classification takes the majority vote of a collection of trees to reduce the decision tree variance.[111] GBT trains decision trees sequentially so that each tree trains based on information from previously trained trees.[112,113] To avoid overfitting, each tree is scaled by a hyperparameter λ, often known as the shrinkage parameter or learning rate that controls the rate the model learns. Recurrent neural networks (RNNs) are a type of NN with recurrent connections and assume that the input data have an ordering, for example, words in a sentence.[114–116] RNN can be seen as a feed-forward NN with a connection from output to input.[115] GRU[117] and LSTM[118] are improved variations of RNN with gating mechanisms to combat the vanishing gradient problem. The improvements help the models to better model long-term temporal dependencies. To tune hyperparameters, grid-search and Bayesian optimization were used in the studies.[47,48,50,53,54] The grid-search method iterates exhaustively through all hyperparameter values within a pre-defined set of values to find the optimal hyperparameter with respect to a validation set. In contrast, the Bayesian optimization method makes informed choices on which values to evaluate using the Bayes formula. The goal of using Bayesian optimization for hyperparameter tuning is to minimize the number of values to evaluate.

All studies reported evaluation results for different algorithms or data types and almost all reported area under the receiver operating characteristic curve (AUC) values except 1.[48] Figure 5 shows differences in AUC values for infection (Figure 5A), sepsis (Figure 5B), septic shock (Figure 5C), and severe sepsis (Figure 5E) when using structured data only, text data only, or a combination of structured and text data. Studies that compared their methods for different hours prior to onset are also included (Figure 5D and F), the lines connecting the points are to visually separate the methods and do not indicate changing AUC values over time. This figure compares data type usage and model performance within an individual study; it should not be used to compare AUC values between subfigures and studies because the studies used different cohorts, sepsis definitions, and hours before onset. Additionally, sepsis, severe sepsis, and septic shock have different manifestations.[119,120] Table 4 summarizes the best and worst AUC values for each study; a full table with additional evaluation metrics is available in Supplementary Table S4 (number of hours before onset for Amrollahi et al[53] was confirmed through personal communications with Shamim Nemati on May 27, 2021 and Fatemeh Amrollahi on June 13, 2021). GBT was the most widely used ML method,[50–52,55] followed by logistic regression,[47,48,52] SVMs,[47,48] RF,[47,52] ridge regression,[49] lasso regression,[54] NB,[47] KNNs,[48] GRU,[50] and LSTM.[53] For hyperparameter tuning, 3 studies used the grid-search method[47,48,54] and 2 used the Bayesian optimization method[50,53] (hyperparameter tuning was provided by personal communication with Ran Liu on September 7, 2021 and Fatemeh Amrollahi on September 7, 2021). Delahanty et al,[51] Hammoud et al,[54] Goh et al,[52] and Qin et al[55] compared their algorithm to scoring systems used in clinical practice, such as SIRS,[121] sequential organ failure assessment (SOFA),[122] quick SOFA (qSOFA),[123] modified early warning system (MEWS),[124] or a targeted real-time early warning score (TREWScore).[78] In addition, Apostolova and Velez[48] evaluated their model on a ground truth set with 200 nursing notes that were manually reviewed by a qualified professional, and Goh et al[52] compared their model with the Rhodes et al[125] sepsis guidelines used by physicians. Furthermore, Horng et al[47] performed additional tests on different patient cohorts for error analysis. Although results are difficult to compare directly because of study heterogeneity, most results suggest that utilizing both structured data and text generally results in better performance for sepsis identification and early detection.

DISCUSSION

Identification, early detection, prediction, and method transferability

Nine studies utilized clinical text for sepsis identification, early detection, or prediction. As all identified studies focus on the identification or early detection of sepsis within a fixed time frame, this

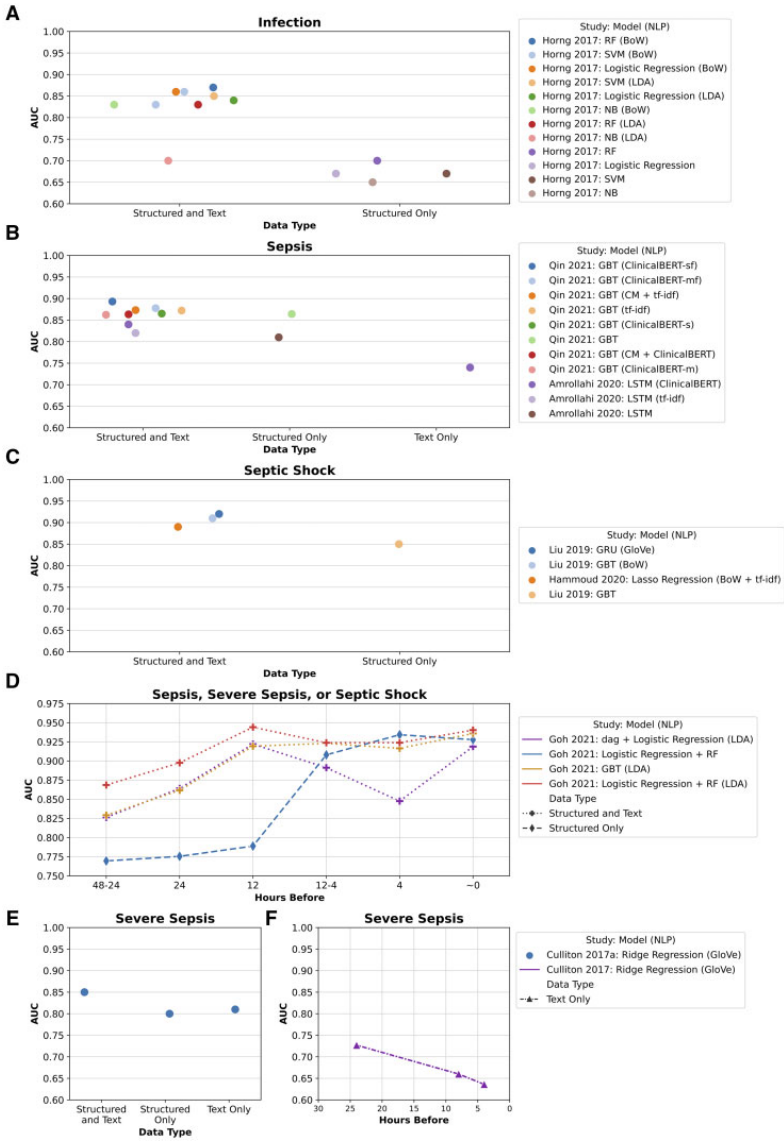


Figure 5. Overview of area under the curve (AUC) values for identification or early detection of infection, sepsis, septic shock, and severe sepsis using different data types (structured data and text, structured data only, and text only).* Each figure contains the study and year, machine learning model,[a] and natural language processing technique[b]. (A) AUC values for infection identification. Horng et al[47] 2017: SVM (BoW) has 2 AUC values; 0.86 when using chief complaints and nursing notes and 0.83 when using only chief complaints. (B) AUC values for early sepsis detection. Amrollahi et al[53] AUC values are from detecting 4 h before sepsis onset, and Qin et al[55] AUC values are the average from detecting 0 to 6 h before sepsis onset. (C) AUC values for early septic shock detection. Hammoud et al[54] AUC values are from detecting 30.64 h before septic shock onset, and Liu et al[50] AUC values are from detecting 6.0 to 7.3 h before septic shock onset. (D) AUC values for early sepsis, severe sepsis, or septic shock detection and sepsis identification in Goh et al.[52] Different symbols separate data types. (E) AUC values for early septic shock detection for Culliton et al[49] using results from the test set. (F) AUC values for early septic shock detection for Culliton et al[49] using results from 3-fold validation. *Disclaimer: AUC values should not be directly compared between studies and different figures for infection, sepsis, severe sepsis, and septic shock. Additionally, the lines connecting points do not indicate AUC values changing over time (Figure 5D and 5F); lines only separate the different methods visually. [a]Machine learning models: dag: dagging (partition data into disjoint subgroups); GBT: gradient boosted trees; GRU: gated recurrent unit; LSTM: long short-term memory; NB: Naïve Bayes; RF: random forest; SVM: support vector machines. [b]Natural language processing techniques: BoW: Bag-of-words; ClinicalBERT: Clinical Bidirectional Encoder Representations from Transformers; ClinicalBERT-m: ClinicalBERT from merging all textual features to get embeddings; ClinicalBERT-sf; finetuned ClinicalBERT from concatenating individual embeddings of each textual feature; CM: Amazon Comprehend Medical service for named entity recognition; GloVe: Global Vectors for Word Representation; LDA: Latent Dirichlet Allocation; tf-idf: term frequency-inverse document frequency.

Table 4. Study outcome overview of best and worst area under the curve values

Study (year)	Hours[a]	Data types[b]		Models[d] (NLP)[e]	AUC[f]
		DVLMC	T[c]		
Horng et al.[47] (2017)	Identify	DV- - -	CC + NN	RF (BoW)	0.87
		DV- - -	–	NB	0.65
Apostolova and Velez[48] (2017)	Identify	- - - - -	NN	SVM (BoW + tf-idf)	–
		- - - - -	NN	Logistic regression + KNN + SVM (PV)	–
Culliton et al.[49] (2017)	−4	- - - - -	CN	Ridge regression (GloVe)	0.64
	−8	- - - - -	CN	Ridge regression (GloVe)	0.66
	−24	- - - - -	CN	Ridge regression (GloVe)	0.73
	−24[g]	-V- -C	CN	Ridge regression (GloVe)	0.85
		-V- -C	–	Ridge regression (GloVe)	0.80
Delahanty et al.[51] (2019)	+1	-VL- -	–	GBT	0.93
	+3	-VL- -	–	GBT	0.95
	+6	-VL- -	–	GBT	0.96
	+12	-VL- -	–	GBT	0.97
	+24	-VL- -	–	GBT	0.97
Liu et al.[50] (2019)	−7	-VLM-	CN	GRU (GloVe)	0.92
	−7.3	-VLM-	CN	GBT (BoW)	0.91
	−6	-VLM-	–	GBT	0.85
Amrollahi et al.[53] (2020)	−4[h]	-VL- -	PN + NN	LSTM (ClinicalBERT)	0.84
		- - - - -	PN + NN	LSTM (ClinicalBERT)	0.74
Hammoud et al.[54] (2020)	−30.6	DVL- -	CN	Lasso regression (BoW + tf-idf)	0.89
Goh et al.[52] (2021)	Identify	DVLM-	PN	Logistic regression + RF (LDA)	0.94
		DVLM-	PN	dag + Logistic regression (LDA)	0.92
	−4	DVLM-	–	Logistic regression + RF	0.93
		DVLM-	PN	dag + Logistic regression (LDA)	0.85
	−6	DVLM-	PN	Logistic regression + RF (LDA)	0.92
		DVLM-	PN	dag + Logistic regression (LDA)	0.89
	−12	DVLM-	PN	Logistic regression + RF (LDA)	0.94
		DVLM-	–	Logistic regression + RF	0.79
	−24	DVLM-	PN	Logistic regression + RF (LDA)	0.90
		DVLM-	–	Logistic regression + RF	0.78
	−48	DVLM-	PN	Logistic regression + RF (LDA)	0.87
		DVLM-	–	Logistic regression + RF	0.77
Qin et al.[55] (2021)	−6 to 0[i]	-VL- -	CN	GBT (ClinicalBERT-sf)	0.89[i]
		-VL- -	–	GBT (ClinicalBERT-m)	0.86[i]

[a]Hours: Identify: not detecting hours before or after; –: hours before; +: hours after an event.

[b]Data types: D: demographics; V: vitals; L: laboratory; M: medications; C: codes; T: text; -'s position in DVLMC indicates which is not used.

[c]Text data types: CC: chief complaints; CN: various types of clinical notes; NN: nursing notes; PN: physician notes; –: no notes.

[d]Machine learning models: dag: dagging (partition data into disjoint subgroups); GBT: gradient boosted trees; GRU: gated recurrent unit; KNN: K-nearest neighbors; LSTM: long short-term memory; NB: Naïve Bayes; RF: random forest; SVM: support vector machines.

[e]Natural language processing (NLP) techniques: BoW: Bag-of-words; ClinicalBERT: Clinical Bidirectional Encoder Representations from Transformers; ClinicalBERT-m: ClinicalBERT from merging all textual features to get embeddings; ClinicalBERT-sf: finetuned ClinicalBERT from concatenating individual embeddings of each textual feature; GloVe: Global Vectors for Word Representation; LDA: Latent Dirichlet Allocation; PV: paragraph vectors; tf-idf: term frequency-inverse document frequency.

[f]Area under the curve (AUC). Apostolova and Velez[48] did not provide metrics for AUC.

[g]Culliton et al[49] performed 2 experiments, these results are from using a test set instead of 3-fold validation.

[h]Number of hours before onset for Amrollahi et al[53] was confirmed through personal communications (with Shamim Nemati on May 27, 2021 and Fatemeh Amrollahi on June 13, 2021).

[i]Qin et al[55] AUC values are an average from 0 to 6 h before sepsis, not the specified hours.

indicates much work is still needed before sepsis prediction can use text from complete patient histories. Studies from this review focus mainly on the ICU and ED, and the addition of continuous measurements of vital signs for sepsis makes generalizability to the ward units limited. However, Culliton et al[49] was successful in detecting sepsis early utilizing only the text from EHR clinical notes, which is a promising approach for all inpatients. Additionally, Horng et al[47] showed that their ML model performed on subsets of specific patient cohorts like pneumonia or urinary tract infection. The different ML methods and NLP techniques from each study may be applicable for different retrospective cohort or case–control studies. Though the studies have varying sepsis definitions, cohorts, ML methods, and NLP techniques, overall, they show that using clinical text and structured data can improve sepsis identification and early detection. Unstructured clinical text predicts sepsis 48–12 h before onset, while structured data predicts sepsis closer to onset (<12 h before).

Sepsis definition impact

In ML, many studies rely heavily on sepsis definitions and ICD-codes to identify patient cohort datasets for sepsis studies.[9,11,13]

Among changing sepsis definitions over time are the 2001 Angus Sepsis ICD-9 abstraction criteria,[79] 2012 Surviving Sepsis Campaign Guidelines,[126] 2016 Sepsis-3 consensus definition,[1] and 2017 Rhee's modified Sepsis-3 definition.[80] Although a consensus sepsis definition exists,[1] not all definition elements will be present in a sepsis patient because sepsis is a very heterogeneous syndrome[127] and the infection site is difficult to identify correctly.[128] Many patients with sepsis are often misdiagnosed with other diseases such as respiratory failure[129] and pneumonia.[129,130] In practice, hospitals also have varying sepsis coding methods.[131–135] As the sepsis definitions change, studies also tend to use the most current definition in their study. A recent study that used different sepsis definitions to generate patient cohorts found significant heterogeneous characteristics and clinical outcomes between cohorts.[136] Similarly, previous work by Liu et al[137] demonstrated that using different infection criteria resulted in a different number of patients and slightly different outcomes. Similar to how changes in the definition and varying coding methods can affect sepsis mortality outcomes,[138] the sepsis definition and codes used in ML studies will likely change the outcome, results, and reporting methods. Thus, future studies should acknowledge that sepsis is a syndrome and clearly characterize each sign of sepsis to reflect the heterogeneity in the definition.

Suggestions for future studies

Predicting sepsis earlier than 12 h prior to sepsis onset can reduce treatment delays and improve patient outcomes.[3,4] Because predictions 48–12 h before sepsis onset appear to rely more on clinical text than structured data, additional NLP techniques should be considered for future ML studies. Additionally, since the sepsis definition used will change the cohort, this indicates opportunities to expand the cohort. Like Apostolova and Velez,[48] who determined their cohort by finding notes describing the use of antibiotics. It should be possible to determine cohorts by using notes describing infection signs (eg, fever, hypotension, or deterioration in mental status), indicators of diseases that sepsis is misdiagnosed with (eg, pulmonary embolism, adrenal insufficiency, diabetic ketoacidosis, pancreatitis, anaphylaxis, bowel obstruction, hypovolemia, colitis, or vasculitis), or medication effect and toxin ingestion, overdose, or withdrawal.[139] NLP methods from infectious diseases known to trigger sepsis can be incorporated to extract infection signs and symptoms from the text for determining potential sepsis signs, patient groups, and risk factors. For instance, many sepsis patients are often admitted with pneumonia, and there are several studies about identifying pneumonia from radiology reports using NLP.[23,140,141] Additionally, heterogeneous sepsis signs or symptoms might be identified by utilizing NLP features for detecting healthcare-associated infections risk patterns[59] or infectious symptoms.[142] Information from other NLP related reviews about using clinical notes can also be applied, such as: challenges to consider,[16] clinical information extraction tools and methods,[18] methods to overcome the need for annotated data,[22] different embedding techniques,[143,144] sources of labeled corpora,[143] transferability of methods,[145] and processing and analyzing symptoms.[146] Moreover, heterogeneous or infectious diseases, with overlapping signs and symptoms of other diseases, can utilize similar sepsis ML and NLP methods to improve detection. The identified studies did not utilize complete patient history data. Thus, future research utilizing complete patient history data can study if sepsis risk can be predicted earlier than 48 h by incorporating sepsis risk factors, such as comorbidities,[7] chronic diseases,[147] patient trajectories,[148] or prior infection incidents.[149]

Limitations

This review has several limitations. The narrow scope of including only studies about utilizing clinical text for sepsis detection or prediction could have missed studies that use other types of text for sepsis detection or prediction. For example, search terms did not include "early warning system," "feature extraction," and "topic modeling." Additionally, search terms did not include possible sources of infection for sepsis, such as bloodstream infection, catheter-associated infection, pneumonia, and postoperative surgical complications. Further, the sensitivity to detect sepsis in text, structured data, or the combined data from these will depend on the timestamps these data recordings have in the EHR. These timestamps may vary depending on the data used to inform the study or the different systems implemented at different hospitals. The articles identified in this review had a homogenous choice of structured data (ie, demographics, vital signs, and laboratory measurements). Of those, laboratory test results have the largest time lag, around 1–2 h to obtain the blood test results.[150] Thus, the good performance of text to detect sepsis in these articles are unlikely explained fully by the time lag between measurement and recording of the structured data. This review thus shows that it is possible to detect sepsis early using text, with or without the addition of structured data.

CONCLUSION

Many studies about sepsis detection exist, but very few studies utilize clinical text. Heterogeneous study characteristics made it difficult to compare results; however, the consensus from most studies was that combining structured data with clinical text improves identification and early detection of sepsis. There is a need to utilize the unstructured text in EHR data to create early detection models for sepsis. The lack of utilizing the complete patient history in early prediction models for sepsis is an opportunity for future ML and NLP studies.

FUNDING

Financial support for this study was provided by the Computational Sepsis Mining and Modelling project through the Norwegian University of Science and Technology Health Strategic Area.

AUTHOR CONTRIBUTIONS

MYY and ØN conceptualized the study and design with substantial clinical insight from LTG. MYY conducted the literature search and initial analysis, LTG verified results, and ØN resolved discrepancies. All authors participated in data analysis and interpretation. MYY drafted the manuscript, which LTG and ØN critically revised.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

ACKNOWLEDGMENTS

We thank those from the Gemini Center for Sepsis Research group for valuable discussions and recommendations related to clinical databases, missing search terms, and presenting results. Specifically, Ms Lise Husby Høvik (RN), Dr Erik Solligård, Dr Jan Kristian Damås, Dr Jan Egil Afset, Dr Kristin Vardheim Liyanarachi, Dr Randi Marie Mohus, and Dr Anuradha Ravi.

CONFLICT OF INTEREST STATEMENT

None declared.

DATA AVAILABILITY

The data underlying this article are available in the article and in its online supplementary material.

REFERENCES

1. Singer M, Deutschman CS, Seymour CW, *et al.* The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA* 2016; 315 (8): 801–10.
2. Fleischmann C, Scherag A, Adhikari NKJ, *et al.*; International Forum of Acute Care Trialists. Assessment of global incidence and mortality of hospital-treated sepsis. Current estimates and limitations. *Am J Respir Crit Care Med* 2016; 193 (3): 259–72.
3. Rivers E, Nguyen B, Havstad S, *et al.* Early goal-directed therapy in the treatment of severe sepsis and septic shock. *N Engl J Med* 2001; 345 (19): 1368–77.
4. Kumar A, Roberts D, Wood KE, *et al.* Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Crit Care Med* 2006; 34: 1589–96.
5. Polat G, Ugan RA, Cadirci E, *et al.* Sepsis and septic shock: current treatment strategies and new approaches. *Eurasian J Med* 2017; 49 (1): 53–8.
6. Arnold C. News feature: the quest to solve sepsis. *Proc Natl Acad Sci USA* 2018; 115 (16): 3988–91.
7. Iskander KN, Osuchowski MF, Stearns-Kurosawa DJ, *et al.* Sepsis: multiple abnormalities, heterogeneous responses, and evolving understanding. *Physiol Rev* 2013; 93 (3): 1247–88.
8. Jawad I, Lukšić I, Rafnsson SB. Assessing available information on the burden of sepsis: global estimates of incidence, prevalence and mortality. *J Glob Health* 2012; 2 (1): 010404.
9. Islam MM, Nasrin T, Walther BA, *et al.* Prediction of sepsis patients using machine learning approach: a meta-analysis. *Comput Methods Programs Biomed* 2019; 170: 1–9.
10. Schinkel M, Paranjape K, Nannan Panday RS, *et al.* Clinical applications of artificial intelligence in sepsis: a narrative review. *Comput Biol Med* 2019; 115: 103488.
11. Wulff A, Montag S, Marschollek M, *et al.* Clinical decision-support systems for detection of systemic inflammatory response syndrome, sepsis, and septic shock in critically Ill patients: a systematic review. *Methods Inf Med* 2019; 58 (S 02): e43–57.
12. Teng AK, Wilcox AB. A review of predictive analytics solutions for sepsis patients. *Appl Clin Inform* 2020; 11 (3): 387–98.
13. Fleuren LM, Klausch TLT, Zwager CL, *et al.* Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive Care Med* 2020; 46 (3): 383–400.
14. Giacobbe DR, Signori A, Del Puente F, *et al.* Early detection of sepsis with machine learning techniques: a brief clinical perspective. *Front Med (Lausanne)* 2021; 8: 617486.
15. Assale M, Dui LG, Cina A, *et al.* The revival of the notes field: leveraging the unstructured content in electronic health records. *Front Med (Lausanne)* 2019; 6: 66.
16. Tayefi M, Ngo P, Chomutare T, *et al.* Challenges and opportunities beyond structured data in analysis of electronic health records. *Wiley Interdiscip Rev Comput Stat* 2021; 13: e1549. doi:10.1002/wics.1549
17. Sheikhalishahi S, Miotto R, Dudley JT, *et al.* Natural language processing of clinical notes on chronic diseases: systematic review. *JMIR Med Inform* 2019; 7 (2): e12239.
18. Wang Y, Wang L, Rastegar-Mojarad M, *et al.* Clinical information extraction applications: a literature review. *J Biomed Inform* 2018; 77: 34–49.
19. Datta S, Bernstam EV, Roberts K. A frame semantic overview of NLP-based information extraction for cancer-related EHR notes. *J Biomed Inform* 2019; 100: 103301.
20. Jackson RG, Patel R, Jayatilleke N, *et al.* Natural language processing to extract symptoms of severe mental illness from clinical text: the Clinical Record Interactive Search Comprehensive Data Extraction (CRIS-CODE) project. *BMJ Open* 2017; 7 (1): e012012.
21. Kreimeyer K, Foster M, Pandey A, *et al.* Natural language processing systems for capturing and standardizing unstructured clinical information: a systematic review. *J Biomed Inform* 2017; 73: 14–29.
22. Spasic I, Nenadic G. Clinical text data in machine learning: systematic review. *JMIR Med Inform* 2020; 8 (3): e17984.
23. Elkin PL, Froehling D, Wahner-Roedler D, *et al.* NLP-based identification of pneumonia cases from free-text radiological reports. *AMIA Annu Symp Proc* November 8–12, 2008; 2008: 172–6; Washington, DC.
24. Jensen K, Soguero-Ruiz C, Oyvind Mikalsen K, *et al.* Analysis of free text in electronic health records for identification of cancer patient trajectories. *Sci Rep* 2017; 7: 46226.
25. Ford E, Carroll JA, Smith HE, *et al.* Extracting information from the text of electronic medical records to improve case detection: a systematic review. *J Am Med Inform Assoc* 2016; 23 (5): 1007–15.
26. Soguero-Ruiz C, Hindberg K, Mora-Jiménez I, *et al.* Predicting colorectal surgical complications using heterogeneous clinical data and kernel methods. *J Biomed Inform* 2016; 61: 87–96.
27. Huddar V, Desiraju BK, Rajan V, *et al.* Predicting complications in critical care using heterogeneous clinical data. *IEEE Access* 2016; 4: 7988–8001.
28. Ribelles N, Jerez JM, Rodriguez-Brazzarola P, *et al.* Machine learning and natural language processing (NLP) approach to predict early progression to first-line treatment in real-world hormone receptor-positive (HR+)/HER2-negative advanced breast cancer patients. *Eur J Cancer* 2021; 144: 224–31.
29. Friedlin J, Overhage M, Al-Haddad MA, *et al.* Comparing methods for identifying pancreatic cancer patients using electronic data sources. *AMIA Annu Symp Proc* November 13–17, 2010; 2010: 237–41; Washington, DC.
30. Van Vleck TT, Chan L, Coca SG, *et al.* Augmented intelligence with natural language processing applied to electronic health records for identifying patients with non-alcoholic fatty liver disease at risk for disease progression. *Int J Med Inform* 2019; 129: 334–41.
31. DeLisle S, Kim B, Deepak J, *et al.* Using the electronic medical record to identify community-acquired pneumonia: toward a replicable automated strategy. *PLoS One* 2013; 8 (8): e70944.
32. Gundlapalli AV, South BR, Phansalkar S, *et al.* Application of natural language processing to VA electronic health records to identify phenotypic characteristics for clinical and research purposes. *Summit Transl Bioinform* March 10–12, 2008; 2008: 36–40; San Francisco, CA.
33. Ananthakrishnan AN, Cai T, Savova G, *et al.* Improving case definition of Crohn's disease and ulcerative colitis in electronic medical records using natural language processing: a novel informatics approach. *Inflamm Bowel Dis* 2013; 19 (7): 1411–20.
34. Carroll RJ, Thompson WK, Eyler AE, *et al.* Portability of an algorithm to identify rheumatoid arthritis in electronic health records. *J Am Med Inform Assoc* 2012; 19 (e1): e162–9.
35. Liao KP, Cai T, Gainer V, *et al.* Electronic medical records for discovery research in rheumatoid arthritis. *Arthritis Care Res (Hoboken)* 2010; 62 (8): 1120–7.
36. Carroll RJ, Eyler AE, Denny JC. Naïve electronic health record phenotype identification for Rheumatoid arthritis. *AMIA Annu Symp Proc* October 22–26, 2011; 2011: 189–96; Washington, DC.
37. Xia Z, Secor E, Chibnik LB, *et al.* Modeling disease severity in multiple sclerosis using electronic health records. *PLoS One* 2013; 8 (11): e78927.
38. DeLisle S, South B, Anthony JA, *et al.* Combining free text and structured electronic medical record entries to detect acute respiratory infections. *PLoS One* 2010; 5 (10): e13377.
39. Zheng H, Gaff H, Smith G, *et al.* Epidemic surveillance using an electronic medical record: an empiric approach to performance improvement. *PLoS One* 2014; 9 (7): e100845.

40. Bhattacharjee P, Edelson DP, Churpek MM. Identifying patients with sepsis on the hospital wards. *Chest* 2017; 151 (4): 898–907.
41. Despins LA. Automated detection of sepsis using electronic medical record data: a systematic review. *J Healthc Qual* 2017; 39 (6): 322–33.
42. Peiffer-Smadja N, Rawson TM, Ahmad R, *et al.* Machine learning for clinical decision support in infectious diseases: a narrative review of current applications. *Clin Microbiol Infect* 2020; 26 (5): 584–95.
43. Freeman R, Moore LSP, García Álvarez L, *et al.* Advances in electronic surveillance for healthcare-associated infections in the 21st century: a systematic review. *J Hosp Infect* 2013; 84 (2): 106–19.
44. de Bruin JS, Seeling W, Schuh C. Data use and effectiveness in electronic surveillance of healthcare associated infections in the 21st century: a systematic review. *J Am Med Inform Assoc* 2014; 21 (5): 942–51.
45. Luz CF, Vollmer M, Decruyenaere J, *et al.* Machine learning in infection management using routine electronic health records: tools, techniques, and reporting of future technologies. *Clin Microbiol Infect* 2020; 26 (10): 1291–9.
46. Moher D, Liberati A, Tetzlaff J, *et al.*; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009; 6 (7): e1000097.
47. Horng S, Sontag DA, Halpern Y, *et al.* Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. *PLoS One* 2017; 12 (4): e0174708.
48. Apostolova E, Velez T. Toward automated early sepsis alerting: identifying infection patients from nursing notes. In: BioNLP 2017. Association for Computational Linguistics August 4, 2017: 257–62; Vancouver, Canada. doi:10.18653/v1/W17-2332
49. Culliton P, Levinson M, Ehresman A, *et al.* Predicting Severe Sepsis Using Text from the Electronic Health Record. In: Workshop on Machine Learning For Health at the Conference on Neural Information Processing Systems (NIPS ML4H 2017), December 8, 2017. arXiv: 1711.11536; Long Beach, CA.
50. Liu R, Greenstein JL, Sarma SV, *et al.* Natural language processing of clinical notes for improved early prediction of septic shock in the ICU. *Annu Int Conf IEEE Eng Med Biol Soc* July 23–27, 2019; 2019: 6103–8; Berlin, Germany. doi:10.1109/EMBC.2019.8857819
51. Delahanty RJ, Alvarez J, Flynn LM, *et al.* Development and evaluation of a machine learning model for the early identification of patients at risk for sepsis. *Ann Emerg Med* 2019; 73 (4): 334–44.
52. Goh KH, Wang L, Yeow AYK, *et al.* Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare. *Nat Commun* 2021; 12 (1): 711.
53. Amrollahi F, Shashikumar SP, Razmi F, *et al.* Contextual embeddings from clinical notes improves prediction of sepsis. *AMIA Annu Symp Proc* November 14–18, 2020; 2020: 197–202; Virtual event.
54. Hammoud I, Ramakrishnan IV, Henry M, *et al.* Multimodal early septic shock prediction model using lasso regression with decaying response. In: 2020 IEEE International Conference on Healthcare Informatics (ICHI). IEEE November 30–December 3, 2020; Virtual conference in Oldenburg, Germany. doi:10.1109/ichi48887.2020.9374377
55. Qin F, Madan V, Ratan U, *et al.* Improving early sepsis prediction with multi modal learning, 2021. arXiv: 2107.11094
56. Bouam S, Girou E, Brun-Buisson C, *et al.* An intranet-based automated system for the surveillance of nosocomial infections: prospective validation compared with physicians' self-reports. *Infect Control Hosp Epidemiol* 2003; 24 (1): 51–5.
57. Koller W, Blacky A, Bauer C, *et al.* Electronic surveillance of healthcare-associated infections with MONI-ICU–a clinical breakthrough compared to conventional surveillance systems. *Stud Health Technol Inform* 2010; 160 (Pt 1): 432–6.
58. Penz JFE, Wilcox AB, Hurdle JF. Automated identification of adverse events related to central venous catheters. *J Biomed Inform* 2007; 40 (2): 174–82.
59. Proux D, Marchal P, Segond F, *et al.* Natural language processing to detect risk patterns related to hospital acquired infections. In: Proceedings of the Workshop on Biomedical Information Extraction: Association for Computational Linguistics; September 18, 2009: 35–41; Borovets, Bulgaria. doi:10.5555/1859776.1859782
60. Bouzbid S, Gicquel Q, Gerbier S, *et al.* Automated detection of nosocomial infections: evaluation of different strategies in an intensive care unit 2000-2006. *J Hosp Infect* 2011; 79 (1): 38–43.
61. Jo Y, Loghmanpour N, Rosé CP. Time series analysis of nursing notes for mortality prediction via a state transition topic model. In: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management (CIKM). Association for Computing Machinery October 18–23, 2015: 1171–80; Melbourne, Australia. doi:10.1145/2806416.2806541
62. Wang T, Velez T, Apostolova E, *et al.* Semantically Enhanced Dynamic Bayesian Network for Detecting Sepsis Mortality Risk in ICU Patients with Infection, 2018. arXiv: 1806.10174
63. Baghdadi Y, Bourrée A, Robert A, *et al.* Automatic classification of free-text medical causes from death certificates for reactive mortality surveillance in France. *Int J Med Inform* 2019; 131: 103915.
64. Guo W, Xu Z, Ye X, *et al.* A time-critical topic model for predicting the survival time of sepsis patients. *Sci Program* 2020; 2020: 1. doi:10.1155/2020/8884539
65. Ribas Ripoll VJ, Vellido A, Romero E, *et al.* Sepsis mortality prediction with the Quotient Basis Kernel. *Artif Intell Med* 2014; 61: 45–52.
66. Halpern Y, Horng S, Choi Y, *et al.* Electronic medical record phenotyping using the anchor and learn framework. *J Am Med Inform Assoc* 2016; 23 (4): 731–40.
67. Vilic A, Petersen JA, Hoppe K, *et al.* Visualizing patient journals by combining vital signs monitoring and natural language processing. *Annu Int Conf IEEE Eng Med Biol Soc* August 16–20, 2016; 2016: 2529–32; Orlando, FL. doi:10.1109/EMBC.2016.7591245
68. Zhu X, Plasek JM, Tang C, *et al.* Embedding, aligning and reconstructing clinical notes to explore sepsis. *BMC Res Notes* 2021; 14 (1): 136.
69. Marks M, Pollock E, Armstrong M, *et al.* Needles and the damage done: reasons for admission and financial costs associated with injecting drug use in a Central London Teaching Hospital. *J Infect* 2013; 66 (1): 95–102.
70. Ippolito P, Larson EL, Furuya EY, *et al.* Utility of electronic medical records to assess the relationship between parenteral nutrition and central line-associated bloodstream infections in adult hospitalized patients. *JPEN J Parenter Enteral Nutr* 2015; 39 (8): 929–34.
71. Liu YZ, Chu R, Lee A, *et al.* A surveillance method to identify patients with sepsis from electronic health records in Hong Kong: a single centre retrospective study. *BMC Infect Dis* 2020; 20 (1): 652.
72. Murff HJ, FitzHenry F, Matheny ME, *et al.* Automated identification of postoperative complications within an electronic medical record using natural language processing. *JAMA* 2011; 306 (8): 848–55.
73. FitzHenry F, Murff HJ, Matheny ME, *et al.* Exploring the frontier of electronic health record surveillance: the case of postoperative complications. *Med Care* 2013; 51 (6): 509–16.
74. Piscitelli A, Bevilacqua L, Labella B, *et al.* A keyword approach to identify adverse events within narrative documents from 4 Italian institutions. *J Patient Saf*. doi:10.1097/PTS.0000000000000783
75. Vermassen J, Colpaert K, De Bus L, *et al.* Automated screening of natural language in electronic health records for the diagnosis septic shock is feasible and outperforms an approach based on explicit administrative codes. *J Crit Care* 2020; 56: 203–7.
76. Saeed M, Villarroel M, Reisner AT, *et al.* Multiparameter intelligent monitoring in intensive care II: a public-access intensive care unit database. *Crit Care Med* 2011; 39 (5): 952–60.
77. Johnson AEW, Pollard TJ, Shen L, *et al.* MIMIC-III, a freely accessible critical care database. *Sci Data* 2016; 3: 160035.
78. Henry KE, Hager DN, Pronovost PJ, *et al.* A targeted real-time early warning score (TREWScore) for septic shock. *Sci Transl Med* 2015; 7 (299): 299ra122.
79. Angus DC, Linde-Zwirble WT, Lidicker J, *et al.* Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. *Crit Care Med* 2001; 29 (7): 1303–10.

80. Rhee C, Dantes R, Epstein L, *et al.*; CDC Prevention Epicenter Program. Incidence and trends of sepsis in US hospitals using clinical vs claims data, 2009-2014. *JAMA* 2017; 318 (13): 1241–9.
81. Reyna MA, Josef C, Seyedi S, *et al.* Early prediction of sepsis from clinical data: the PhysioNet/Computing in Cardiology Challenge 2019. In: 2019 Computing in Cardiology (CinC) September 8–11, 2019: 1–4; Singapore. doi:10.23919/CinC49843.2019.9005736
82. Wagner MM, Hogan WR, Chapman WW, *et al.* Chief complaints and ICD codes. *Handbook of Biosurveillance* 2006: 333.
83. Lucian J., Davis JFM. History and physical examination. *Murray and Nadel's Textbook of Respiratory Medicine* 2016: 263.doi:10.1016/B978-1-4557-3383-5.00016-6
84. Aghili H, Mushlin RA, Williams RM, *et al.* Progress notes model. In: Proceedings of AMIA Annual Fall Symposium October 25–29, 1997: 12–6; Nashville, TN.
85. Hellesø R. Information handling in the nursing discharge note. *J Clin Nurs* 2006; 15 (1): 11–21.
86. Ioanna P, Stiliani K, Vasiliki B. Nursing documentation and recording systems of nursing care. *Health Sci J* 2007; (4): 7.https://www.hsj.gr/medicine/nursing-documentation-and-recording-systems-of-nursing-care.php?aid=3680 Accessed June 29, 2021.
87. Flink M, Ekstedt M. Planning for the discharge, not for patient self-management at home - an observational and interview study of hospital discharge. *Int J Integr Care* 2017; 17 (6): 1.
88. Stopford E, Ninan S, Spencer N. How to write a discharge summary. *BMJ* 2015; 351: h2696.
89. Sørby ID, Nytrø Ø. Does the electronic patient record support the discharge process? A study on physicians' use of clinical information systems during discharge of patients with coronary heart disease. *Health Inf Manag* 2006; 34 (4): 112–9.
90. Kind AJH, Smith MA. Documentation of mandated discharge summary components in transitions from acute to subacute care. In: Henriksen K, Battles JB, Keyes MA, *et al.*, eds. *Advances in Patient Safety: New Directions and Alternative Approaches (Vol. 2: Culture and Redesign. AHRQ Publication No. 08-0034-2)*. Rockville, MD: Agency for Healthcare Research and Quality (US); 2011.
91. Jensdóttir A-B, Jónsson P, Noro A, *et al.* Comparison of nurses' and physicians' documentation of functional abilities of older patients in acute care–patient records compared with standardized assessment. *Scand J Caring Sci* 2008; 22 (3): 341–7.
92. Penoyer DA, Cortelyou-Ward KH, Noblin AM, *et al.* Use of electronic health record documentation by healthcare workers in an acute care hospital system. *J Healthc Manag* 2014; 59: 130–44.
93. Furniss D, Lyons I, Franklin BD, *et al.* Procedural and documentation variations in intravenous infusion administration: a mixed methods study of policy and practice across 16 hospital trusts in England. *BMC Health Serv Res* 2018; 18 (1): 270.
94. Sohn S, Wang Y, Wi C-I, *et al.* Clinical documentation variations and NLP system portability: a case study in asthma birth cohorts across institutions. *J Am Med Inform Assoc* 2018; 25 (3): 353–9.
95. McPherson K. International differences in medical care practices. *Health Care Financ Rev* 1989; 1989: 9–20.
96. Georgopoulos BS. Organization structure and the performance of hospital emergency services. *Ann Emerg Med* 1985; 14 (7): 677–84.
97. McKinlay J, Link C, Marceau L, *et al.* How do doctors in different countries manage the same patient? Results of a factorial experiment. *Health Serv Res* 2006; 41 (6): 2182–200.
98. Sørby ID, Nytrø Ø. Analysis of communicative behaviour: profiling roles and activities. *Int J Med Inform* 2010; 79 (6): e144-51–e151.
99. Røst TB, Tvedt CR, Husby H, *et al.* Identifying catheter-related events through sentence classification. *Int J Data Min Bioinform* 2020; 23 (3): 213–33.
100. Harris ZS. Distributional Structure. *Word World* 1954; 10 (2–3): 146–62.
101. Le Q, Mikolov T. Distributed representations of sentences and documents. In: *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32*. JMLR.org June 21–26, 2014: II-1188–96; Beijing, China. doi:10.5555/3044805.3045025
102. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. *J Mach Learn Res* 2003; 3: 993–1022. doi:10.5555/944919.944937
103. Mikolov T, Sutskever I, Chen K, *et al.* Distributed Representations of Words and Phrases and their Compositionality. 2013. arXiv:1310.4546
104. Pennington J, Socher R, Manning C. GloVe: Global Vectors for Word Representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics October 25–29, 2014: 1532–43; Doha, Qatar. doi:10.3115/v1/D14-1162
105. Mikolov T, Chen K, Corrado G, *et al.* Efficient Estimation of Word Representations in Vector Space. 2013. arXiv: 1301.3781
106. Devlin J, Chang M-W, Lee K, *et al.* BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Association for Computational Linguistics June 2–7, 2019: 4171–86; Minneapolis, MN. doi:10.18653/v1/N19-1423
107. Alsentzer E, Murphy J, Boag W, *et al.* Publicly available clinical BERT Embeddings. In: Proceedings of the 2nd Clinical Natural Language Processing Workshop. Association for Computational Linguistics June 7, 2019: 72–8; Minneapolis, MN. doi:10.18653/v1/W19-1909
108. Koller D, Friedman N. *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, MA and London, England: MIT Press, 2009.
109. Cortes C, Vapnik V. Support-vector networks. *Mach Learn* 1995; 20 (3): 273–97.
110. Fix E, Hodges JL. Nonparametric discrimination: consistency properties. *Randolph Field, Texas, Project* 1951: 21–49.
111. Breiman L. Random forests. *Mach Learn* 2001; 45 (1): 5–32.
112. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat* 1999; 29: 1189–232. http://www.jstor.org/stable/2699986 Accessed May 3, 2021.
113. Friedman JH. Stochastic gradient boosting. *Comput Stat Data Anal* 2002; 38 (4): 367–78.
114. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature* 1986; 323 (6088): 533–6.
115. Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation In: Rumelhart DE, McClelland JL, eds. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*. MIT Press 1986: 318–62. https://ieeexplore.ieee.org/document/6302929 Accessed May 3, 2021.
116. Mikolov T, Karafiát M, Burget L, *et al.* Recurrent neural network based language model. In: *INTERSPEECH* 2010. ISCA September 26–30, 2010; Makuhari, Chiba, Japan.doi: 10.21437/Interspeech.2010-343
117. Cho K, van Merrienboer B, Gulcehre C, *et al.* Learning phrase representations using RNN Encoder-Decoder for Statistical Machine Translation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics October 25–29, 2014: 1724–34; Doha, Qatar. doi:10.3115/v1/D14-1179
118. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997; 9 (8): 1735–80.
119. Cawcutt KA, Peters SG. Severe sepsis and septic shock: clinical overview and update on management. *Mayo Clin Proc* 2014; 89 (11): 1572–8.
120. Hotchkiss RS, Moldawer LL, Opal SM, *et al.* Sepsis and septic shock. *Nat Rev Dis Primers* 2016; 2: 16045.
121. Bone RC, Balk RA, Cerra FB, *et al.* Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine. *Chest* 1992; 101 (6): 1644–55.
122. Vincent JL, Moreno R, Takala J, *et al.* The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med* 1996; 22 (7): 707–10.

123. Seymour CW, Liu VX, Iwashyna TJ, *et al.* Assessment of clinical criteria for sepsis: for the third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA* 2016; 315 (8): 762–74.
124. Subbe CP, Kruger M, Rutherford P, *et al.* Validation of a modified early warning score in medical admissions. *QJM* 2001; 94 (10): 521–6.
125. Rhodes A, Evans LE, Alhazzani W, *et al.* Surviving sepsis campaign: international guidelines for management of sepsis and septic shock: 2016. *Intensive Care Med* 2017; 43 (3): 304–77.
126. Dellinger RP, Levy MM, Rhodes A, *et al.*; Surviving Sepsis Campaign Guidelines Committee including the Pediatric Subgroup. Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2012. *Crit Care Med* 2013; 41 (2): 580–637.
127. Vincent J-L. The clinical challenge of sepsis identification and monitoring. *PLoS Med* 2016; 13 (5): e1002022.
128. Abe T, Tokuda Y, Shiraishi A, *et al.*; JAAM SPICE Study Group. In-hospital mortality associated with the misdiagnosis or unidentified site of infection at admission. *Crit Care* 2019; 23 (1): 202.
129. Rothberg MB, Pekow PS, Priya A, *et al.* Variation in diagnostic coding of patients with pneumonia and its association with hospital risk-standardized mortality rates: a cross-sectional analysis. *Ann Intern Med* 2014; 160 (6): 380–8.
130. Bewick T, Simmonds M, Chikhani M, *et al.* Pneumonia in the context of severe sepsis: a significant diagnostic problem. *Eur Respir J* 2008; 32 (5): 1417–8.
131. Rhee C, Murphy MV, Li L, *et al.*; for the Centers for Disease Control and Prevention Epicenters Program. Comparison of trends in sepsis incidence and coding using administrative claims versus objective clinical data. *Clin Infect Dis* 2015; 60 (1): 88–95.
132. Tidswell R, Inada-Kim M, Singer M. Sepsis: the importance of an accurate final diagnosis. *Lancet Respir Med* 2021; 9 (1): 17–8.
133. Lopansri BK, Miller RR III, Burke JP, *et al.* Physician agreement on the diagnosis of sepsis in the intensive care unit: estimation of concordance and analysis of underlying factors in a multicenter cohort. *J Intensive Care* 2019; 7: 13.
134. Rhee C, Kadri SS, Danner RL, *et al.* Diagnosing sepsis is subjective and highly variable: a survey of intensivists using case vignettes. *Crit Care* 2016; 20: 89.
135. Rhee C, Klompas M. Sepsis trends: increasing incidence and decreasing mortality, or changing denominator? *J Thorac Dis* 2020; 12 (Suppl 1): S89–100.
136. Yu SC, Betthauser KD, Gupta A, *et al.* Comparison of sepsis definitions as automated criteria. *Crit Care Med* 2021; 49 (4): e433–43.
137. Liu R, Greenstein JL, Granite SJ, *et al.* Data-driven discovery of a novel sepsis pre-shock state predicts impending septic shock in the ICU. *Sci Rep* 2019; 9 (1): 6145.
138. Walkey AJ, Shieh M-S, Liu VX, *et al.* Mortality measures to profile hospital performance for patients with septic shock. *Crit Care Med* 2018; 46 (8): 1247–54.
139. Vincent J-L. The challenge of early identification of the hospital patient at risk of septic complications. *Ann Transl Med* 2017; 5 (3): 56.
140. Mendonça EA, Haas J, Shagina L, *et al.* Extracting information on pneumonia in infants using natural language processing of radiology reports. *J Biomed Inform* 2005; 38 (4): 314–21.
141. Dublin S, Baldwin E, Walker RL, *et al.* Natural language processing to identify pneumonia from radiology reports. *Pharmacoepidemiol Drug Saf* 2013; 22 (8): 834–41.
142. Matheny ME, Fitzhenry F, Speroff T, *et al.* Detection of infectious symptoms from VA emergency department and primary care clinical documentation. *Int J Med Inform* 2012; 81 (3): 143–56.
143. Wu S, Roberts K, Datta S, *et al.* Deep learning in clinical natural language processing: a methodical review. *J Am Med Inform Assoc* 2020; 27 (3): 457–70.
144. Khattak FK, Jeblee S, Pou-Prom C, *et al.* A survey of word embeddings for clinical text. *J Biomed Informatics: X* 2019; 100: 100057.
145. Névéol A, Zweigenbaum P; Section Editors for the IMIA Yearbook Section on Clinical Natural Language Processing. Expanding the diversity of texts and applications: findings from the section on clinical natural language processing of the international medical informatics association yearbook. *Yearb Med Inform* 2018; 27 (1): 193–8.
146. Koleck TA, Dreisbach C, Bourne PE, *et al.* Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. *J Am Med Inform Assoc* 2019; 26 (4): 364–79.
147. Wang HE, Shapiro NI, Griffin R, *et al.* Chronic medical conditions and risk of sepsis. *PLoS One* 2012; 7 (10): e48307.
148. Prescott HC, Carmichael AG, Langa KM, *et al.* Paths into sepsis: trajectories of presepsis healthcare use. *Ann Am Thorac Soc* 2019; 16 (1): 116–23.
149. Delano MJ, Ward PA. The immune system's role in sepsis progression, resolution, and long-term outcome. *Immunol Rev* 2016; 274 (1): 330–53.
150. Faisal M, Scally A, Richardson D, *et al.* Development and external validation of an automated computer-aided risk score for predicting sepsis in emergency medical admissions using the patient's first electronically recorded vital signs and blood test results. *Crit Care Med* 2018; 46 (4): 612–8.

Paper B

Preliminary Processing and Analysis of an Adverse Event Dataset for Detecting Sepsis-Related Events

Melissa Y. Yan, Lise Husby Høvik, André Pedersen, Lise Tuset Gustad, and Øystein Nytrø

B

Published in:
IEEE International Conference on Bioinformatics and Biomedicine, (BIBM), December 2021, pp. 1605–1610. DOI: 10.1109/BIBM52615.2021.9669410.

Preliminary Processing and Analysis of an Adverse Event Dataset for Detecting Sepsis-Related Events

Melissa Y. Yan*
Norwegian University of Science and Technology
Trondheim, Norway
melissa.yan@ntnu.no

Lise Husby Høvik*
St. Olavs hospital, Trondheim University Hospital
Trondheim, Norway
lise.hovik@ntnu.no

André Pedersen
Norwegian University of Science and Technology
SINTEF
Trondheim, Norway
andre.pedersen@sintef.no

Lise Tuset Gustad
Norwegian University of Science and Technology
Trondheim, Norway
lise.t.gustad@ntnu.no

Øystein Nytrø
Norwegian University of Science and Technology
Trondheim, Norway
nytroe@ntnu.no

***Abstract*—Adverse event (AE) reports contain notes detailing procedural and guideline deviations, and unwanted incidents that can bring harm to patients. Available datasets mainly focus on vigilance or post-market surveillance of adverse drug reactions or medical device failures. The lack of clinical-related AE datasets makes it challenging to study healthcare-related AEs. AEs affect 10% of hospitalized patients, and almost half are preventable. Having an AE dataset can assist in identifying possible patient safety interventions and performing quality surveillance to lower AE rates. The free-text notes can provide insight into the cause of incidents and lead to better patient care. The objective of this study is to introduce a Norwegian AE dataset and present preliminary processing and analysis for sepsis-related events, specifically peripheral intravenous catheter-related bloodstream infections. Therefore, the methods focus on performing a domain analysis to prepare and better understand the data through screening, generating synthetic free-text notes, and annotating notes.**

***Index Terms*—Adverse events, Healthcare knowledge representation, Natural language processing, Quality improvement, Sepsis**

I. Introduction

Sepsis is the most common cause of death among hospitalized patients [1] and contributes to 30% to 50% of hospitalized deaths [2]. Caused by a dysregulated host response to an infection, sepsis can lead to multi-organ failure and death [3]. Bloodstream infections (BSIs) occur when bacteria enter the bloodstream [4]. A particularly lethal bacterium that commonly causes BSIs is *Staphylococcus aureus* (*S. aureus*) [5], a gram-positive bacteria frequently found on the skin. A range from 7.6% to 35% of *S. aureus* BSIs are due to peripheral intravenous catheters (PIVCs) [6]. PIVCs are inserted in a peripheral vein to administer IV fluids, medications, and blood transfusions. They are the most frequently used device in hospitals [7]; over one billion PIVCs are estimated to be inserted annually worldwide [8] and up to 80% of patients admitted to a hospital will receive at least one PIVC [9].

Improper management of PIVCs can lead to increased patient mortality risk via BSIs [6]. Four gateways to BSIs are described with PIVCs; migration of microbes down the catheter tract, via the catheter hub, by contaminated infusate, or by an existing infection where bacteria circulating the bloodstream can attach to the catheter [10]. Although frequently used, PIVCs are often not documented in clinical records [7]. Additionally, sepsis is poorly documented in departments outside the intensive care unit [11]. This makes retrospective and real-time systematic quality surveillance for PIVC difficult. However, failure related to PIVCs are more frequently reported in AE reports. Hence, the main motivation of this project was to use an AE dataset to facilitate systematic monitoring and quality of care improvements related to PIVCs for reducing sepsis and BSI cases.

II. Background

A. About Adverse Events

An estimated 1 in 10 hospitalized patients worldwide are affected by an adverse event (AE), and nearly 50% are preventable [12]. Commonly reported AEs include surgical or medication procedural errors, hospital-acquired infections, pressure ulcers, and falls [13]. AE incidents can be recorded in electronic health records (EHRs) or separate reporting systems. Current methods for detecting AEs include manual chart review and screening using ICD codes, keyword search, and natural language processing (NLP) [14]. Using the Global Trigger tool [15], various studies track health care quality indicators to identify triggers and measure AE rates by manually reviewing medical records [13]. In addition to the manual approach, other studies focus on developing automated trigger tools, such as extracting EHR data using NLP [16] and monitoring nursing notes for infection signs [17].

* Authors contributed equally to this work.

Yan, M.Y., Høvik, L.H., Pedersen, A., Gustad, L.T., Nytrø, Ø, Preliminary Processing and Analysis of an Adverse Event Dataset for Detecting Sepsis-Related Events
IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2021, Houston, TX, USA, December 9-12, 2021, pages 1605–1610. IEEE, December 2021

TABLE I. ADVERSE EVENT DATABASES (DB) WORLDWIDE

Type	Purpose	Database
Drug	Vigilance	Canada Vigilance Adverse Reaction Online DB; EU Drug Regulating Authorities Pharmacovigilance (EudraVigilance); German ABDA[a] DB; Japanese Adverse Drug Event Report (JADER) DB; Korean Adverse Event Reporting System (KAERS); UK MHRA[b] Interactive Drug Analysis Profiles (iDAPs); World Health Organization (WHO) VigiBase via VigiAccess
Drug	Post-market surveillance	US FDA[c] Adverse Event Reporting System (FAERS) DB; US Vaccine Adverse Event Reporting System (VAERS)
Device	Post-market surveillance	German Medical Devices Information and DB System (DMIDS); US Manufacturer and User Facility Device Experience (MAUDE)
Drug & Device	Post-market surveillance	Australian DB of Adverse Event Notifications (DAEN)
All	Near-miss or AE	Japan Council for Quality Health Care (JQ) project

[a] ABDA: Federal Union of German Associations of Pharmacists (Bundesvereinigung Deutscher Apothekerverbände)
[b] MHRA: Medicines and Healthcare products Regulatory Agency
[c] FDA: Food and Drug Administration

B. Purpose of Adverse Events

The primary purpose of documenting AEs is to promote patient safety. Along with collaborators worldwide, the World Health Organization has been promoting methods which contribute to effectively learning from AEs [18]. As AE documentation shifts from legal consequences with personal responsibility towards a learning perspective, it is becoming possible to better understand the causes resulting in AEs, and thus identify possible interventions to improve patient safety within hospitals [12].

C. Available Datasets and Databases

Available AE datasets and databases worldwide mainly focus on vigilance or post-market surveillance of drugs or devices (see Table I). As the focus is on either adverse drug events and side effects or device failures, it is unlikely that many of the databases will capture clinical healthcare-related AEs. To the best of our knowledge, the only publicly available AE database containing clinical healthcare-related AEs is the Japan Council for Quality Health Care's Project to Collect Medical Near-Miss/Adverse Event Information [19].

However, the Japanese healthcare system varies from Norwegian healthcare in terms of treatment, health system organization, and strategies to ensure the quality of care [20]. For instance, Norway has general practitioners who act as gatekeepers to specialist treatment; this is relatively new in Japan. In addition, most Norwegian hospitals are government-owned, whereas only 15% of Japanese hospitals are government-owned. Furthermore, Norwegian hospitals are obligated to participate and measure quality and safety improvement, and there is a national program for tracking health care indicators of survival and infection rates. In contrast, the Japanese government promotes hospitals to report quality indicators on their websites, and only advanced treatment Japanese hospitals are required to report AEs.

D. Objective

This paper aims to present a Norwegian AE dataset and preliminary results for characterizing a dataset on detecting sepsis-related events, to demonstrate further research potential using a dataset currently undergoing preparation for release. Various clinical events were found by inspecting the AE dataset with the initial motivation of identifying PIVC-related BSIs. This included events related to sepsis and phlebitis, which is inflammation of a vein near the skin's surface and can be an indicator of infection. Furthermore, falls and device failures were deemed relevant to interpret the AE dataset.

III. NORWEGIAN ADVERSE EVENTS

In Norway, a retrospective review of EHRs estimated that one-third of all hospital deaths were due to AEs [21]. Further exploration into two Norwegian hospitals identified that 11.2% of AEs were life-shortening. From these, 82.4% of the incidences were related to healthcare-associated infection. In general, comparing statistics based on AE-events from registries, EHR reviews, and automated methods is challenging.

A. Health Care Policy

Many countries have their own AE-related legislation to monitor the safety of drugs and medical devices. In Norway, the Regulations on Medicinal Products maintain drug safety and marketing[1], whereas the Medical Equipment Act regulates medical equipment safety and post-market surveillance[2]. Additionally, under the Norwegian Specialized Health Services Act of 1999, all health and care services are obligated to notify the Norwegian Board of Health Supervision of unexpected incidents related to patient injury and death[3]. Furthermore, according to the Regulations on Management and Quality Improvement in the Health and Care Service, those services are required to manage quality improvement and patient safety systematically by reviewing deviations (i.e., AE), evaluating implemented preventative measures, and rectifying activities[4].

B. Norwegian Adverse Event Dataset Description

There are 18 555 AE reports from the electronic incident reporting registry system at St. Olavs hospital, Trondheim University Hospital in Trondheim, Norway between September 30, 2015 and December 31, 2019. Intentionally written for a specific purpose and directed at someone specific, these reports are not written routinely by a clinician and differ in quality, purpose, and structure from EHR clinical notes. These reports describe various events in addition to AEs, such as procedural and guideline deviations, near-miss events that could have

[1] https://lovdata.no/dokument/LTI/forskrift/2009-12-18-1839
[2] https://lovdata.no/dokument/NL/lov/2020-05-07-37
[3] https://lovdata.no/dokument/NL/lov/1999-07-02-61
[4] https://lovdata.no/dokument/SF/forskrift/2016-10-28-1250

Yan, M.Y., Høvik, L.H., Pedersen, A., Gustad, L.T., Nytrø, Ø, Preliminary Processing and Analysis of an Adverse Event Dataset for Detecting Sepsis-Related Events
IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2021, Houston, TX, USA, December 9-12, 2021, pages 1605–1610. IEEE, December 2021

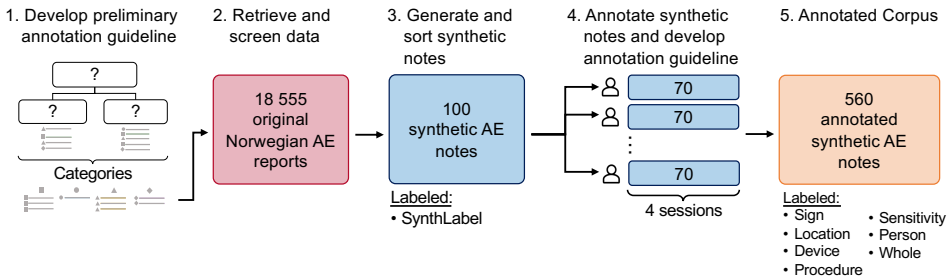


Fig. 1. Steps for preprocessing adverse event (AE) data. **1)** The preliminary guideline was developed from clinical questions of interest, which are used to create categories. **2)** The categories are used to screen the original 18 555 Norwegian AE reports. **3)** A total of 100 unique synthetic AE notes were generated based on notes from the original reports, and each note was given a **SynthLabel** label indicating if it contained infection, faulty device malfunctioning, or fall information. **4)** Eight annotators each annotated 70 notes over four sessions using a guideline that was revised after each session. **5)** This resulted in 560 annotated synthetic AE notes labeled with seven categories (i.e., **Sign**, **Location**, **Device**, **Procedure**, **Sensitivity**, **Person**, and **Whole**) used to capture and represent documented healthcare knowledge.

harmed patients, misunderstandings, resource needs, and patients with poor behavior who pose a risk to others. Each report has: an identifier, title, registration date, changed date, report to and from units, booleans for security-related or patient-related event, event type and severity, clinical division, and an unstructured free-text note. Lastly, it also contains a status indicating if the incident is open or closed; a closed status indicates specific solutions for patient safety problems have been developed and implemented. Use of AE notes for the purpose of this study was approved by the Norwegian Regional Committees for Medical and Health Research Ethics (REK), approval no 26814.

IV. Materials and Methods

The original 18 555 Norwegian AE reports dataset was used to create a synthetic dataset annotated for PIVC-related BSI events (see Fig. 1). The synthetic notes were annotated to capture data, information, and knowledge in the text at different levels; word- or phrase-level indicates an annotation that spans a word or phrase, whereas note-level indicates an annotation representing the span of the whole text. This resulted in 100 synthetic notes with **SynthLabel** note-level labels (i.e., infection-related, faulty device malfunction-related, and fall-related incidents) and 560 annotated synthetic notes.

A. Preliminary Annotation Guideline Development

The preliminary annotation guideline was developed based on the proposed clinical question: "Is there a connection between BSIs and PIVCs at the hospital?" The clinical question was simplified to:

- How can sepsis or BSIs be identified when the symptoms are similar to other diseases?
- How can poorly documented PIVCs be identified?

Those questions were then modified based on the clinical perspectives of the nurses; for example, some catheters are documented distinctly (for data extraction), whereas others can be distinguished based on anatomical insertion site (for information extraction) or procedures (for knowledge extraction). This resulted in the following domain-specific *questions of interest*:

- What are the different signs of infections, specifically for BSIs, sepsis, or infected PIVCs?
- What are the signs for different types of catheters?
- Where are the anatomical insertion sites of catheters?
- What events can be related to catheter use?

Domain-specific questions were answered by nurses who provided a list of keywords, phrases, sentences, and examples from the clinic. As shown in Fig. 2(a), these answers were sorted into four categories (technically known as entities in annotation or classes in ontologies) for word- or phrase-level labels: **Sign**, **Location**, **Device**, and **Procedure**. Next, a total of 700 randomly selected notes, from the original AE dataset, were manually screened to ensure that the four categories related to catheters and BSIs could be found and occurred frequently enough for downstream analysis (see Fig. 2(c)). After screening, three additional categories (i.e., **Sensitivity**, **Person**, and **Whole**) were included to ensure that sensitive data was correctly anonymized, actions related to an individual could be determined, and a note-level label was available. This resulted in seven categories:

1) **Sign**: infection signs
2) **Location**: anatomical insertion sites
3) **Device**: signs of catheter types
4) **Procedure**: procedures, interventions, or activities related to catheters
5) **Sensitivity**: protected health information
6) **Person**: individuals (i.e., patient, clinician, or relative)
7) **Whole**: note-level label indicating whether the note contains infection, BSI, sepsis, faulty device malfunctioning, catheter, PIVC, or sensitive information.

Each category can form a hierarchy with more specific subcategories (e.g., the **Device** category contains a general subcategory "catheter" that has a more specific "PIVC" subcategory). In addition, six relationships (see Fig. 2(b)) were added to link categories 1-4 together to ensure that information was not lost for downstream analysis (e.g., infection sign at a specific

Yan, M.Y., Høvik, L.H., Pedersen, A., Gustad, L.T., Nytrø, Ø, Preliminary Processing and Analysis of an Adverse Event Dataset for Detecting Sepsis-Related Events
IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2021, Houston, TX, USA, December 9-12, 2021, pages 1605–1610. IEEE, December 2021

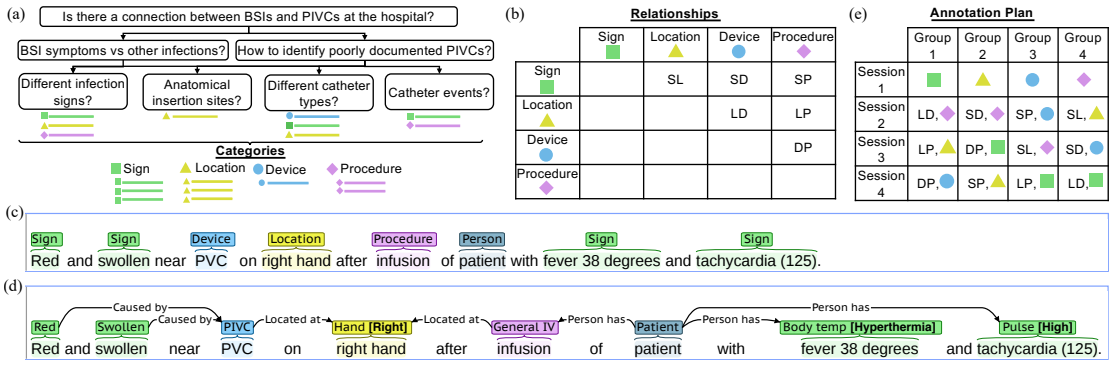


Fig. 2. Annotation guideline development and annotation. (a) The clinical question of interest was simplified into domain-specific *questions of interest* which were answered by clinicians and sorted into four different main categories (i.e., **Sign**, **Location**, **Device**, and **Procedure**). (b) To capture knowledge about peripheral intravenous catheters (PIVCs) and bloodstream infections for downstream analysis, relationships linking categories to each other were included in the guideline. There are six relationships: sign-location, sign-device, sign-procedure, location-device, location-procedure, and device-procedure (i.e., SL, SD, SP, LD, LP, and DP). (c) Randomly selected adverse event notes were manually screened to ensure that the four main categories were detectable. (d) During annotation, annotators used the Brat rapid annotation tool to label notes in more detail using subcategories and attributes. For instance, "right hand" which was previously labeled as **Location** in (c) in now labeled using **Location**'s subcategory "Hand" and given the attribute [Right]. Relationships are also used to link one label to another (e.g., "Red" is linked to "PIVC" using the sign-device relationship "Caused by"). (e) Categories and relationships were sorted to create an annotation plan with four sessions and four groups of two annotators each.

location). Fig. 2(d) provides an example of how relationships link categories together and how detailed information can be provided by using subcategories and attributes. Using categories, relationships, and screening results as examples and counterexamples, a preliminary annotation guideline was created. The preliminary annotation guideline describes how to annotate each category and relationship to remove annotator confusion and disagreements.

B. Synthetic Adverse Event Dataset Generation

The 100 synthetic notes were generated and validated by a nurse, and thereafter divided into 10 sets with 10 notes each for the four main categories and six possible relationships. The 10 sets were sorted into four groups such that the sets for the four main categories were annotated by each group once and the sets for the six relationships were annotated at least twice by a different group. This was done to assess guideline revision improvements among different annotators using the same set of notes. The combination of these sets resulted in an annotation plan with four annotation sessions and four groups each with two annotators (see Fig. 2(e)). Thus, each annotator would annotate 10 notes in the first session and 20 notes in the remaining three sessions for a total of 70 notes.

C. Annotation Guideline Development and Annotation

Synthetic notes were annotated in four annotation sessions. In each session, two annotators annotated notes using the annotation guideline and Brat rapid annotation tool (BRAT) [22]. Annotations were evaluated by group using the inter-annotator agreement (IAA) F_1-score and assessed for whether clinical question information was captured. Then, ambiguities and annotator comments were discussed with nurses and incorporated into annotation guidelines revisions, and the process was repeated (guidelines for each session are available online[5]).

D. Annotated Dataset

Annotation by eight annotators produced a dataset of 560 notes stored in the BRAT standoff format. Each note has a note-level label (i.e., **Whole**). All AE notes can have overlapping note-level topics. Additionally, each note can have word- or phrase-level labels for the remaining six categories (i.e., **Sign**, **Location**, **Device**, **Procedure**, **Sensitivity**, and **Person**). Each word- or phrase-level label can have additional attribute information and can be linked to other labels to form relationships.

E. Preprocessing

For purposes of this study, only note-level labels were used. The 560 notes were converted into a comma-separated value file. The file contains basic information for each note, such as annotation session number, annotator identifier, filename, and text. Annotations included all word- or phrase-level labels from the seven categories and any annotator provided attributes. **Whole** category labels were separated into individual columns to identify note-level labels indicating whether a note contains infection, BSI, sepsis, faulty device malfunction, catheter, PIVC, or sensitive information. In addition, **SynthLabel** note-level labels were also separated into individual columns for infection, faulty device malfunction, or fall. Two additional merged labels, "Merged Infections" and "Merged Device Fails", were formed by combining parts of note-level labels **Whole** and **SynthLabel**. "Merged Infections" was comprised of **SynthLabel** label infection-related

[5]https://folk.ntnu.no/melissay/ae-guidelines/

and **Whole** category labels infection-related, BSI-related, and sepsis. "Merged Device Fails" was comprised of **SynthLabel** label device malfunction-related and **Whole** category labels device malfunction.

As multiple annotators labeled the same note, a max-voting strategy was conducted to produce a ground truth. No ties occurred between annotators (e.g., four annotators assigned 1 and four other annotators assigned 0).

F. Experiment

To assess the usefulness of the annotated dataset, experiments were conducted on a selection of tasks using a machine learning pipeline. Firstly, two datasets were defined: the training set of 18 555 original Norwegian notes and the test set of 560 annotated synthetic notes.

Each note was preprocessed using the following pipeline: 1) The common, redundant phrase "Hele_Notater" and other stop words were removed. 2) Capitalization was converted to lowercase. 3) Redundant characters such as newlines and quotation marks were removed. 4) Rare words with less than three occurrences were discarded. 5) Only notes with more than n_{min} and less than n_{max} words were kept. 6) Notes having less than l_{min} characters were discarded.

The following topic analysis pipeline was used to perform classification: 1) A word count vectorization was applied, keeping the top N words. Only unigrams and bigrams were generated. 2) Latent Dirichlet Allocation (LDA) [23] was then applied using K number of topics, trained concurrently using 16 workers for M iterations. 3) The word vectorizer and the LDA model were then trained using the training set only. 4) As LDA is an unsupervised method, it does not produce classification labels directly. Hence, the topic with the highest overlap with the task's labels in the test set was assigned for each respective task. This enabled evaluation of the unsupervised pipeline without manually choosing which topic(s) corresponded to each respective task(s), which is infeasible for a large number of topics.

Manually tuning relevant hyperparameters such as the number of topics K and the number of iterations M for the LDA model is challenging. Thus, an automatic hyperparameter search utilizing Bayesian optimization was conducted for 1000 iterations. To initialize the Bayesian search, the first 20 iterations were a random search. The test set's macro-averaged F_1-score was used as the objective function.

Models were trained using an Intel Core Processor with 32 cores and 128 GB of RAM. Implementation was done in Python 3.6. The topic model and feature extractor were implemented using scikit-learn (v0.16.1) [24]. Bayesian hyperparameter optimization was conducted using scikit-optimize (v0.8.1) [25]. The source code used in this study is made openly available on GitHub[6].

V. Preliminary Results

The resultant F_1-scores varied considerably between tasks (see Table II). Overall, the modeling technique performed well on all tasks, but performed best on the Fall and the Catheters tasks. It performed poorer on the infection and the merged tasks.

TABLE II. Test set performance of a selection of tasks using the hyperparameters chosen by the Bayesian optimization.

Task	F_1-score	Hyperparameters				
		N[a]	K[b]	n_{min}[c]	n_{max}[d]	l_{min}[e]
Infection	0.791	9268	35	1	48	34
Fall	0.997	5405	13	1	29	48
Device failure	0.895	1000	100	10	45	15
PIVC	0.877	10000	87	8	39	50
Catheters	1.000	1000	37	1	31	50
Merged Infections	0.843	4072	40	5	23	22
Merged Device Fails	0.743	1050	26	9	43	27

[a] N most frequent occurring words. [b] K number of topics. [c] n_{min} lower bound for number of words in a note. [d] n_{max} upper bound for number of words in a note. [d] l_{min} lower bound for number of characters in a note.

Hyperparameters chosen by the Bayesian search also differed between tasks, but the results had some patterns. Optimal performance on individual tasks was achieved using different sets of hyperparameters. Hence, using a single model for all tasks would result in overall degraded performance on individual tasks. Having a large number of topics K and a large number of words N were beneficial for detecting rarer and likely more challenging tasks.

VI. Discussion

This study presents a new AE dataset, a corresponding annotated dataset for PIVC-related BSIs, and preliminary data characterization results. The dataset is currently in development, but the plan is to make it openly available in the future. Initial experiments using a machine learning technique on a selection of tasks showed promising results.

The original 18 555 Norwegian AE reports dataset from a large representative university hospital are intended to be processed by hospital administration for quality improvement instead of responsibility, legal, or commercial reasons. The partially structured reports are written by health care personnel in complete sentences; this differs greatly from EHR clinical text, which are grammatically incomplete and brief [26]. The AE data may thus highlight patient safety issues that require addressing at an organizational or local level as well as drive national policy. Hence, this clinical dataset differs in quality, purpose, and structure from EHR clinical text.

To prepare for PIVC-related BSI studies using reasoning tasks and supervised machine learning, this study developed an annotation guideline and a corresponding annotated corpus which represents and captures PIVCs and BSIs documented in AE notes. Further work is required to develop an ontology based on the guideline as a framework to test the representation and reasoning about PIVC-related BSI. There are plans to develop PIVC-related BSI classifiers using word- and phrase-level annotations. Additionally, a previous study predicted central venous catheter events using sentences from clinical text with limited training data [26]. Thus, the annotated synthetic

[6]https://github.com/andreped/adverse-events

data will be preprocessed further to easily use sentence-level annotations for detecting sepsis-related events.

Due to the limited dataset size, unsupervised methods like LDA were preferred, as they tend to be more robust on smaller datasets. However, hyperparameter selection in the classification pipeline was tuned on the test set. It was not possible to use the training set for tuning, as the note-level annotations were only present in the test set. Using only unsupervised objectives for tuning is challenging, as topics might be distributed in numerous ways. Therefore, to obtain appropriate classification performance, guiding hyperparameter selection in a supervised manner was necessary. However, as the model was tuned on the same data used for evaluation, the model might have overfitted. In future work, trained models should be evaluated on independent test data.

To increase data accessibility, the AE dataset can be translated into other languages. Additional prospective work includes cross-lingual annotations, such that word- or phrase-level annotations and insights can be used in other languages.

VII. Conclusion

The Norwegian AE dataset is a resource for quality control improvement in hospitals. In addition to AEs, the dataset contains honest and open reporting about clinically relevant events and improvement suggestions which offers insight for quality assurance and patient safety in healthcare. This differs vastly from other available datasets focusing on adverse drug events and faulty devices malfunctioning. We want to collaborate with other research groups in order to use this dataset to improve patient safety and care quality.

Acknowledgment

We would like to thank the annotators. This research was funded by the Computational Sepsis Mining and Modelling project through the Norwegian University of Science and Technology (NTNU) Health Strategic Area and approved by the Norwegian Regional Committees for Medical and Health Research Ethics (REK), approval no. 26814. Technical support and the platform for performing analyses were provided by NTNU's HUNT Cloud.

References

[1] C. Deutschman and K. Tracey, "Sepsis: Current Dogma and New Perspectives," *Immunity*, vol. 40, pp. 463–475, Apr. 2014.

[2] C. Rhee, R. Dantes, L. Epstein, D. J. Murphy, C. W. Seymour, T. J. Iwashyna, S. S. Kadri, D. C. Angus, R. L. Danner, A. E. Fiore, and et al., "Incidence and Trends of Sepsis in US Hospitals Using Clinical vs Claims Data, 2009-2014," *JAMA: the J. of the Amer. Med. Assoc.*, vol. 318, no. 13, p. 1241–1249, Oct. 2017.

[3] M. Singer, C. S. Deutschman, C. W. Seymour, M. Shankar-Hari, D. Annane, M. Bauer, R. Bellomo, G. R. Bernard, J.-D. Chiche, C. M. Coopersmith, and et al., "The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)," *JAMA: the J. of the Amer. Med. Assoc.*, vol. 315, no. 8, p. 801–810, Feb. 2016.

[4] L. E. Huerta and T. W. Rice, "Pathologic Difference between Sepsis and Bloodstream Infections," *The J. of Appl. Lab. Med.*, vol. 3, no. 4, pp. 654–663, Oct. 2019.

[5] J. Paulsen, A. Mehl, Å. Askim, E. Solligård, B. O. Åsvold, and J. K. Damås, "Epidemiology and outcome of Staphylococcus aureus bloodstream infection and sepsis in a Norwegian county 1996-2011: an observational study," *BMC infectious diseases*, vol. 15, p. 116, Mar. 2015.

[6] L. A. Mermel, "Short-term Peripheral Venous Catheter-Related Bloodstream Infections: A Systematic Review," *Clin. Infectious Diseases*, vol. 65, no. 10, p. 1757–1762, Oct. 2017.

[7] E. Alexandrou, G. Ray-Barruel, P. J. Carr, S. A. Frost, S. Inwood, N. S. Higgins, F. Lin, L. Alberto, L. A. Mermel, and C. M. Rickard, "Use of short peripheral intravenous catheters: Characteristics, management, and outcomes worldwide," *J. of Hospital Med.*, vol. 13 5, 2018.

[8] E. Alexandrou, G. Ray-Barruel, P. J. Carr, S. Frost, S. Inwood, N. Higgins, F. Lin, L. Alberto, L. Mermel, and C. M. Rickard, "International prevalence of the use of peripheral intravenous catheters," *J. of Hospital Med.*, vol. 10, no. 8, p. 530–533, Aug. 2015.

[9] W. Zingg and D. Pittet, "Peripheral venous catheters: an under-evaluated problem," *Int. J. of Antimicrobial Agents*, vol. 34 Suppl. 4, p. S38–42, 2009.

[10] L. Zhang, S. Cao, N. Marsh, G. Ray-Barruel, J. Flynn, E. Larsen, and C. M. Rickard, "Infection risks associated with peripheral vascular catheters," *J. of Infection Prevention*, vol. 17, no. 5, pp. 207–213, Sep. 2016.

[11] J. M. Rohde, A. J. Odden, C. Bonham, L. Kuhn, P. N. Malani, L. M. Chen, S. A. Flanders, and T. J. Iwashyna, "The epidemiology of acute organ system dysfunction from severe sepsis outside of the intensive care unit," *J. of Hospital Med.*, vol. 8, no. 5, p. 243–247, May 2013.

[12] E. N. de Vries, M. A. Ramrattan, S. M. Smorenburg, D. J. Gouma, and M. A. Boermeester, "The incidence and nature of in-hospital adverse events: a systematic review," *Qual. & Saf. in Health Care*, vol. 17, no. 3, p. 216–223, Jun. 2008.

[13] P. D. Hibbert, C. J. Molloy, T. D. Hooper, L. K. Wiles, W. B. Runciman, P. Lachman, S. E. Muething, and J. Braithwaite, "The application of the global trigger tool: a systematic review," *Int. J. for Qual. in Health Care*, vol. 28, no. 6, p. 640–649, Dec. 2016.

[14] H. J. Murff, V. L. Patel, G. Hripcsak, and D. W. Bates, "Detecting adverse events for patient safety research: a review of current methodologies," *J. of Biomed. Inform.*, vol. 36, no. 1-2, p. 131–143, Feb. 2003.

[15] F. A. Griffin and R. K. Resar, *IHI Global Trigger Tool for Measuring Adverse Events*, 2nd ed. Cambridge, MA: Inst. for Healthcare Improvement, 2009.

[16] S. N. Musy, D. Ausserhofer, R. Schwendimann, H. U. Rothen, M.-M. Jeitziner, A. W. Rutjes, and M. Simon, "Trigger tool-based automated adverse event detection in electronic health records: Systematic review," *J. of Med. Internet Res.*, vol. 20, no. 5, p. e198, May 2018.

[17] E. Apostolova and T. Velez, "Toward Automated Early Sepsis Alerting: Identifying Infection Patients from Nursing Notes," in *BioNLP 2017*, Aug. 2017, pp. 257–262.

[18] I. Larizgoitia, M.-C. Bouesseau, and E. Kelley, "Who efforts to promote reporting of adverse events and global learning," *J. of Public Health Res.*, vol. 2, no. 3, p. e29, Dec. 2013.

[19] "Japan council for quality health care (JQ) project to collect medical near-miss/adverse event information." [Online]. Available: https://www.med-safe.jp/contents/english/index.html

[20] R. Tikkanen, R. Osborn, E. Mossialos, A. Djordjevic, and G. Wharton, "2020 international health care system profiles," Dec. 2020. [Online]. Available: https://www.commonwealthfund.org/international-health-policy-center/system-profiles

[21] P. Bjark, E. Hansen, and E. Lingaas, "Helsetjenesteassosiert infeksjon som årsak til død i sykehus," *Tidsskrift for den Norske legeforening*, vol. 140, Apr. 2020.

[22] P. Stenetorp, S. Pyysalo, G. Topić, T. Ohta, S. Ananiadou, and J. Tsujii, "BRAT: a Web-based Tool for NLP-assisted Text Annotation," in *Proc. Demonstrations at 13th Conf. EACL 2012*, Apr. 2012, pp. 102–107.

[23] N. Halko, P.-G. Martinsson, and J. Tropp, "Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions," *SIAM Review*, vol. 53, pp. 217–288, Jan. 2011.

[24] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine Learning in Python," *J. of Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[25] T. Head, M. Kumar, H. Nahrstaedt, G. Louppe, and I. Scherbatyi, "scikit-optimize," Sep. 2020, v0.8.1. [Online]. Available: http://doi.org/10.5281/zenodo.4014775

[26] T. B. Røst, C. R. Tvedt, H. Husby, I. A. Berg, and Ø. Nytrø, "Identifying catheter-related events through sentence classification," *Int. J. of Data Mining and Bioinf.*, vol. 23, no. 3, p. 213–233, Jan. 2020.

Yan, M.Y., Høvik, L.H., Pedersen, A., Gustad, L.T., Nytrø, Ø, Preliminary Processing and Analysis of an Adverse Event Dataset for Detecting Sepsis-Related Events
IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2021, Houston, TX, USA, December 9-12, 2021, pages 1605–1610. IEEE, December 2021

Paper C

Method for Designing Semantic Annotation of Sepsis Signs in Clinical Text

Melissa Y. Yan, Lise Tuset Gustad, Lise Husby Høvik, and Øystein Nytrø

Published in:
Proceedings of the 5th Clinical Natural Language Processing Workshop, (ClinicalNLP@ACL), July 2023, pp. 236–246. DOI: 10.18653/v1/2023.clinicalnlp-1.29.

Method for Designing Semantic Annotation of Sepsis Signs in Clinical Text

Melissa Y. Yan[1], **Lise Tuset Gustad** [2,3], **Lise Husby Høvik**[3,4], **Øystein Nytrø**[1]
[1] Dept. of Computer Science, Norwegian University of Science and Technology
[2] Faculty of Nursing and Health Sciences, Nord University
[3] Dept. of Circulation and Medical Imaging, Norwegian University of Science and Technology
[4] Clinic of Anaesthesia and Intensive Care, St. Olavs Hospital, Trondheim University Hospital
{melissa.yan, lise.t.gustad, lise.hovik, nytroe}@ntnu.no

Abstract

Annotated clinical text corpora are essential for machine learning studies that model and predict care processes and disease progression. However, few studies describe the necessary experimental design of the annotation guideline and annotation phases. This makes replication, reuse, and adoption challenging.

Using clinical questions about sepsis, we designed a semantic annotation guideline to capture sepsis signs from clinical text. The clinical questions aid guideline design, application, and evaluation. Our method incrementally evaluates each change in the guideline by testing the resulting annotated corpus using clinical questions. Additionally, our method uses inter-annotator agreement to judge the annotator compliance and quality of the guideline. We show that the method, combined with controlled design increments, is simple and allows the development and measurable improvement of a purpose-built semantic annotation guideline. We believe that our approach is useful for incremental design of semantic annotation guidelines in general.

1 Introduction

Annotated clinical text corpora provide natural language processing (NLP) and machine learning (ML) studies the data necessary to find patterns, classify, and predict patient risk and disease progression. Compared to models that only utilize structured data from the electronic health record (EHR), many studies and reviews have shown that model performance can increase by incorporating unstructured clinical text (Soguero-Ruíz et al., 2016; Huddar et al., 2016; Culliton et al., 2017; Assale et al., 2019; Sheikhalishahi et al., 2019; Spasic and Nenadic, 2020).

Pre-existing annotated clinical corpora include the Medical Information Mart for Intensive Care (MIMIC-III) (Johnson et al., 2016), the Clinical E-Science Framework (CLEF) (Roberts et al., 2007), and the Informatics for Integrating Biology and the Bedside (i2b2) challenges and National NLP Clinical Challenges (n2c2) (Uzuner and Stubbs, 2015; Luo et al., 2020). However, studies utilizing pre-existing annotated corpora must limit their research questions to the specific purpose(s) for which the corpus was annotated. Otherwise, the annotations required to answer a research question might be missing or too general. Thus, many studies opt to develop their own annotated clinical corpus tailored to capture and extract the necessary information for their research (Yim et al., 2015; Rama et al., 2018; South et al., 2009; Oliveira et al., 2022).

Methods with lower requirements for supervision, such as information extraction, commonly use keyword search, rule-based algorithms, and ML to detect clinical cases. However, those methods might not consider the context of the clinical case (Ford et al., 2016). For example, different documented signs within a specific situation can describe a medical condition that is not named. Hence, medical expertise is necessary for making annotation judgments and capturing clinical knowledge within the text (Xia and Yetisgen-Yildiz, 2012). Retrieving domain-specific patient knowledge to ascertain or answer clinical questions includes extracting data, information, and knowledge. Data are attributes (e.g., names or dates), information gives meaning to data (e.g., location, cause, and time), and knowledge interprets information based on one's role and responsibility (e.g., clinical document's purpose and effect) (Gudea, 2005).

Making a quality annotated corpus is an iterative process that includes designing an annotation guideline, annotating text with the guideline, and refining the guideline based on inter-annotator agreement (IAA) (Roberts et al., 2009; Xia and Yetisgen-Yildiz, 2012; Deleger et al., 2012; Savkov et al., 2016; Oliveira et al., 2022). Although studies describe how annotated clinical corpora were made, few studies are explicit about the design process.

Proceedings of the 5th Clinical Natural Language Processing Workshop, pages 236–246

We believe that the acquisition and transformation of clinical questions about the patient cohort into corresponding corpus requirements for retrieving information from the actual text of the annotated corpus should drive the annotation process.

2 Related Work

This section provides an overview of studies that describe the design process leading to an annotation guideline and annotated clinical corpus. Studies that share their annotation challenges or offer improvements are also included.

The CLEF Corpus was semantically annotated to help develop and evaluate the CLEF information extraction system (Roberts et al., 2007, 2009). Free-text documents in the corpus are histopathology reports, imaging reports, and clinical narratives (i.e., discharge summaries, reports, case notes, audits, letters, or narratives to the general practitioner, consultant, referrer, or patient). Initially, templates for the documents using ontology-based entities and relationships were manually filled-in. However, the templates did not directly align with text, and ontology complexity made it time-consuming to fill templates. Thus, Roberts et al. (2009) iteratively developed an annotation guideline based on a simplified version of the original ontology and template definitions. Following established standard NLP annotation methodology (Boisen et al., 2000), 2 clinicians annotated 31 documents over 5 sessions, and a third annotator resolved disagreements (Roberts et al., 2009). However, due to workload and time constraints, resigning annotators could have impacted the corpus quality and size. Thus, Roberts et al. (2009) proposed solutions such as pre-annotated documents and a reduced annotation scope.

The i2b2 challenges have annotated corpora for various purposes. For example, in the i2b2 NLP challenge of extracting patient medication from discharge summaries, 79 annotators from 20 teams annotated 251 discharge summaries in a community annotation experiment (Uzuner et al., 2010). The annotation guideline was developed iteratively in 2 phases before the community annotation. For several iterations in phase 1, university students annotated discharge summaries that were measured for IAA and asked questions to aid revisions. This produced a guideline and 17 annotated discharge summaries for phase 2. Finally, during phase 2, teams annotated discharge summaries using the guideline and addressed inconsistencies within the 17 annotated discharge summaries to produce a refined guideline.

The i2b2 temporal relations corpus contains 310 discharge summaries annotated by 8 annotators (Sun et al., 2013). The annotation guideline was based on the TimeML event and temporal expression specification language (Pustejovsky et al., 2003) and the Temporal Histories of Your Medical Event (THYME) project annotation guidelines. The corpus development process included: a guideline development pilot study, data selection, pre-annotation, annotator training session, 2 annotators annotating pre-annotated documents, an adjudicator who resolved disagreements, and evaluation.

The 2014 i2b2/UTHealth de-identification corpus annotation guideline focuses on removing Protected Health Information (PHI) in longitudinal medical records for automatic de-identification system development (Stubbs and Uzuner, 2015). Introduced PHI subcategories enable downstream analyses to adjust the scope or focus on specific categories. Additionally, they compared parallel and serial annotation processes on pre-annotated and unannotated corpora and found that the process does not affect annotation quality (Stubbs and Uzuner, 2017).

Xia and Yetisgen-Yildiz (2012) utilized a variation of the typical annotation process for 3 different studies. Each study's corpus focused on a specific clinical report, such as radiology, chest x-ray, or intensive care unit reports. The process included: defining a study based on clinical needs, selecting data, gaining ethical approval, writing annotation guidelines, creating annotation tools, annotating, building a system with the corpus, and testing if the system meets clinical needs. Physicians were guideline designers and annotators, whereas NLP researchers provided technical support and built NLP systems with the corpora. Suggestions for improvement included more NLP researcher involvement, consideration for guideline granularity versus annotation time, marking rationale or evidence for a label, and estimating time commitment.

Deleger et al. (2012) developed their annotation guideline by building off a previous guideline. The rest of the methods were similar: defining annotation tasks, selecting data from stratified random sampling, and annotating with 2 annotators. During the annotation process, 2 annotators annotated the same documents, IAA was measured, and con-

sensus sessions were held to resolve disagreements and update the guideline. Using the same annotation process, they built gold standard corpora from clinical trial announcements, US Food and Drug Administration (FDA) drug labels, and EHR clinical notes. This included clinical notes such as discharge summaries, referrals, reports, and notes for consultations, procedures, plans, or progress.

Interested in capturing infections caused by central venous catheters, a nurse specializing in infection annotated 2 745 of 22 174 inspected notes (Røst et al., 2018). Before inspection, duplicate notes were removed. The guideline was a table containing events for annotation. Defined by computer scientists, nurses, and an NLP domain expert, the annotation labels formed a hierarchy starting with generalized events at the top level and more specific events below. They also provided information about data access restrictions to promote patient confidentiality and clinical record extraction. Record extraction included physician and nurse notes for admissions, care, plans, evaluations, transfers, and discharge summaries.

In this study, we focus on a method of incremental annotation guideline design by intertwining acquisition with testing of corpus requirements and corresponding annotation phases. This ensures that the guideline produces an annotated corpus that fulfills corpus requirements derived from clinical questions, even if the clinical questions are not answerable by the actual data. To the best of our knowledge, there lacks a study that describes this approach in detail.

3 Objective

This study aims to describe our method for designing a semantically annotated corpus for signs of sepsis by starting from clinical questions that formulate the corpus requirements. Hence, the main contributions are: (1) providing a detailed description of the guideline design process before annotation, (2) illustrating the systematic and iterative annotation process taken, and (3) discussing insights from the design and annotation process.

3.1 Clinical Problem

Sepsis leads to life-threatening multi-organ failure and is caused by a dysregulated host immune response to an infection (Singer et al., 2016). One infectious agent is the *Staphylococcus aureus* (*S. aureus*) bacteria found on skin that is known to cause serious bloodstream infections (BSIs). There is a known overlap between sepsis and BSI, as BSI is found in 30–58% of sepsis patients depending on which sepsis definition is used (Phua et al., 2013; Mellhammar et al., 2021). An estimated 7.6%–35% of *S. aureus* BSIs are related to peripheral intravenous catheters (PIVCs), and the presence of phlebitis can indicate infection via PIVC (Mermel, 2017). A PIVC is a medical device inserted into a vein for administering intravenous (IV) fluids, medication, and blood transfusions. Unfortunately, improperly managed PIVCs can become gateways that lead to phlebitis, BSI, or sepsis (Zhang et al., 2016).

Despite the high sepsis mortality rates and routine usage of PIVCs, both sepsis and PIVCs are poorly documented in clinical text and rarely available as structured data in the EHR (Rohde et al., 2013; Alexandrou et al., 2018). This makes it challenging for hospitals to perform retrospective systematic quality surveillance of PIVC-related BSIs to lower sepsis incidents. Additionally, the lack of explicit documentation inhibits the opportunities for clinicians to learn from and improve PIVC care practices to lower BSI and sepsis rates.

4 Original Adverse Event Dataset

We had access to 18 555 Norwegian adverse event (AE) reports extracted from a hospital's electronic incident reporting system (Yan et al., 2021). Extracted AE reports described procedural deviations, misunderstandings, resource needs, and risky patient behavior. Each report has structured data (i.e., identifier, registration date, reporting hospital unit, if the event is patient-related or security-related, event type, and event severity) and an unstructured free-text note.

5 Semantic Annotation Design Process

This section presents the semantic annotation design process leading up to the annotation process and guideline development. A summary can be found in Figure 1.

5.1 Clarify and Operationalize Clinical Questions to Form Corpus Requirements

Curious about PIVC-related BSI or phlebitis that can lead to sepsis and opportunities to improve patient care, nurses proposed the clinical question: "Is there a connection between PIVCs and BSIs or PIVCs and phlebitis at the hospital?" Thus, the

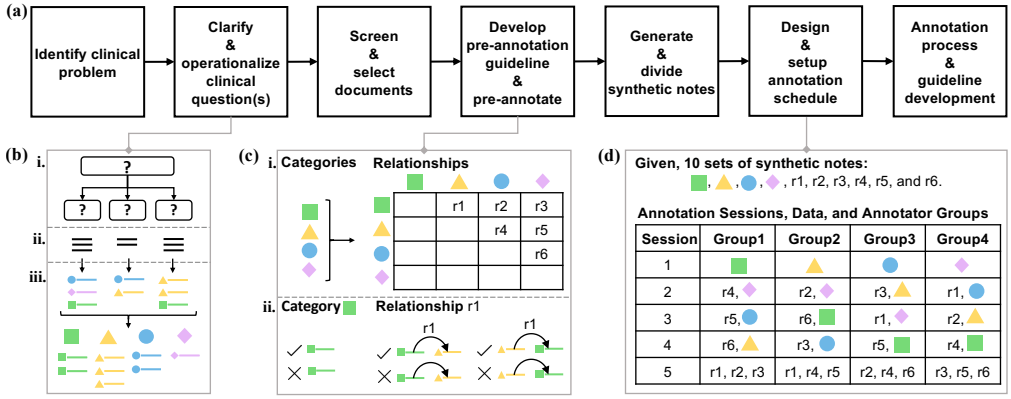


Figure 1: Semantic annotation design process. **(a)** Overview of the process until annotation and guideline development. **(b)** Clarify and operationalize clinical questions into corpus requirements to form annotation categories or entities. **i.** Clarify and operationalize clinical questions by expanding them to derive corpus requirements. **ii.** List examples to answer each question. **iii.** Sort examples into different categories to form the annotation categories. **(c)** Develop the pre-annotation guideline and pre-annotate. **i.** Find relationships using unique category combinations. **ii.** Create the pre-annotation guideline using concrete examples and counterexamples for categories and relationships. **(d)** Determine the annotation sessions and annotator groups to create a schedule. Divide synthetic notes into sets based on the number of categories and relationships. Each group annotates each category at least once in a different session. Additionally, each relationship is annotated at least twice by a different group throughout the sessions. Thus, the sets can be reused in different sessions by different groups, and guideline revisions can be tested on a different group using the same data.

clinical need is to identify PIVC-related BSI and phlebitis or sepsis signs, preferably by automatically classifying patients with PIVCs requiring follow-up care. Through iterative discussions with nurses and computer scientists, the clinical question was clarified to ensure data, information, and knowledge could be extracted to answer the clinical question (Figure 1 (**b**)i). Thus, the clinical question was clarified by expanding it into:

1. How can sepsis or BSIs be identified when the symptoms are similar to other diseases?
2. What signs or symptoms does PIVC-related phlebitis have?
3. How can poorly documented PIVCs be identified?

Those clinical questions were further modified based on the nurses' perspectives. For example, certain types of catheters are distinctly documented (for data extraction). Other catheters can be distinguished based on anatomical insertion sites (for information extraction) or procedures (for knowledge extraction). This resulted in the following questions that also operationalize and form the corpus requirements:

1. What are the different documented signs of infections or phlebitis, specifically those related to PIVCs, BSIs, or sepsis?
2. What can distinguish catheter types in the notes?
3. Where are the documented anatomical insertion sites of catheters?
4. What procedures, interventions, and activities can be related to catheter use from text content or report structured data (e.g., ward type or care situation)?

Figure 2 shows how clinical questions guide the design, application, and evaluation of the annotated corpus, annotation guideline, and corpus requirements.

Creating an annotated clinical corpus is time-consuming and labor-intensive (Wei et al., 2018). However, discussions revealed that we could not reuse a corpus and needed a new annotation guideline. Corpus requirements provided the annotation purpose and can be viewed as "information requests" to develop procedures for extracting data, information, and knowledge through annotation. Extracted data can be facts and observations, such

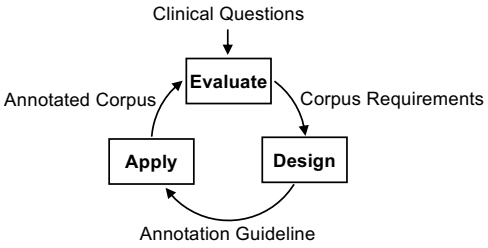


Figure 2: Design driven by clinical questions. Evaluating clinical questions forms the annotated corpus requirements used to design the annotation guideline. Annotators apply the annotation guideline to make a new (sub)corpus. The corpus is evaluated using inter-annotator agreement for annotator compliance and guideline comprehension. Clinical questions are used separately afterward to evaluate the corpus and requirements.

as dates, signs, or symptoms (e.g., purple skin). Information extracted can be phrases for specific signs and symptoms of a case (e.g., purple skin is a sign of a bruise). Furthermore, knowledge extracted can be other signs or symptoms that indicate something not necessarily mentioned (e.g., bruise color can indicate the stage).

Clarifying and operationalizing clinical questions helped determine corpus requirements about documented patient features, patient states, and care features. Including clinicians and computer scientists when clarifying questions was essential because it helped identify requirements for representing knowledge populated by text processing. Furthermore, these questions can be used to evaluate if the annotated corpus can answer the clinical questions.

5.2 Form Annotation Categories or Entities

Clinicians provided examples for the corpus requirements by listing keywords, phrases, and sentences (Figure 1 **(b)**ii). Computer scientists asked clarifying questions to resolve confusion and ambiguity. They also inquired about clinical actions versus actual documented actions to understand what is documented in the text. After generating a list of answers, answers were sorted into different categories (technically known as entities) through discussions (Figure 1 **(b)**iii). Each category is a label for a single word or phrase.

Answers were sorted into the 4 categories: **Sign**, **Location**, **Device**, and **Procedure**. Two additional categories, **Sensitivity** and **Person**, were included to ensure that data is de-identified and that the 4 categories can be linked to an individual. Thus, the 7 main categories are as follows:

1. **Sign**: infection signs
2. **Location**: anatomical insertion sites
3. **Device**: signs of catheter types
4. **Procedure**: catheter acts or interventions
5. **Sensitivity**: potential patient identifiers
6. **Person**: role (e.g., patient or clinician)
7. **Whole**: AE note topic label for validation (i.e., has patient identifier or is about infection, BSI, sepsis, faulty device, catheter, and/or PIVC).

Excluding the **Whole** category, the remaining 6 categories each form a hierarchy with more specific subcategories underneath. Subcategories are used to capture more detailed granularity from the text (e.g., the **Device** category contains a "Catheter" subcategory with different specific catheter types as subcategories).

Concrete categories made understanding the clinical annotation task easier and less ambiguous for the multidisciplinary research group. Having discussions and generating a list with clinicians helped determine the categories and subcategories needed to extract necessary data, information, and knowledge.

5.3 Screen and Select Notes

To ensure that categories specified above are present in notes, 700 randomly selected AE notes were manually screened and categorized by a computer scientist and nurse. Categorizing notes included providing a comment about the categorization rationale and marking potentially ambiguous notes. In addition, the potentially ambiguous notes were clarified in discussions and used as examples for properly annotating notes. Screening notes identified documented information that could satisfy corpus requirements and help answer clinical questions in downstream analyses. Additionally, it provided examples that drove preliminary guideline development in the next section.

5.4 Develop Pre-Annotation Guideline and Pre-Annotate

Initially, 6 possible relationships were found using a table with unique category combinations (Fig-

ure 1 **(c)**i). Then, those 6 relationships were discussed within the research group to evaluate which were required and merged. This resulted in the following 4 relationships for linking categories:

1. **Person** $\xrightarrow{\text{Person has}}$ **Sign**, **Location**, **Device**, or **Procedure**
2. **Procedure** $\xrightarrow{\text{Procedure uses}}$ **Device**
3. **Sign** $\xrightarrow{\text{Caused by}}$ **Device** or **Procedure**
4. **Sign**, **Device**, or **Procedure** $\xrightarrow{\text{Located nearby/on/in}}$ **Location**

Before actual annotation, the preliminary annotation guideline underwent a pre-annotation phase. Two pre-annotation guidelines were created to assess the utility and decide how detailed an annotation guideline should be for consistent annotation. The low granularity guideline was a Word document that provided brief instructions, a hierarchical list of categories, and only annotation examples for 2 categories (i.e., **Sensitivity** and **Person**). In contrast, the high granularity guideline was a static HTML webpage with interactive instructions for using the annotation tool and had links to corresponding sections for each category or relationship. Each category and the relationships in the high granularity guideline contained 1 concrete annotation example and counterexamples as needed (Figure 1 **(c)**ii). A nurse and a computer scientist used both pre-annotation guidelines to annotate 15-27 notes. Afterward, the research group determined a high granularity annotation guideline was more informative and easier to use with the annotation tool.

Capturing relationships between categories ensures that data is not lost in downstream analysis (e.g., infection signs at a specific location). It can also provide additional support to answer the clinical questions. By merging relationships, the complexity of annotation options was simplified and reduced. It is ideal to reduce the complexity of annotation because making the annotation task too difficult and time-consuming can result in annotators resigning (Roberts et al., 2009). The pre-annotation phase allowed the research group to manually evaluate, discuss, revise, and improve the guideline before use. This included the suitable granularity level and ease of use for the annotators.

5.5 Generate and Divide Synthetic Notes

Synthetic notes appear real and could be real. 100 unique synthetic clinical text notes were manually generated through 2 methods. The first method combines parts of the original notes to create a similar synthetic AE note with manually anonymized patient identifiers, and the content was verified by a nurse. Whereas in the second method, a nurse manually created a note based on possible clinical scenarios with synthetic patients to ensure some notes contained information about catheters and/or infections. The mean, minimum, maximum and median tokens per AE note in the corpus were 45, 4, 316, and 36, respectively. Generating synthetic notes took a couple of workdays for the nurse.

Afterward, the 100 unique synthetic notes were divided into 10 distinct sets with 10 notes each. Each set corresponds to either a category or relationship. The categories utilized in dividing the sets are those related to catheters or infections (i.e., **Sign**, **Location**, **Device**, and **Procedure**). The relationships utilized are the 6 initial possible relationship combinations.

AE notes often contain excessive and potentially identifying information irrelevant for annotating catheter-related events. Thus, relevant and closely related AE notes were selected and combined to use annotator time efficiently. Generating synthetic notes ensures the data is anonymized and usage is optimized, as clinical data is scarce. Additionally, it provides more data for ML analyses and makes the data more easily accessible to other researchers. Separating synthetic data into different categories or relationships ensures that specific labels will be annotated within the dataset. Different sets could be given to different annotators to reuse data and test if annotation guideline revisions improved IAA.

5.6 Design and Set Up Annotation Schedule

The same 4 categories and 6 initial relationships used to divide synthetic notes into 10 sets were used to design the annotation schedule (Figure 1 **(d)**). Categories were separated into groups, and relationships were added such that each group would annotate a relationship that excluded the group's category. Additionally, relationships within the groups were organized such that each relationship was annotated at least twice by 2 different annotator groups to evaluate revisions. This resulted in 4 annotator groups, each with 5 annotation sessions that

used a different set of notes and could annotate in parallel. Each group had 2 annotators so that IAA could be measured. This design defined the annotation schedule, the number of annotation groups needed, and how to reuse synthetic notes for guideline development. Furthermore, parallelization for each session helped reduce the project timeline.

6 Annotation Process and Annotation Guideline Development

Following the schedule, synthetic notes were annotated by 4 annotator groups over 5 sessions using a systematic, iterative annotation process for guideline refinement. In each session, 2 annotators from each group annotated notes based on an annotation guideline using the Brat rapid annotation tool (BRAT) (Stenetorp et al., 2012). Afterward, annotations were evaluated for IAA and manually inspected to assess if annotations could fulfill corpus requirements and answer the clinical questions. Text was tokenized and annotation labels were assigned to tokens before measuring the IAA F_1-score. Disagreements and ambiguities were discussed within the research group, and comments from annotators were incorporated. Next, a computer scientist revised the guideline based on discussions. Finally, the process was repeated with a new set of notes and the revised guideline. Figure 3 shows an example sentence annotated by 2 different annotators.

7 General Results from Sessions 1–5

Over 5 sessions, 8 annotators annotated 100 unique synthetic AE notes to produce 770 annotated synthetic AE notes. From session 1, it was clear that subcategory and attribute names should not be used in more than one category, and synonyms should be avoided. For example, simultaneously having "Name" as both a **Sensitivity** subcategory and an attribute for the **Person** category raised questions. Furthermore, annotators left relationships, attributes, and notes unannotated because they felt those notes were irrelevant to answering the clinical questions.

The need for annotating relationships, attributes, and all notes for ML was addressed in session 2. Red font emphasized guideline revisions, and the guideline began with an "Overview of Updated Instructions" section to aid annotators in identifying revisions. In sessions 2 and 3, the main revisions were correcting and including missing subcategories to address annotator concerns.

Session 4 provided a structured terminology for the guideline. A terminology was developed from the guideline to give structure and provide users quick insight into the annotated corpus for downstream analysis (Yan et al., 2023). This restructured the annotation guideline for session 5 by removing ambiguities and allowed AE note querying to answer the clinical questions. For example, the new **Observation** category encompasses the **Sign** category's signs and symptoms and the **Procedure** subcategory "Device malfunction signs." The computer scientist who revised the guidelines misinterpreted clinical knowledge and made incorrect assumptions in the previous sessions, so the terminology and restructured guideline were validated by nurses to ensure medical concepts were used correctly before session 5. The session 1–4 annotation guidelines were made available online[1] for Yan et al. (2021), and the session 5 annotation guideline was added online for this study. IAA for different sessions are in Figure 4.

8 Discussion

8.1 Design and Annotation Process

The annotation guideline development design process focuses on identifying the effect of the guideline on different categories, corpus content, and clinical questions. Categories were developed to answer different clinical questions and focus on localized guideline changes. Revising parts of specific category hierarchies made it possible to make controlled changes to specific subcategories in the annotation guideline and observe the impact on the annotated corpus, IAA, and clinical questions.

The annotation process greatly influences and drives guideline development. Clinical questions led to corpus requirements that developed the annotation guideline, which is applied on the annotated corpus and evaluated by the clinical questions. In turn, evaluating the annotated corpus also either indicates if it is possible to fulfill corpus requirements to answer clinical questions or detects a lack of corpus content needed for the clinical questions. Using the iterative process, we uncovered corpus requirements that the corpus content could not fulfill and could revise the requirements to drive guideline development and annotation.

[1] https://folk.ntnu.no/melissay/ae-guidelines/

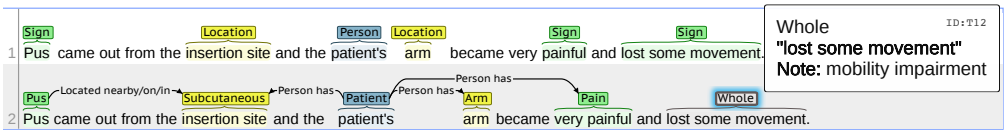


Figure 3: Annotation example for 2 different annotators. Annotator1 on top annotated using only the main categories, whereas Annotator2 on the bottom used subcategories to capture more detail and relationships to link categories. Although the **Whole** category is for indicating if an AE note contains information related to the clinical questions, Annotator2 has misused this label to leave a comment and indicate the phrase is about "mobility impairment". Actual AE notes only contain annotations from 1 annotator, and annotators cannot see the annotations from others.

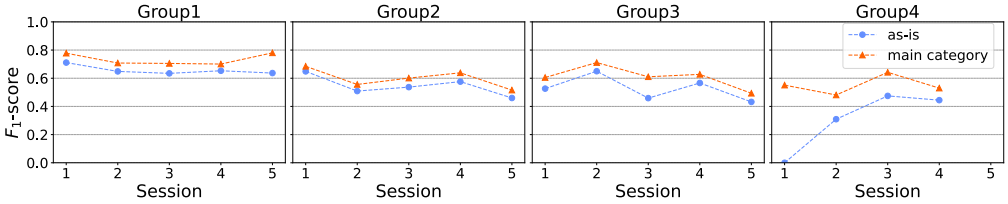


Figure 4: F_1-score for 4 annotator groups over the 5 sessions. The "as-is" F_1-score was calculated using annotator provided labels. Whereas, the "main category" F_1-score converted the labels to the main categories of **Sign**, **Location**, **Device**, **Procedure**, **Sensitivity**, **Person**, or **Whole**. Group4 session 5 has no F_1-score because an annotator withdrew.

8.2 Inter-annotator Agreement

There are several possible reasons for changes in Figure 4's F_1-score. An annotator often misused the **Whole** category to leave comments about clinical knowledge, while this is clinically insightful, it decreases the IAA (e.g., Figure 3 Annotator2 misusing the **Whole** category). The guideline complexity increased and reduced annotator compliance (i.e., sessions 1-5 had 89, 88, 105, 110, and 137 subcategories, respectively). As shown by the "as-is" F_1-score decrease in session 5, the guideline likely became too complex after session 4 revisions. The annotator from Group4 probably withdrew because of the increasing clinical complexity. Another annotator gave feedback that they were uncertain if they annotated some notes correctly. So, increasing the guideline and notes can overwhelm annotators (i.e., sessions 1-5 had 10, 20, 20, 20, and 30 notes, respectively). Group1 was a medical and nursing student, Group2 was a nurse and medical student, Group3 were nurses, and Group4 was a nurse and computer scientist. In general, students followed guidelines well, even if it contained incorrect medical concepts. Thus, paired annotators could have different clinical expertise that impacted results.

Granularity can have an effect on IAA, but granularity can be adjusted to identify problematic subcategories and utilized by those performing downstream analyses. Lower granularity in the annotation guideline leads to higher agreement because it reduces the complexity and level of detail. Annotators usually agree on which main category to annotate a word or phrase, but they had difficulties choosing certain subcategories. For example, in Figure 3 Annotator1 annotated with the main categories whereas Annotator2 was more detailed and annotated almost the same words with subcategories from the same main categories (e.g., "Pus" was annotated by Annotator1 with the **Sign** category and by Annotator2 with **Sign**'s subcategory "Pus"). This is also shown in Figure 4 for Group4 in session 1, where the "as-is" F_1-score is 0, but the "main category" F_1-score is 0.55. It is also possible to perform IAA on different subcategories within a subcategory to identify the most problematic areas after guideline revisions. The granularity in the annotated corpus can also be utilized and adjusted in downstream analyses based on the level of detail required by researchers.

9 Conclusion

Our method captures knowledge about sepsis signs in clinical text. We control changes in the annotation guideline by using hierarchical categories and continuous evaluation. Through applying a systematic, iterative annotation process, we evaluated the

changes using the clinical questions and IAA. The clinical questions evaluate corpus quality, and IAA evaluates annotator compliance and guideline complexity. As the guideline is designed to answer different clinical questions, it is possible to adjust the granularity level as needed to answer different clinical questions. By detailing our design process and annotation process, we hope our method can aid other researchers who cannot utilize pre-annotated corpora in developing an annotated corpus for their research.

Limitations

This method for designing and annotating clinical text for a specific clinical use case can be beneficial for researchers needing to annotate a corpus. However, there are some limitations. First, the experiences are based on a specific clinical case and focus on the qualitative aspects. Details of certain parts of the design and annotation process will likely need to be adjusted based on resources available to other researchers. This can include the data selected for annotation, the number of annotators available, and the annotators' level of expertise. For instance, the use case in the design process is based on using 8 annotators to annotate 100 synthetic AE notes over 5 sessions. Second, expertise and additional time are required to generate synthetic notes for annotation. Finally, future work is still needed to replicate the described design and annotation process on other forms of clinical text and problems.

Ethical Considerations

To protect patient privacy when designing and annotating clinical text, synthetic AE notes were manually generated and verified by a nurse to ensure the data is anonymized. Additionally, the annotation guideline includes the **Sensitivity** category to allow annotators to label potential information in the synthetic notes that could identify a patient. This process was described to provide an example for researchers who need to annotate sensitive data.

The Norwegian Regional Committees for Medical and Health Research Ethics (REK) has approved the use of medical data in this study (REK approval no. 26814; 2018/1201/REKmidt). To ensure annotators are protected, collecting and processing personal annotator data has also been approved by the Norwegian Centre for Research Data (NSD reference no. 142683). Furthermore, the annotators have consented to the use of their specified personal information (i.e., profession and years of experience) and their annotations.

Acknowledgements

The authors would like to thank the 8 annotators, this project would not be possible without them. Research is funded by the Computational Sepsis Mining and Modelling project through the Norwegian University of Science and Technology (NTNU) Health Strategic Area.

References

Evan Alexandrou, Gillian Ray-Barruel, Peter J. Carr, Steven A. Frost, Sheila Inwood, Niall Higgins, Frances Lin, Laura Alberto, Leonard Mermel, Claire M. Rickard, and O.M.G. Study Group. 2018. Use of short peripheral intravenous catheters: Characteristics, management, and outcomes worldwide. *Journal of hospital medicine*, 13.

Michela Assale, Linda Greta Dui, Andrea Cina, Andrea Seveso, and Federico Cabitza. 2019. The revival of the notes field: Leveraging the unstructured content in electronic health records. *Frontiers in Medicine*, 6:66.

Sean Boisen, Michael Crystal, Richard M. Schwartz, Rebecca Stone, and Ralph M. Weischedel. 2000. Annotating resources for information extraction. In *Proceedings of the Second International Conference on Language Resources and Evaluation, LREC 2000*, Athens, Greece. European Language Resources Association.

Phil Culliton, Michael Levinson, Alice Ehresman, Joshua Wherry, Jay S Steingrub, and Stephen I Gallant. 2017. Predicting severe sepsis using text from the electronic health record. In *Workshop on Machine Learning For Health at the Conference on Neural Information Processing Systems (NIPS ML4H 2017)*, Long Beach, CA, USA.

Louise Deleger, Qi Li, Todd Lingren, Megan Kaiser, Katalin Molnar, Laura Stoutenborough, Michal Kouril, Keith Marsolo, and Imre Solti. 2012. Building gold standard corpora for medical natural language processing tasks. *AMIA Annual Symposium proceedings. AMIA Symposium.*, 2012:144–153.

Elizabeth Ford, John A. Carroll, Helen E. Smith, Donia Scott, and Jackie Cassell. 2016. Extracting information from the text of electronic medical records to improve case detection: a systematic review. *Journal of the American Medical Informatics Association*, 23:1007 – 1015.

Sorin Gudea. 2005. Data, information, knowledge: a healthcare enterprise case study. *Perspectives in Health Information Management*, 2:8.

Vijay Huddar, Bapu Koundinya Desiraju, Vaibhav Rajan, Sakyajit Bhattacharya, Shourya Roy, and Chandan K. Reddy. 2016. Predicting complications in critical care using heterogeneous clinical data. *IEEE Access*, 4:7988–8001.

Alistair E. W. Johnson, Tom J. Pollard, Lu Shen, Li wei H. Lehman, Mengling Feng, Mohammad M. Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G. Mark. 2016. MIMIC-III, a freely accessible critical care database. *Scientific Data*, 3.

Yen-Fu Luo, Sam Henry, Yanshan Wang, Feichen Shen, Ozlem Uzuner, and Anna Rumshisky. 2020. The 2019 n2c2/UMass Lowell shared task on clinical concept normalization. *Journal of the American Medical Informatics Association*, 27(10):1529–e1.

Lisa Mellhammar, Fredrik Kahn, Caroline Whitlow, Thomas Kander, Bertil Christensson, and Adam Linder. 2021. Bacteremic sepsis leads to higher mortality when adjusting for confounders with propensity score matching. *Scientific reports*, 11(1):6972.

Leonard A. Mermel. 2017. Short-term peripheral venous catheter-related bloodstream infections: A systematic review. *Clinical Infectious Diseases*, 65(10):1757–1762.

Lucas Emanuel Silva E Oliveira, Ana Carolina Peters, Adalniza Moura Pucca da Silva, Caroline Pilatti Gebeluca, Yohan Bonescki Gumiel, Lilian Mie Mukai Cintho, Deborah Ribeiro Carvalho, Sadid Al Hasan, and Claudia Maria Cabral Moro. 2022. SemClinBr - a multi-institutional and multi-specialty semantically annotated corpus for Portuguese clinical NLP tasks. *Journal of Biomedical Semantics*, 13(1):13.

Jason Phua, Wang Ngerng, Kay See, Chee Tay, Timothy Kiong, Hui Lim, Mei Chew, Hwee Yip, Adeline Tan, Haji Khalizah, Rolando Capistrano, Kang Lee, and Amartya Mukhopadhyay. 2013. Characteristics and outcomes of culture-negative versus culture-positive severe sepsis. *Critical Care*, 17(5):R202.

James Pustejovsky, José M. Castaño, Robert Ingria, Roser Saurí, Robert J. Gaizauskas, Andrea Setzer, Graham Katz, and Dragomir R. Radev. 2003. TimeML: Robust specification of event and temporal expressions in text. In *New Directions in Question Answering, Papers from 2003 AAAI Spring Symposium*, pages 28–34, Stanford University, Stanford, CA, USA. AAAI Press.

Taraka Rama, Pål Brekke, Øystein Nytrø, and Lilja Øvrelid. 2018. Iterative development of family history annotation guidelines using a synthetic corpus of clinical text. In *Proceedings of the Ninth International Workshop on Health Text Mining and Information Analysis*, pages 111–121, Brussels, Belgium. Association for Computational Linguistics.

Angus Roberts, Robert J. Gaizauskas, Mark Hepple, Neil Davis, George Demetriou, Yikun Guo, Jay Kola, Ian Roberts, Andrea Setzer, Archana Tapuria, and Bill Wheeldin. 2007. The CLEF corpus: Semantic annotation of clinical text. *AMIA Annual Symposium proceedings. AMIA Symposium*, pages 625–629.

Angus Roberts, Robert J. Gaizauskas, Mark Hepple, George Demetriou, Yikun Guo, Ian Roberts, and Andrea Setzer. 2009. Building a semantically annotated corpus of clinical texts. *Journal of Biomedical Informatics*, 42(5):950–966.

Jeffrey M. Rohde, Andrew J. Odden, Catherine Bonham, Latoya Kuhn, Preeti N. Malani, Lena M. Chen, Scott A. Flanders, and Theodore J. Iwashyna. 2013. The epidemiology of acute organ system dysfunction from severe sepsis outside of the intensive care unit. *Journal of hospital medicine*, 8:243–247.

Thomas Brox Røst, Christine Raaen Tvedt, Haldor Husby, Ingrid Andås Berg, and Øystein Nytrø. 2018. Capturing Central Venous Catheterization Events in Health Record Texts*. In *IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2018*, pages 488–495, Madrid, Spain. IEEE Computer Society.

Aleksandar Savkov, John A. Carroll, Rob Koeling, and Jackie Cassell. 2016. Annotating patient clinical records with syntactic chunks and named entities: the harvey corpus. *Language Resources and Evaluation*, 50:523 – 548.

Seyedmostafa Sheikhalishahi, Riccardo Miotto, Joel T Dudley, Alberto Lavelli, Fabio Rinaldi, and Venet Osmani. 2019. Natural language processing of clinical notes on chronic diseases: Systematic review. *JMIR Medical Informatics*, 7(2):e12239.

Mervyn Singer, Clifford S. Deutschman, Christopher Warren Seymour, Manu Shankar-Hari, Djillali Annane, Michael Bauer, Rinaldo Bellomo, Gordon R. Bernard, Jean-Daniel Chiche, Craig M. Coopersmith, and et al. 2016. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*, 315(8):801–810.

Cristina Soguero-Ruíz, Kristian Hindberg, I. Mora-Jiménez, José Luis Rojo-álvarez, Stein Olav Skrøvseth, Fred Godtliebsen, Kim Mortensen, Arthur Revhaug, Rolv-Ole Lindsetmo, Knut Magne Augestad, and Robert Jenssen. 2016. Predicting colorectal surgical complications using heterogeneous clinical data and kernel methods. *Journal of Biomedical Informatics*, 61:87–96.

Brett R South, Shuying Shen, Makoto Jones, Jennifer Garvin, Matthew H Samore, Wendy W Chapman, and Adi V Gundlapalli. 2009. Developing a manually annotated clinical document corpus to identify phenotypic information for inflammatory bowel disease. *BMC Bioinformatics*, 10 Suppl 9:S12.

Irena Spasic and Goran Nenadic. 2020. Clinical text data in machine learning: Systematic review. *JMIR Medical Informatics*, 8(3):e17984.

Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. 2012. brat: a web-based tool for NLP-assisted text annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 102–107, Avignon, France. Association for Computational Linguistics.

Amber Stubbs and Özlem Uzuner. 2015. Annotating longitudinal clinical narratives for de-identification: The 2014 i2b2/UTHealth corpus. *Journal of Biomedical Informatics*, 58 Suppl:S20–S29.

Amber Stubbs and Özlem Uzuner. 2017. De-identification of medical records through annotation. In Nancy Ide and James Pustejovsky, editors, *Handbook of Linguistic Annotation*, pages 1433–1459. Springer Netherlands, Dordrecht.

Weiyi Sun, Anna Rumshisky, and Ozlem Uzuner. 2013. Annotating temporal information in clinical narratives. *Journal of Biomedical Informatics*, 46 Suppl:S5–S12.

Özlem Uzuner, Imre Solti, Fei Xia, and Eithon Cadag. 2010. Community annotation experiment for ground truth generation for the i2b2 medication challenge. *Journal of the American Medical Informatics Association*, 17(5):519–523.

Özlem Uzuner and Amber Stubbs. 2015. Practical applications for natural language processing in clinical research: The 2014 i2b2/UTHealth shared tasks. *Journal of Biomedical Informatics*, 58 Suppl:S1–S5.

Qiang Wei, Amy Franklin, Trevor Cohen, and Hua Xu. 2018. Clinical text annotation - what factors are associated with the cost of time? *AMIA Annual Symposium proceedings. AMIA Symposium*, 2018:1552–1560.

Fei Xia and Meliha Yetisgen-Yildiz. 2012. Clinical corpus annotation: challenges and strategies. In *Proceedings of the Third Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM'2012) in conjunction with the International Conference on Language Resources and Evaluation (LREC 2012)*, pages 32–39.

Melissa Y. Yan, Lise Tuset Gustad, Lise Husby Høvik, and Øystein Nytrø. 2023. Terminology and ontology development for semantic annotation: A use case on sepsis and adverse events. *Semantic Web*, 14(5):811–871.

Melissa Y. Yan, Lise Husby Høvik, André Pedersen, Lise Tuset Gustad, and Øystein Nytrø. 2021. Preliminary processing and analysis of an adverse event dataset for detecting sepsis-related events. In *IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2021, Houston, TX, USA, December 9-12, 2021*, pages 1605–1610. IEEE.

Wen-wai Yim, Sharon Kwan, and Meliha Yetisgen. 2015. In-depth annotation for patient level liver cancer staging. In *Proceedings of the Sixth International Workshop on Health Text Mining and Information Analysis*, pages 1–11, Lisbon, Portugal. Association for Computational Linguistics.

Li Zhang, Siyu Cao, Nicole Marsh, Gillian Ray-Barruel, Julie Flynn, Emily Larsen, and Claire M Rickard. 2016. Infection risks associated with peripheral vascular catheters. *Journal of Infection Prevention*, 17(5):207–213.

Paper D

Terminology and ontology development for semantic annotation: A use case on sepsis and adverse events

Melissa Y. Yan, Lise Tuset Gustad, Lise Husby Høvik, and Øystein Nytrø

Published in:
Semantic Web, May 2023, volume 14, number 5, pp. 811–871. DOI: 10.3233/SW-223226.

Semantic Web 14 (2023) 811–871
DOI 10.3233/SW-223226
IOS Press

Terminology and ontology development for semantic annotation: A use case on sepsis and adverse events

Melissa Y. Yan [a,*], Lise Tuset Gustad [b,c,d], Lise Husby Høvik [d,e] and Øystein Nytrø [a]

[a] *Department of Computer Science, Norwegian University of Science and Technology, Trondheim, Norway*
E-mails: melissa.yan@ntnu.no, nytroe@ntnu.no
[b] *Faculty of Nursing and Health Sciences, Nord University, Levanger, Norway*
E-mail: lise.t.gustad@nord.no
[c] *Department of Medicine, Levanger Hospital, Nord-Trøndelag Hospital Trust, Levanger, Norway*
E-mail: lise.gustad@hnt.no
[d] *Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, Trondheim, Norway*
E-mails: lise.t.gustad@ntnu.no, lise.hovik@ntnu.no
[e] *Clinic of Anaesthesia and Intensive Care, St. Olavs hospital, Trondheim University Hospital, Trondheim, Norway*
E-mail: lise.husby.hovik@stolav.no

Editors: Haridimos Kondylakis, FORTH-ICS, Greece; Praveen Rao, University of Missouri, USA; Kostas Stefanidis, Tampere University, Finland
Solicited reviews: Andre Lamurias, Aalborg University, Denmark; two anonymous reviewers

Abstract. Annotations enrich text corpora and provide necessary labels for natural language processing studies. To reason and infer underlying implicit knowledge captured by labels, an ontology is needed to provide a semantically annotated corpus with structured domain knowledge. Utilizing a corpus of adverse event documents annotated for sepsis-related signs and symptoms as a use case, this paper details how a terminology and corresponding ontology were developed. The Annotated Adverse Event NOte TErminology (AAENOTE) represents annotated documents and assists annotators in annotating text. In contrast, the complementary Catheter Infection Indications Ontology (CIIO) is intended for clinician use and captures domain knowledge needed to reason and infer implicit information from data. The approach taken makes ontology development understandable and accessible to domain experts without formal ontology training.

Keywords: Ontology development, clinical knowledge representation and reasoning, semantic annotation, sepsis, adverse events

1. Introduction

Many natural language processing (NLP) studies rely on annotated corpora to create models for text classification, information extraction, named entity recognition, question answering, summarization, and text generation.

*Corresponding author. E-mail: melissa.yan@ntnu.no.

Often, semantic annotation is done to capture domain knowledge within the text. Annotated corpora are frequently generated by annotators based on an annotation guideline, which provides the standard and rules for how to label text using specified terms. This annotation guideline is usually similar to a terminology, unless the corpus was annotated using an ontology. To further enrich an annotated corpus by capturing, reasoning, and inferring the underlying associated domain knowledge, an ontology is needed.

To demonstrate semantic annotation terminology and ontology development, the use case is based on clinician-presented needs for identifying sepsis from adverse events (AEs). Improperly cared for peripheral intravenous catheter (PIVC) medical devices can lead to unwanted and unintentional events that harm patients, such as AEs like phlebitis, bloodstream infections (BSIs), and sepsis. However, PIVCs are poorly documented in clinical records because of routine use among inpatients, and sepsis is also poorly documented outside the intensive care units (ICUs). The lack of explicitly documented concepts makes it challenging to directly detect and annotate PIVC-related phlebitis, BSI, and sepsis for quality surveillance to improve care. Thus, indications for the presence of PIVCs, phlebitis, infections, and sepsis are annotated instead. Those annotations are structurally preserved as a terminology, and the clinical knowledge required to reason about the indications is represented in an ontology. Additional details for the use case are provided in Section 2.

This paper provides a detailed and concrete description of the methodology utilized for ontology development of an annotated corpus based on a use case from the clinical domain. The main contributions presented are:

1. Describing the development process for constructing a terminology that can represent an annotated corpus. Specifically, a terminology for indexing annotated AE documents.
2. Presenting the development process for the terminology's corresponding ontology, which represents domain knowledge and allows inference of implicit knowledge in a specific domain. The corresponding ontology in the use case represents clinical domain knowledge specifically for annotated catheter-related and infection-related signs in AE documents.
3. Releasing a terminology and ontology that can be applicable to identifying and reasoning about sepsis in an AE corpus.

This paper significantly extends the papers [66] and [67], by adding an ontology with instances and including evaluation of the correctness and ability to answer competency questions. In addition, instances from the annotated corpus in [67] were added into the terminology, the terminology was evaluated using competency questions, and an ontology was developed to answer competency questions with clinical knowledge.

Based on the presented use case in Section 2 and objective in Section 2.2, an Annotated Adverse Event NOte TErminology (AAENOTE) and corresponding Catheter Infection Indications Ontology (CIIO) were developed. Section 5 details the terminology construction and development process to represent annotated documents, and Section 6 presents the results and evaluations for the AAENOTE. To address shortcomings in the terminology for annotated documents, Section 7 describes the ontology development process for domain knowledge representation of documented content. Additionally, Section 8 presents the results and evaluations for the CIIO. Finally, Section 9 discusses the findings, limitations, representations, accessibility, and utility.

2. Use case background and motivation

2.1. Sepsis from peripheral intravenous catheter-related phlebitis and infection adverse events

As the most commonly used medical device in hospitals, PIVCs are inserted into the peripheral vein to administer intravenous (IV) fluids, IV medications, and blood transfusions [1]. Improper management of PIVCs or the infusions connected to the PIVC can lead to phlebitis, which is either infectious, mechanical, or chemical inflammation of the vein [12,23,47]. Independent of cause, all PIVC phlebitis share many symptoms like redness and swelling near a patient's infusion insertion site for infectious, mechanical, or chemical phlebitis, making it difficult to distinguish. Furthermore, all PIVC-related phlebitis causes AEs like significant pain, PIVC failure that delays treatment, and compromises future venous access. Infectious phlebitis may lead to BSI due to: 1. migration of bacteria at the insertion site, 2. bacteria migrating through the catheter tract or catheter hub, 3. contaminated infusate, or 4. bacteria

from an existing infection in the bloodstream attaching to the catheter [69]. BSIs can potentially cause sepsis and occur when bacteria enter the bloodstream [26]. *Staphylococcus aureus* (*S. aureus*) is a lethal bacteria frequently found on skin that commonly causes BSIs [41], defined as a dysregulated host immune response to infection that results in organ failure and a mortality rate of 20% [53]. Approximately 7.6% to 35% of *S. aureus* BSIs are caused by PIVCs [32].

Even though PIVCs are frequently used, they are routinely not documented in clinical records [1]. Moreover, sepsis is also poorly documented outside the ICUs [49]. This lack of documentation makes it challenging to perform retrospective and real-time systematic quality surveillance of PIVC-related phlebitis or BSIs to identify learning opportunities for improving PIVC care to lower phlebitis-related, BSI-related, and PIVC-related AE incidents. Hence, AE reports or documents, which are customarily used to report PIVC failures, were selected as the clinical text for this project. To capture documented observable patient states and infer underlying knowledge of PIVC-related phlebitis or BSI from clinical text, an ontology that models clinical knowledge representation and reasoning is necessary.

2.2. Use case objective

The use case objective was to develop a model for representing and reasoning about PIVC-related BSIs in the unstructured free-text of AE reports, describe the development process, and discuss the discoveries and limitations. From the research question "is there a connection between BSIs and PIVCs at the hospital?", competency question requirements for an ontology representing and reasoning about PIVC-related BSIs were identified by clinicians as follows:

1. Does patientA have phlebitis, and was it infectious, chemical, or mechanical phlebitis?
2. Does patientA have an infection?
3. Does patientA have a BSI?
4. How many patients have an infection or BSI?
5. Which patients have sepsis?
6. Does patientB have a catheter?
7. Does patientB have a PIVC?
8. How many catheters does patientB have, where are they, and why does patientB need them?
9. Does patientC have an infection and catheter? If so, was patientC's infection associated with a catheter?

3. Related work

3.1. Annotation, tagging, and ontologies for natural language processing

Many studies focus on the relationship between annotation, tagging, and ontologies for NLP development. These studies include annotating corpora, NLP extraction, and classification tasks. Below are some studies that have shared their findings, issues, and possible solutions.

Annotated using the Uberon multi-species anatomical ontology [34], the Colorado Richly Annotated Full-Text (CRAFT) Corpus is a resource for NLP development which is also semantically annotated with concepts from eight Open Biomedical Ontologies (OBO) and terminologies [4,6]. However, while annotating with the OBOs, they discovered that the OBOs are not developed for annotation because there are overlapping terms within the different OBOs, context-specific definitions, and semantic ambiguities. Additionally, some OBOs do not follow the OBO Foundry principle of using relations from the OBO Relation Ontology (RO) [57] to link concepts [5]. Therefore, to improve OBOs for semantic annotation of biomedical documents, the researchers proposed desirable ontology implementations such as, but not limited to, integrating overlapping OBOs terms, resolving ontology-specific ambiguities, and expanding relations [5].

A study comparing how anatomy ontologies are used for annotations discovered annotation and ontology issues [61]. Annotations from three public datasets were compared to anatomical terms in the Foundational Model

of Anatomy (FMA) [50,51] and Uberon [34] ontologies using the Zooma and Ontology Mapper software tools. Manual and semi-automated preprocessing were done to normalize terms, but there were few matches between the ontologies and annotations, mainly because of strict matching. Additionally, the user-provided annotation labels resulted in mismatches, such as annotating a phrase with multiple ontology terms or using an abbreviation or adjective for an anatomical part instead of the anatomical ontology term. Ontology issues include missing anatomical synonyms used by the annotators and differing anatomical terms in the ontologies because the ontologies are designed for different purposes and made by different design decisions. The study concluded that mapping terms to an ontology requires a large amount of time, effort, and manual curation. Furthermore, an ontology's design decisions and scope will affect users trying to match annotations to an ontology, and ontologies must be used to understand their potential.

The Unified Medical Language System (UMLS) is a collection of standard biomedical terminology [8], and it has been used to process text by extracting concepts, relations, and knowledge (i.e., link or annotate text with standard terminology) [2]. A software capable of finding and linking biomedical text to terminology concepts in the UMLS Metathesaurus is MetaMap [3]. However, the developers of MetaMap mention that improvement is required for detecting similar names, acronyms, and abbreviations and resolving ambiguities by possibly distinguishing concepts using word sense disambiguation.

In [37], an overview of studies using knowledge bases for entity coreference resolution were discussed. Among those studies was the OntoNotes project, which annotated a multilingual corpus for different levels of semantic structure in the text [25,44,45]. One of the annotation levels includes linking OntoNotes word senses to the Omega ontology [25,46,68]. Near-synonymous word sense pools were created by specialists who grouped sense distinctions from WordNet and dictionaries based on similar definitions. This enables machines to automatically tag senses more accurately and improves inter-annotator agreement due to difficulties determining WordNet distinctions directly in the text [68]. Before each sense pool was linked to a concept in the Omega ontology [42], each sense pool was verified by machine and humans [68].

3.2. Ontology development methods and evaluation

There are many ontology development methods, such as: the Enterprise ontology's Uschold and King [62], the TOronto Virtual Enterprise (TOVE) ontology's Grüninger and Fox [20], METHONTOLOGY [14], the On-To-Knowledge Methodology (OTKM) [60], and NeOn [59]. Of those methods, Uschold and King [62] and Grüninger and Fox [20] follow a sequential sequence of phases, whereas METHONTOLOGY [14], OTKM [60], and NeOn [59] are iterative. Whether sequential or iterative, the previously mentioned methods and 2 reviews [10,43] have shown that ontology development typically includes the phases: specification, conceptualization, formalization, implementation, and maintenance. During those phases, the knowledge acquisition, evaluation, and documentation phases also commonly occur either as a separate phase or concurrently with other phases. Appendix A provides a summary of the methods for each phase.

During and between phases, ontology evaluation judges an ontology's content to a reference, such as requirement specifications, competency questions [20], or the real-world [18,19]. Evaluation includes: (1) verification that the ontology has the correct informal natural language definition and formal ontology language definition, and (2) validation that the ontology represents the world it was created for [18,19]. In theory, there are many criteria for evaluation, but in practice, most studies only use the expressiveness and practical usefulness criteria [11]. Expressiveness is the number of competency questions answerable by the ontology [11,20,38], and practical usefulness is the number of problems an ontology can be applied to [11,38].

3.3. Relevant ontology resources

There lacks an ontology specifically for sepsis-related BSI, infection signs, anatomical locations, medical devices, and procedures. However, pre-existing ontologies can contain relevant concepts. For example, the Infectious Disease Ontology (IDO) [27] has sepsis and hospital-acquired infection entities. Sign and symptom entities are present in the Ontology for General Medical Science (OGMS) [39], and vital sign entities exist in the Vital Sign Ontology (VSO) [17,64]. Anatomical locations can be described using anatomical entities of the Foundational

Table 1
Overview of relevant resources for this study

Type	Resource	Relevant Concepts or Terms
Ontology	Infectious Disease Ontology (IDO)	Sepsis and hospital-acquired infection entities
	Ontology for General Medical Science (OGMS)	Sign and symptom entities
	Vital Sign Ontology (VSO)	Vital sign entities
	Foundational Model of Anatomy Ontology (FMA)	Anatomical entities
	Biological Spatial Ontology (BSPO)	Anatomical spatial location descriptor entities
	Ontology of Adverse Events (OAE)	Adverse event entities
	Open Biological and Biomedical Ontology (OBO) Relation Ontology	Relationship object properties
Terminology	National Cancer Institute Thesaurus (NCIT) terminology	Procedure and medical device terms
	International Classification for Nursing Practice (ICNP) terminology	Terms
Taxonomy	Nursing Interventions Classification (NIC) taxonomy	Terms
	NANDA International Nursing Diagnoses Classification taxonomy	Terms
Clinical Guideline	1998 Visual Infusion Phlebitis Scale	Visual infusion phlebitis grading scale
	2021 Infusion Therapy Standards of Practice Updates	Updated infusion therapy practice standards

Model of Anatomy Ontology (FMA) [15,51] and anatomical spatial location descriptor entities from the Biological Spatial Ontology (BSPO) [7]. Because AE reports are used, the adverse event entities in the Ontology of Adverse Events (OAE) [21,40] might also be relevant. Furthermore, relationship object properties in the Open Biological and Biomedical Ontology (OBO) Relation Ontology [48,57] could be used to link different entities together to capture more information.

In addition to ontologies, there are also potential relevant terminologies and taxonomies. For example, there are different procedure, medical device, and catheter terms in the National Cancer Institute Thesaurus (NCIT) [35,36]. Potential relevant standardized nursing practice language is found in the International Classification for Nursing Practice (ICNP) terminology [13], Nursing Interventions Classification (NIC) taxonomy [9], and NANDA International Nursing Diagnoses Classification taxonomy [22]. Furthermore, infusion phlebitis-related information can be obtained from the 1998 Visual Infusion Phlebitis Scale [30] and the 2021 Infusion Therapy Standards of Practice Updates [28]. Concepts or terms from these resources can be used to expand the ontology if deemed necessary by ontology users. The relevant ontologies, terminologies, taxonomies, and clinical guidelines can be found in Table 1.

4. Materials

4.1. Synthetic dataset

Documents for annotation are from an AE synthetic dataset. The documents are based on unstructured free-text AE notes within the extracted AE reports from the electronic incident reporting system at St. Olavs hospital, Trondheim University Hospital in Trondheim, Norway, between September 2015 to December 2019 [67]. The synthetic dataset contains 100 AE notes or documents manually created and verified by a nurse to ensure clinical data is anonymized. The Norwegian Regional Committees for Medical and Health Research Ethics (REK) has granted ethical approval to use AEs in this paper (approval no 2018/1201/REKmidt, 26814).

4.2. Annotated synthetic dataset

The synthetic documents were annotated by 8 annotators with clinical backgrounds over 4 annotation sessions [67]. Each annotator annotated 10 documents in session 1 and 20 documents in the remaining 3 sessions (i.e., 70 documents annotated over 4 annotation session). This resulted in 560 annotated synthetic AE documents, as shown in Fig. 1 (i.e., 8 annotators * 70 annotated documents over 4 sessions = 560 total annotated synthetic AE documents). In each annotation session, annotators followed the annotation guideline and used the Brat rapid annotation tool (BRAT) [58] to annotate the documents. Then, documents were evaluated and manually screened to

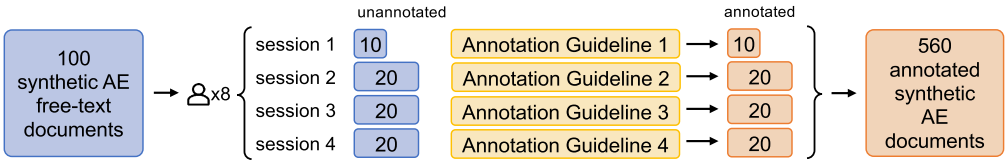


Fig. 1. From unannotated documents to an annotated corpus for populating instances in a terminology. 100 synthetic adverse event (AE) unstructured free-text documents were manually generated. Those synthetic documents were annotated by 8 annotators over 4 annotation sessions using revised annotation guidelines. Each annotator annotated 70 documents (i.e, 10 documents in session 1 and 20 documents in the remaining 3 sessions) to produce a total of 560 annotated synthetic AE documents.

identify ambiguities for revising the annotation guideline. This process was repeated 3 additional times with a new set of documents and a revised annotation guideline.

4.3. Annotation guideline

An annotation guideline was developed based on the clinical research question: Is there a connection between BSIs and PIVCs at the hospital? Discussions with nurses provided insight into how catheters can be distinguished explicitly by the name or implicitly based on the anatomical insertion site or procedure mentioned. This formed into four domain-specific *questions of interest*:

1. What are the different signs of infections, specifically for BSIs, sepsis, or infected PIVCs?
2. What are the signs for different types of catheters?
3. Where are the anatomical insertion sites of catheters?
4. What procedures, interventions, and activities can be related to catheter use?

Answers to domain-specific questions were then sorted into the following 7 main categories:

1. **Sign**: infection signs.
2. **Location**: anatomical insertion sites.
3. **Device**: signs of catheter types.
4. **Procedure**: procedures, interventions, or activities related to catheters.
5. **Sensitivity**: protected health information.
6. **Person**: individuals mentioned, such as patient, clinician, or relative.
7. **Whole**: label representing the span of the whole document and given to indicate if the document contains information about infection, BSI, sepsis, faulty device malfunctioning, catheter, PIVC, or has sensitive protected health information.

All categories except **Whole** have a hierarchy with more specific subcategories to capture detailed granularity from text (e.g., the **Person** category has a **Patient** subcategory). Furthermore, to capture relationships between categories for downstream analysis (e.g., infection sign at a specific location), the following four relationships to link categories were included:

1. **Person** $\xrightarrow{\text{Person has}}$ **Sign**, **Location**, **Device**, or **Procedure**.
2. **Procedure** $\xrightarrow{\text{Procedure uses}}$ **Device**.
3. **Sign** $\xrightarrow{\text{Caused by}}$ **Device** or **Procedure**.
4. **Sign**, **Device**, or **Procedure** $\xrightarrow{\text{Located nearby/on/in}}$ **Location**.

A preliminary annotation guideline was created using the seven categories and four relationships. Annotation guidelines from each of the 4 annotation sessions are available online[1] [67].

[1] https://folk.ntnu.no/melissay/ae-guidelines/

5. Terminology development for annotations

In annotations, categories are known as entities for labeling a span of words or phrases. Whereas in ontologies and terminologies, categories are known as classes. To separate the annotation guideline from the terminology and ontology, the annotation categories and entities are in **bold font**, and the terminology and ontology classes are in `typewriter font`.

5.1. Design decision for annotations

The terminology was developed using the bottom-up approach based on the annotation guideline refinement process from 4 iterations. Competency questions were not used to create this terminology. This terminology is meant to assist annotators who want to label text and allow users interested in performing downstream analyses to adjust the granularity of labels. The objective is solely to represent the annotated corpus and provide structure to the terminology used by annotators. Thus, included individuals are based on concrete examples from the annotated corpus. A simplified example of how annotation labels in annotated documents are added to the terminology as individuals is provided in Fig. 2.

Instead of reusing and re-defining existing ontologies, it was easier and simpler to develop a terminology based on what is documented in the data. For instance, although the FMA contains relevant anatomical parts, the ontology was too complex and detailed to be incorporated easily into the terminology to fit the use case's purpose. Additionally, the purpose was to include only concrete items documented in the terminology and not provide terminology for all existing items. By opting to simplify the terminology, it was easier to build the terminology directly based on the annotation guideline and then modify the terminology to incorporate feedback from discussions with clinicians.

5.2. Convert annotation guideline to terminology

The categories, attributes, and relationships in annotation guidelines described in Section 4.3 correspond to classes, data properties, and object properties in terminologies (Table 2 and Fig. 3).

The terminology was developed from the annotation guideline by translating each hierarchy of entities into a class hierarchy, using attribute information to add data properties, and converting relationships into object properties. During development, the terminology was modified to remove ambiguities by adding new class hierarchies and

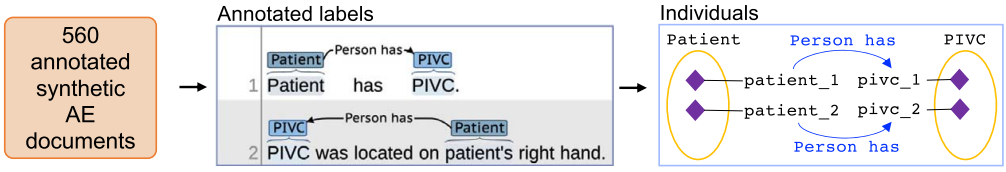


Fig. 2. Using an annotated corpus to populate individuals in a terminology. Each of the 560 documents were translated into an individual, and each label within a document was also translated into an individual. In the simplified example, an annotated document has 2 **patient** labels, 2 **PIVC** labels, and 2 $\xrightarrow{\text{Person has}}$ relationships linking the labels. Each label is converted into an individual (i.e., a purple diamond) of the corresponding class (i.e., `Patient` or `PIVC` yellow circle). Then the labels are linked using the $\xrightarrow{\text{Person has}}$ object property, similarly to how the $\xrightarrow{\text{Person has}}$ relationship links labels in the annotated text.

Table 2
Convert annotation guideline to terminology

Annotation Guideline	Terminology
Entity (category) hierarchies	Class hierarchies
Attributes which provide detailed entity information	Data properties
Relationships between entities	Object properties

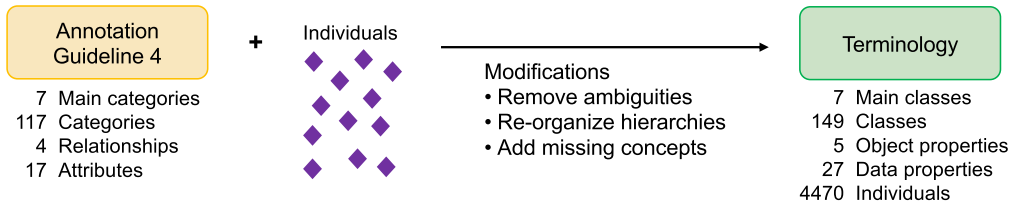


Fig. 3. Terminology development. The annotation guideline from the fourth annotation session was converted into a terminology. Annotation categories were converted into ontology classes, relationships into object properties, and attributes into data properties. Then the individuals of documents and labels were added. Additional modifications were incorporated as needed, such as removing ambiguities, re-organizing hierarchies, and adding missing concepts. This resulted in the AAENOTE which models and provides an index of annotated documents.

Table 3

Annotation Guideline vs Annotated Adverse Event NOte TErminology (AAENOTE). Annotation categories and entities are in **bold font** and the terminology and ontology classes are in `typewriter font`

Annotation Categories	Category Description	Sub-categories	Terminology Classes	Class Description	Subclasses
Sign	Infection signs	29	`Observation`	Documented clinical observation including symptoms, infection signs, and device malfunctions	41
Location	Anatomical insertion sites	17	`Anatomical location`	Anatomical location	25
Device	Signs of catheter types	16	`Medical device`	Treatment equipment or part	19
Procedure	Procedures, interventions, or activities related to catheters	31	`Procedure`	Procedure, intervention, or activity for catheter-related versus non-catheter related	36
Sensitivity	Protected health information	14	`Identifier`	Protected health information	14
Person	Individuals	3	`Person`	Individual	3
Whole	Label representing whole text indicating the note contains infection, BSI, sepsis, faulty device malfunctioning, catheter, PIVC, or sensitive information	0	`Annotated document`	Representation of an AE note's filename, annotation session, annotator, and annotated labels	4

modifying class names, object properties, and data properties. Specifically, to remove ambiguity between symptoms, infection signs, and device malfunction signs, the `Observation` class was introduced to encompass them. Additionally, the **Sensitive** category was revised to the `Identifier` class and the $\xrightarrow{\text{Caused by}}$ relationship was revised to the $\xrightarrow{\text{Is observed with}}$ relationship. A summary of the converted 7 main classes can be found in Table 3.

The results from all 4 annotation sessions were included as individuals in the terminology, but the terminology only reflects results based on the last annotation guideline. To accommodate revisions in the annotation guideline, annotation categories that were revised in the guideline are updated in the terminology correspondingly. For instance, removed annotation categories are reflected by changing the granularity of the removed category to a higher level. Annotation categories can also be re-organized to become subclasses of a different class. Moreover, newly added annotation categories are directly added as new classes in the terminology. An example is depicted in Fig. 4.

Although the terminology was not developed to answer competency questions, the competency questions were still used to determine what could be found in annotated documents. To answer competency questions, the annotated documents were imported into the terminology as individuals.

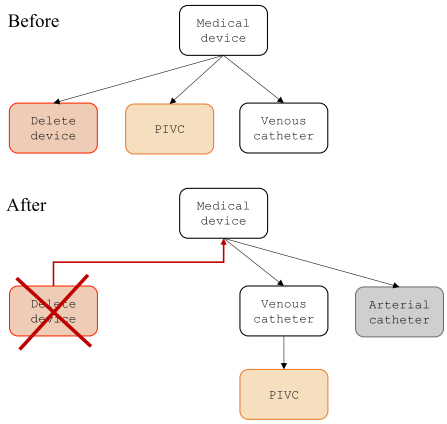


Fig. 4. Handling annotation guideline revisions in the terminology. If the **Delete device** annotation category in red is removed from the annotation guideline, then the `Delete device` class is removed from the terminology and all individuals of the `Delete device` terminology class are now part of the superclass `Medical device`. If the **PIVC** (peripheral intravenous catheter) annotation category in orange is re-organized to become the sub-category of **Venous catheter**, then the `PIVC` class is now a subclass of `Venous catheter` and the individuals remain as part of the `PIVC` class. If the **Arterial catheter** annotation category in gray is added, then the `Arterial catheter` class is added to the terminology and the corresponding individuals will be added as well.

6. Terminology results for annotations

6.1. Annotated Adverse Event NOte TErminology (AAENOTE)

Annotated AE documents and their annotations are modeled by the Annotated Adverse Event NOte TErminology (AAENOTE). To increase accessibility, the terminology is in both English and Norwegian. There are 149 classes, 5 object properties, 27 data properties, and 4470 individuals. The 7 classes which form the main hierarchies are:

1. `Observation`: Any sign or symptom that can be monitored.
2. `Anatomical location`: Any anatomical body part, organ, or relative position of the body.
3. `Medical device`: Any instrument, device, or equipment used for a medical purpose.
4. `Procedure`: Any procedure, intervention, or activity related to catheters.
5. `Identifier`: Protected health information that can be used to identify an individual.
6. `Person`: An individual, such as a patient, clinician, or relative.
7. `Annotated document`: Annotated adverse event document metadata and labels.

Relationships between the 7 class hierarchies can be formed using the following 5 object properties:

1. `Person` $\xrightarrow{\text{Person has}}$ `Observation`, `Anatomical location`, `Medical device`, or `Procedure`.
2. `Procedure` $\xrightarrow{\text{Procedure uses}}$ `Medical device`.
3. `Observation` $\xrightarrow{\text{Is observed with}}$ `Medical device` or `Procedure`.
4. `Observation`, `Medical device`, or `Procedure` $\xrightarrow{\text{Located nearby/on/at/in}}$ `Anatomical location`.
5. `Annotated document` $\xrightarrow{\text{Has label}}$ `Anatomical location`, `Identifier`, `Medical device`, `Observation`, `Person`, or `Procedure`.

An example showing AAENOTE representing an annotated document using parts of the class hierarchies and class properties is shown in Fig. 5. The complete class hierarchies of AAENOTE are in Appendix B.1.

Each annotated AE note or document is an individual of the `Annotated document` class and can have object properties linking the AE document to individual labels from the other 6 class hierarchies. Additionally, each AE

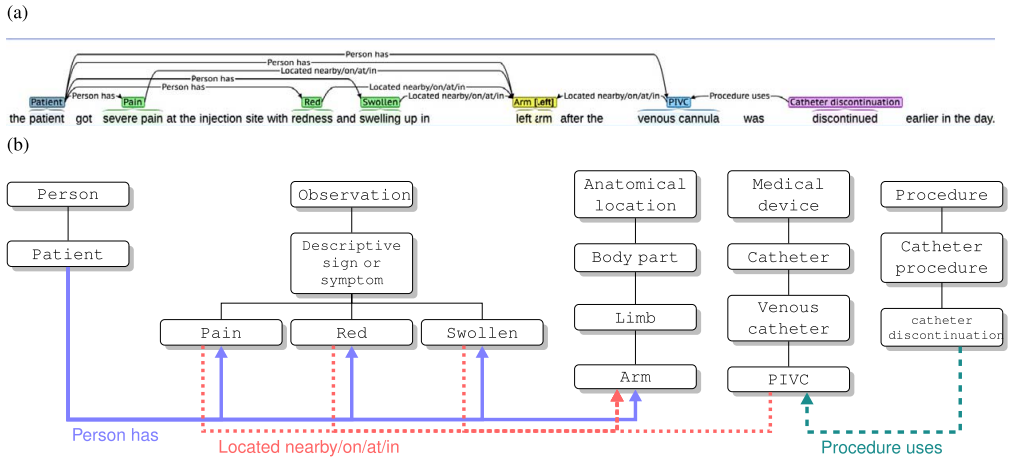


Fig. 5. Annotated Adverse Event NOte TErminology (AAENOTE) representing an annotated document. (a) Example of an annotated document with annotation categories and relationships that link the categories together. (b) Part of the terminology class hierarchies used within the annotation example are shown in the white boxes. The 3 annotated relationships (i.e., $\xrightarrow{\text{Person has}}$, $\xrightarrow{\text{Located nearby/on/at/in}}$, and $\xrightarrow{\text{Procedure uses}}$) are represented by the 3 object properties that link the classes together. Object properties are shown using the thicker colored lines with arrows.

document has data properties for the filename, annotation session, and annotator. The individuals of the other 6 hierarchies can also have object properties and data properties if an annotator provides that information.

6.2. AAENOTE evaluation

The purpose of AAENOTE is to model and provide semantic meaning to annotated AE notes or documents. The terminology was not developed based on competency questions, but it would be interesting to see what competency questions could be answered. Hence, AAENOTE was evaluated using the competency questions as requirements. Competency questions using AAENOTE can only be answered based on explicitly annotated classes or subclasses. Words or phrases that lack annotation are excluded from this terminology. Thus, only the annotation category labels provided by annotators are included as individuals of the corresponding classes.

Knowledge represented by the terminology can either be found explicitly, based on the direct classes and relationships, or be inferred implicitly, based on underlying concrete knowledge and indirect classes and relationships. For example, the competency question "Does patientA have an infection?" can be answered explicitly by finding an individual of the patient class who has an infection (i.e., `Patient` $\xrightarrow{\text{Person has}}$ `Infection`). It can also be answered implicitly by finding an anatomical location that has an infection (i.e., `Infection` $\xrightarrow{\text{Located nearby/on/at/in}}$ `Anatomical location`) because an anatomical location, in this terminology, must be part of a person. However, if the infection is not explicitly mentioned, this terminology cannot implicitly determine what other observations combined indicate an infection.

SPARQL query results vary depending on the clinician's interest in knowing how many instances a patient has for one class or a combination of that one class with other classes. For example, to find how many patients explicitly have an infection, the query can be written to find all instances where either: 1. individuals of the patient class have the object property "person has" to an individual of the infection class (i.e., `Patient` $\xrightarrow{\text{Person has}}$ `Infection`), or 2. the individuals of the patient class have the object property "person has" to an individual of the infection class and/or an individual of the observation class or it's subclass (i.e., `Patient` $\xrightarrow{\text{Person has}}$ `Infection` and other `Observation(s)`). The different number of instances is shown in Table 4. The first query has 14 instances where a patient has an infection regardless of other observations. In contrast, the second query divides

Table 4
AAENOTE SPARQL query result: patient has infection

			Person	Observation												
Query Result	Instances	# Observations	patient	blood pressure	body temp	c reactive protein	conciousness level	infection	neurological and physiological	observation	pulse	red	respiratory rate	sepsis	swollen
1	14	1	✓					✓							
2	1	1	✓					✓							
3	1	2	✓				✓	✓							
4	1	2	✓					✓				✓			
5	1	2	✓					✓					✓		
6	1	3	✓		✓	✓		✓							
7	1	3	✓		✓			✓					✓		
8	1	3	✓			✓		✓	✓						
9	1	3	✓				✓	✓		✓					
10	2	3	✓					✓				✓			✓
11	1	4	✓			✓	✓	✓		✓					
12	1	5	✓	✓	✓	✓		✓			✓				
13	1	6	✓	✓	✓	✓		✓			✓			✓	
14	1	8	✓	✓	✓	✓		✓	✓	✓	✓		✓		

Query result 1: `Patient` $\xrightarrow{\text{Person has}}$ `Infection`.
Query result 2–14: `Patient` $\xrightarrow{\text{Person has}}$ `Infection` and other `Observation(s)`.

the 14 instances into Query Result 2–14 to show the number of instances where a patient has an infection with different combinations of other observations. To implicitly find patients who have an infection, the infection must be located at an anatomical location of a person. Queries about infections at an anatomical location can be written using `Infection` $\xrightarrow{\text{Located nearby/on/at/in}}$ `Anatomical location` or using `Infection` and other `Observation(s)` $\xrightarrow{\text{Located nearby/on/at/in}}$ `Anatomical location` if the clinician is curious about additional observations that were documented with the infection. In AAENOTE, there is only 1 instance where an anatomical location (i.e., skin) has an infection, as shown in Query Result 1 of Table 5. However, it is more informative for clinicians to look at additional potential observations that can indicate infection around a certain location; Table 5 Query Results 2–10 provide other observations located on skin. In AAENOTE, there lacks clinical knowledge required to find indications of infections and catheters.

Overall, explicit queries and basic implicit queries in AAENOTE can answer 7 of the 9 competency questions. However, these queries still lack the clinical knowledge needed to include more implicit queries by combining additional observations, anatomical locations, and/or procedures to identify indications. In Appendix B, Table 7 provides the terminology classes and relationships used to form explicit and implicit queries to explicitly and implicitly answer each competency question. Additionally, concrete underlying knowledge used to make inferences is also provided. Results can be found in Appendix B.

7. Ontology development for domain knowledge

7.1. Design decision for domain knowledge

In this use case, the clinicians' need is to focus on identifying and inferring a patient's state based on documented observations in AE documents related to PIVCs and BSIs. A patient's underlying state can be measured by monitoring devices that measure vital signs (e.g., blood pressure, pulse, and respiratory rate) or exhibited by observable

Table 5
AAENOTE SPARQL results for observations located at the skin

Query Result	Instances	Observations					
		edema	infection	pain	red	swollen	warm
1	1		✓				
2	3			✓			
3	3					✓	
4	2			✓		✓	
5	2			✓	✓	✓	
6	2				✓		
7	2				✓	✓	
8	1						✓
9	1				✓		✓
10	1	✓			✓	✓	

Query result 1: `Infection` $\xrightarrow{\text{Located nearby/on/at/in}}$ `Anatomical location`.
Query result 2–10: `Observation` $\xrightarrow{\text{Located nearby/on/at/in}}$ `Skin`.

signs and symptoms (e.g., pain, fever, chills, and mobility impairment). Those measurable and observable signs and symptoms are then documented by clinicians in the electronic health record (EHR) to record patient conditions and communicate with other clinicians. When an AE incident could have or has happened, clinicians will go through the documentation to recall what occurred and report it in a separate AE document.

To limit the scope of modeling the clinical knowledge ontology, 15 documents were used as examples to form the classes and individuals. Each document was split into sentences to identify catheter and infection indications at the sentence-level and document-level. At the sentence-level, individual sentences were presented to clinicians who determined what observations, anatomical locations, or procedures within the text are needed to determine catheter and infection indication. Only clinician-identified sentences with indications were included as individuals in the ontology. At the document-level, individual sentences from a document were presented together, allowing clinicians to identify indications based on additional information from a more complete documented story. Presenting the document as separate sentences allowed clinicians to identify concepts within a limited example to determine what can and cannot be determined based on limited information. Whereas, allowing a clinician to see the whole document presented more possibilities and helped identify necessary data combinations for indications of catheters and infections.

The focus of the ontology includes catheter indications and the clinicians have identified that it is important to identify infusion phlebitis. Thus, infusion phlebitis was included in the ontology as rules based on the 1998 Visual Infusion Phlebitis Scale [30] mentioned in the 2021 Infusion Therapy Standards of Practice Updates [28]. Furthermore, causality is not within the scope because the exact reason for chemical and mechanical reactions resulting in infection-like signs are more likely found at the body's cellular or genetic-level in pathophysiology studies [29,55] and unlikely to be documented in AE documents.

Anatomical locations in this ontology were kept simple and similar to the AAENOTE. Clinical guidelines for catheter insertion into anatomical locations are very specific (e.g., a central venous catheter is inserted in the jugular vein until it reaches the superior vena cava [31]) because clinical guidelines provide instructions on how to perform a task properly. However, clinical documentation is more general (e.g., central venous catheter in the chest) because this is common clinical knowledge, and the documentation is written for other clinicians to understand. To match the ontology with available documented data, this ontology relies on general anatomical location terminology. If clinicians deem it necessary, clinical guidelines can be included in a separate ontology focused on identifying catheter locations based on clinical guidelines and the FMA. Inclusion of clinical guidelines to identify specific catheter insertion sites and placement requires anatomical knowledge. For instance, to identify a central venous

catheter's general anatomical location using a clinical guideline and the FMA anatomy ontology, the ontology would need to:

1. Identify the jugular vein insertion site and the superior vena cava placement.
2. Infer that the jugular vein is in the neck and the superior vena cava is present within the superior and middle mediastinum [65], annotated as anatomical location chest.
3. Convert the terms into more general terms that match the available data (i.e., vena cava is in the chest).

7.2. Representing domain knowledge

Discussions with clinicians about example documents and indications formulated the classes, object properties, data properties, and rules within the ontology. Then, the provided indications were sorted and summarized to match the ontology closely. Afterward, indications were verified by clinicians and included in the ontology using SPARQL queries. A list of indications can be found in Appendix C.3.1 to Appendix C.3.7.

8. Ontology results for domain knowledge

8.1. Catheter Infection Indications Ontology (CIIO)

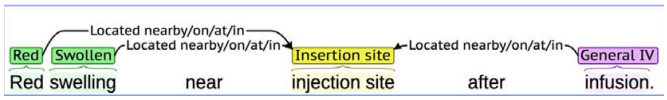
The Catheter Infection Indications Ontology (CIIO) represents clinical knowledge for signs of infections and catheters to identify PIVC-related BSIs and was developed to accompany the AAENOTE. Similar to AAENOTE, this ontology is also in both English and Norwegian. There are 57 classes, 10 object properties, 16 data properties, and 187 individuals. The 7 classes which form the main hierarchies are:

1. `Observation`: Any sign or symptom that can be monitored.
2. `Anatomical location`: Any anatomical body part, organ, or relative position of the body.
3. `Medical device`: Any instrument, device, or equipment used for a medical purpose.
4. `Procedure`: Any procedure, intervention, or activity related to catheters.
5. `Person`: An individual.
6. `Document`: Unstructured free-text report consisting of sentences documented to represent observable patient states.
7. `Sentence`: A set of words documented to represent observable patient states.

Relationships between the 7 class hierarchies can be formed using the following 10 object properties:

1. `Person`'s subclass `Patient` $\xrightarrow{\text{Patient has}}$ `Observation`, `Anatomical location`, `Medical device`, or `Procedure`.
2. `Procedure` $\xrightarrow{\text{Procedure uses}}$ `Medical device`.
3. `Observation` $\xrightarrow{\text{Is observed with}}$ `Medical device` or `Procedure`.
4. `Observation`, `Medical device`, or `Procedure` $\xrightarrow{\text{Located nearby/on/at/in}}$ `Anatomical location`.
5. `Document` or `Sentence` $\xrightarrow{\text{Contains}}$ `Observation`, `Anatomical location`, `Medical device`, `Procedure`, or `Person`. Additionally, only `Document` $\xrightarrow{\text{Contains}}$ `Sentence`.
6. `Procedure`'s subclass `General IV` $\xrightarrow{\text{Is combined with}}$ a different `General IV`.
7. `Procedure`'s subclass `General IV` $\xrightarrow{\text{Is IV for}}$ `Procedure`'s subclass `Infusion`.
8. `Procedure`'s subclass `General IV` $\xrightarrow{\text{Medication should have been}}$ a different `General IV`.
9. `Observation`, `Anatomical location`, `Medical device`, `Procedure`, or `Person` $\xrightarrow{\text{Is documented in}}$ `Document` or `Sentence`. Additionally, only `Sentence` $\xrightarrow{\text{Is documented in}}$ `Document`.
10. `Anatomical location` $\xrightarrow{\text{Location has}}$ `Observation`, `Medical device`, or `Procedure`.

(a)



(b)

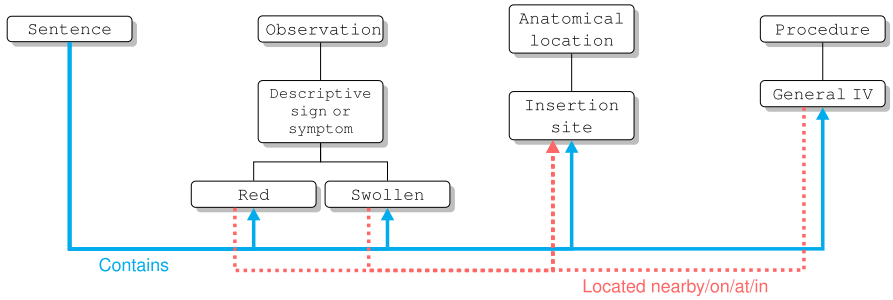


Fig. 6. Catheter Infection Indications Ontology (CIIO) clinical knowledge representation. (a) A sentence from a document used to identify documented clinical knowledge. Annotations are based on terms from the Annotated Adverse Event NOte TErminology (AAENOTE). (b) CIIO has a `Sentence` class and $\xrightarrow{\text{Contains}}$ relationship to link a sentence to documented observable patient states (i.e., AAENOTE terms). AAENOTE terms were used in CIIO to conceptualize the classes and object properties that represent documented knowledge and can be used in reasoning to identify catheters and infections. Part of the ontology class hierarchies and object properties used in knowledge representation are shown. Classes are in white boxes and object properties are shown using thicker colored lines with arrows.

An example showing how a sentence is represented using the class hierarchies and class properties of CIIO to model documented clinical knowledge is shown in Fig. 6. The complete class hierarchies of CIIO are in Appendix C.1.

Each sentence documented in the report is an individual of the `Sentence` class. `Sentence` class individuals contain `Observation`, `Anatomical location`, `Medical device`, `Procedure`, or `Person` individuals present within the text. An individual of the `Document` class contains the `Sentence` individuals that form it and the content from those sentences. Similar to AAENOTE, individuals of `Observation`, `Anatomical location`, `Medical device`, `Procedure`, and `Person` can also have object properties and data properties.

8.2. CIIO evaluation

Designed to capture and reason about clinical catheter-related and infection-related signs and symptoms documented in an AE report, the CIIO provides the missing clinical domain knowledge for the AAENOTE. CIIO can answer 8 of the 9 competency questions based on assumptions and indications. The assumptions are that 1 AE document represents 1 patient and all sentences within a document are likely to describe concepts within the same event (Appendix C.2). Indications for catheters and infections are provided in (Appendix C.3). Additionally, the ontology classes and relationships used to answer each competency question is detailed in Appendix C.4.

9. Discussion

9.1. Ontology development method comparison

The clinical problem drove this study, and the objective was not to apply an ontology development method. Hence, a specific ontology development method was not applied. However, certain steps taken are similar to the pre-existing methods and this study does include the typical phases of specification, conceptualization, formalization,

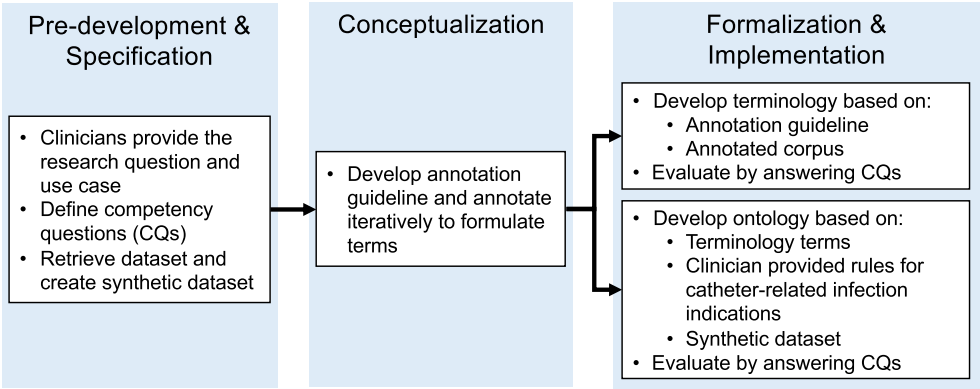


Fig. 7. Development phases for Annotated Adverse Event NOte TErminology (AAENOTE) and Catheter Infection Indications Ontology (CIIO).

implementation, maintenance, knowledge acquisition, evaluation, and documentation. An overview of the process is shown in Fig. 7, and similarities to other methods can be found in Appendix A.

During the pre-development and specification phases of the terminology, clinicians provided the research question and use case. Those were utilized to define the competency questions. Additionally, the AE dataset was retrieved, and an AE synthetic dataset was created. The conceptualization phase was performed by iteratively developing the annotation guideline and annotation sessions. Afterward, the formalization and implementation phases of the terminology were developed iteratively based on the annotation guideline and using instances from the annotated corpus to answer competency questions for evaluation. Knowledge acquisition occurred during all phases with insight, guidance, and feedback from clinicians. Documentation is provided in the annotation guidelines, the annotated corpus, and the evaluation of competency questions. The annotation guidelines document changes in terms over time, and the annotated corpus documents knowledge acquisition from the text. Answers to each competency question are documented using natural language for clinicians and SPARQL queries for computer scientists.

Ontology development is similar to the terminology's pre-development, specification, and conceptualization phases. However, the formalization and implementation phases differ. The ontology iteratively incorporated clinical knowledge that can be annotated in AE documents using terminology terms to answer competency questions for evaluation. Knowledge acquisition was provided through the annotated corpus, clinician-provided catheter indication rules, and clinician-provided publications containing phlebitis rules. Additionally, clinicians iteratively reviewed and verified documented sentences to match the rules and competency questions. Ontology documentation includes the assumptions (Appendix C.2), rules for catheter and infection indications (Appendix C.3), and how competency questions were answered (Appendix C.4).

Although UMLS includes clinical terminology (i.e., SNOMED CT and ICD-10), the terms are often a combination of different concepts. For example, phlebitis has many options and is combined with different locations, such as "phlebitis of the lower limb vein," "phlebitis of the portal vein," and "retainal phlebitis." Additionally, swollen has many options, such as "foot swelling," "swollen nose," and "tongue swelling." The June 10, 2022 version of SNOMED CT has 361,907 classes.[2] It would require extensive time and effort to determine which classes are suitable for our purpose and to maintain a pre-determined class hierarchy. Introducing UMLS terminology would make it difficult for annotators to determine which term to use, introduce ambiguities for the use case, and decrease the precision needed. Furthermore, the UMLS MetaMap is software that finds and links biomedical text to terminology concepts [3]. Unfortunately, that software is for biomedical text and not clinical text. Clinical text differs from biomedical text because it is often ungrammatical and ambiguous, with many shorthand abbreviations and

[2] https://bioportal.bioontology.org/ontologies/SNOMEDCT/?p=summary

acronyms [33]. Additionally, the MetaMap developers have mentioned that detecting abbreviations and acronyms needs improvements [3].

9.2. AAENOTE scope and limitations

The clinician asks questions about the condition of a physical patient and the patient's PIVCs, but AAENOTE is about words in the AE document regarding a patient event. The correspondence between clinical condition and document content is represented by CIIO. So, the answer to a question about a patient's condition will be answered using the document's annotated text. Understanding a query means translating it from clinical concepts to concepts within the document's content. Here, the terminology is used to fit and answer questions. SPARQL queries can answer most competency questions, and the results can be used as consistency checks. For example, SPARQL can be used to count and make quantitative queries about the number of catheters and devices. Likewise, qualitative results enabled clinicians to verify if results matched their expectations of clinical events (i.e., the anatomical location of specific catheters) or why the AE was reported (i.e., incorrect medical devices used in a particular procedure).

There are several limitations to the AAENOTE. Although this terminology does not cover sepsis, it does cover events that could lead to sepsis. This terminology lacks the clinical knowledge required to answer several competency questions more in-depth. Moreover, it is not always possible to determine what a patient has because of the document's content or provided annotations. Most documents do not explicitly mention a patient because these are AE documents, and it is often implied that the adverse event has happened to a patient. Annotators will often not link the patient to all possible observations, anatomical locations, medical devices, or procedures because typically, one AE document refers to one event or patient. Furthermore, referent tracking and resolution are not handled by AAENOTE. Thus, multiple mentions of a label or individual do not indicate whether it is the same item or a different item. For example, given the annotated document in Fig. 2, the terminology cannot determine if the 2 `Patient` individuals refer to the same patient or not because each label is an individual. Similarly, the same also applies to the 2 `PIVC` individuals. In this example, a query counting how many patients have a PIVC will answer 2. However, based on the context, both sentences in the document likely refer to the same patient and PIVC and the answer should be 1.

9.3. CIIO scope and limitations

The CIIO is an abstract ontology with instances populated using the terminology. Only parts of the AAENOTE necessary for creating queries with clinician-provided indications were included or extended. This provides flexibility, allows for easier ontology maintenance, and separates the needs of clinicians who use CIIO and annotators who use AAENOTE.

Based on assumptions and indications, competency questions can be answered using SPARQL queries. The queries retrieve documents and translate the content for the user by identifying concepts necessary to answer the competency questions. Thus, the retrieved documents and concepts can provide sufficient information for clinicians to further decipher retrieved answers. For example, the exact reason why a patient needs a catheter cannot be determined by the query unless there is a direct relationship (i.e., `Procedure` $\xrightarrow{\text{Procedure uses}}$ `Medical device`) because the list of indications does not provide reasons for catheter usage. However, the clinician can view the retrieved list of medical devices and procedures within a document to determine why the medical devices were required.

The lack of detailed documentation inhibits the query's ability to answer certain questions. This includes why a patient needs a catheter as previously stated, counting catheters within a patient, and where the catheters are located in a patient. Counting the exact number of catheters per document is not possible because multiple sentences within a document could be describing the same catheter or multiple procedures could use the same catheter. The exact anatomical location of catheters per document cannot be determined for several reasons. First, multiple sentences within a document could be describing the same catheter at the same location but with more general terms (i.e., arm

instead of hand). Second, the location's position not being documented makes it difficult to distinguish if a body part is on the same side. Finally, various procedures can be performed at the same location. For example, given an example document, "The patient received IV fluids in elbowA and IV antibiotics in right handB. Right armC showed signs of phlebitis." Here, handB is part of armC because both are on the right side, but elbowA might or might not be part of armC. Additionally, armC is likely a more general term for handB. Furthermore, an additional anatomical ontology is needed to infer the possible locations based on catheter type.

9.4. Purpose of separate terminology and ontology

Even though the ontology uses terms from the terminology, the terminology and ontology are separate. They are separate because of their different purposes and functionalities. Additionally, separation provides downstream analysis flexibility for researchers. It also simplifies evaluation and allows for easier maintenance. Furthermore, separation enables a better understanding of the terminology's and ontology's limitations.

The terminology and ontology were developed for different purposes using different methods. AAENOTE is intended to be useful for annotators who are annotating documentation and a way to provide them a structured terminology with varying granularity. In comparison, CIIO is intended to be used by clinicians to clinically reason about a patient state. The design process of AAENOTE is heavily based on the explicit terms used in annotations and not on competency questions. It uses the bottom-up method and the annotation guideline development process to capture semantic annotations. In contrast, the design process of CIIO is based on the competency questions, which focus on patient states. Designed with a top-down method, it is based on concepts naturally used by clinicians to describe patients. Thus, the terminology and ontology have different purposes and functionalities.

Separating the terminology from the ontology enables annotators to annotate concepts with standard terms and clinicians to reason about the annotated concepts. Here, the ontology does not impact the terminology annotators can use. Instead, the ontology provides knowledge for the terminology. Thus, our methods avoid the significant amount of time, effort, and manual curation previously required to map terms to an ontology [61]. Instead, our ontology utilizes concepts in the terminology and is limited by the competency questions, clinical guidelines used, and clinician-provided rules. In downstream analyses, researchers can freely choose to use the terminology to quickly retrieve documents with specific annotations, the ontology to reason and infer clinical knowledge, or both.

The terminology indicates concepts annotated in documents. Using terms from the terminology ensures that the included clinical knowledge within the ontology represents the knowledge documented in the text that can be annotated. As the ontology develops further, it is possible to conceptualize additional terms required to answer the competency questions. Those terms can then be added to the terminology and annotation guideline for additional data curation.

In this paper, separating the terminology made it easier and quicker to evaluate syntax and semantics because the ontology only has 187 instances compared to the terminology's 4470 instances. Mixing the indexed annotation terminology with a clinical knowledge ontology would be outside the ontology's scope, decrease ontology reusability, and increase the complexity of ontology maintenance. Additionally, the terminology can cover a broader scope of documents not in the ontology. Finally, using competency questions to evaluate the terminology and ontology separately reveals the distinct limitations of both. The inability to answer competency questions can be due to either the lack of knowledge or lack of necessary content within the data.

9.5. Representing annotated data and revisions

Annotating data provides data meaning, and the corresponding annotation guideline and terminology provide additional structured semantic meaning. Additionally, the terminology can represent knowledge and disambiguate annotation entities and relationships. Each annotation session uses a slightly different annotation guideline that has been revised based on the previous annotation session. Hence, revisions in the annotation guidelines include added, re-organized, and removed categories. Since results from 4 different annotation sessions are included as individuals in the AAENOTE, this indicates the terminology can handle different versions of annotated data while preserving semantic meaning. It is also possible to easily customize the granularity (i.e., superclasses, classes, or subclasses) and extend or retract the terminology based on clinician needs without breaking the terminology.

To alleviate the problem with overlapping terms and ambiguities experienced by [5] and remove mismatches between the annotations and the terminology experienced by [61], all annotators in our study could only use provided annotation labels from the terminology. Using concrete concepts from the annotation guideline based on what can be found in the documentation instead of other pre-existing ontologies lowers the complexity and simplifies the terminology.

9.6. Representing annotated documents the way clinicians view patients

The AAENOTE is a terminology that provides an index of what is annotated in a clinical document. It is not used to design a language's syntax, grammar, or terms because AAENOTE is a terminology for understanding the language and underlying meanings. Instead, it is the interpreted formalized language that has been translated into basic statements for reasoning. To capture relevant information, the underlying document was represented by annotated labels and relationships for the task of question answering and text understanding instead of solely retrieving information. Hence, the terminology focuses only on items of interest and is blind to items not within the terminology.

The corresponding CIIO is an ontology that models clinical knowledge missing from AAENOTE. It provides the missing clinical knowledge required to reason about the presence of catheters and infections documented in clinical text. Although the data modeled is documented text, it enables clinicians to think about the data as an individual patient because they already do this routinely when documenting patient states.

9.7. Understandability and accessibility for domain experts

The approach in this study made ontology development understandable and accessible for the domain experts without formal ontology training. Furthermore, the employed approach made it possible for clinicians to understand and be part of the design process. In practice, the approach was a necessity to progress in developing the ontology to incorporate clinical knowledge.

9.8. Clinical utility

The collected competency questions and requirements are largely met. Thus, the main objective of developing a terminology and ontology that clinicians and hospital systems can use to get a systematic overview of identifying and reasoning about PIVC-related phlebitis, infection, and sepsis in an AE corpus has been met. Furthermore, our ontology is a step toward automated and continuous quality control that move beyond today's focus on repeated point prevalence quality controls, like the Peripheral Intravenous Catheter mini Questionnaire (*PIVC-miniQ*) [24].

The developed ontology is of value for sepsis because of its purpose, clinician involvement during development, and intended use. The ontology focuses on identifying indicators of catheter-related phlebitis or infections that can lead to sepsis by utilizing the clinicians' documentation and perspectives. Throughout the whole development, clinicians were involved as the users, domain experts, and data annotators. Furthermore, clinical knowledge within the ontology was captured similarly to how clinicians ask questions, document observations, and view documents as patients. The intent is to eventually implement the ontology into a quality surveillance system to automatically detect the presence of PIVC-related phlebitis and BSIs to improve PIVC care and lower sepsis incidents. Thus, only documented content can be included as data, and the ontology must directly correspond to and represent concepts documented within the AE documents from the clinician's perspective.

10. Future work

For the sepsis-related use case, the synthetic AE dataset used for annotations is a placeholder for the real Norwegian AE dataset and clinical records from the EHR. Future work includes utilizing the current AAENOTE to annotate the real Norwegian AE dataset and clinical records. And to evaluate if clinical knowledge from the CIIO can still be applied and expanded on new data. Additionally, the ontologies could be applied to AE documents at other Norwegian hospitals to assess how similar documentation and knowledge are between different hospitals. The

ontologies can be directly translated to other Scandinavian languages (e.g., Swedish and Danish) and applied similarly at other Scandinavian hospitals. The design and representations are largely language-independent and should be easy to transform for English clinical text about adverse events. After all, international literature suggests that the phenomena related to PIVC and devices are language-independent [1]. It would also be possible to provide multi-language querying over multi-language AE documents to enable cross-language repositories [63]. Furthermore, supervised machine learning methods can be employed to identify PIVC-related BSIs and classify patients requiring additional monitoring.

11. Conclusion

The development process resulted in a terminology and an ontology, specifically, the Annotated Adverse Event NOte TErminology (AAENOTE) which models annotated classes in annotated documents and the Catheter Infection Indications Ontology (CIIO) which models clinical knowledge for catheter and infection indications. Although there is a clinical focus here, the methodology for creating a terminology from an annotation guideline for semantically annotated data and a domain knowledge ontology to represent knowledge can be utilized in other domains to provide additional semantic meaning to annotated datasets in other domains.

12. Data availability

The AAENOTE, CIIO, and SPARQL queries for this paper are in the GitHub repository branch "swj" of https://github.com/melissayan/aaenote_and_ciio. Detailed specifications for AAENOTE and CIIO were generated using WIzard for DOCumenting Ontologies (WIDOCO) [16] and are available in English and Norwegian at https://folk.ntnu.no/melissay/ontology/index.html.

Acknowledgements

The authors would like to acknowledge the eight annotators for producing the annotated clinical adverse event (AE) corpus. Funding was provided by the Computational Sepsis Mining and Modelling project through the Norwegian University of Science and Technology (NTNU) Health Strategic Area. Additionally, ethical approval was granted by the Regional Committees for Medical and Health Research Ethics (REK), approval no. 26814.

Appendix A. Ontology development methods and evaluation similarities

There are various ontology development methods and the steps in this study have some similarities to other methods Table 6.

Appendix B. Annotated adverse event NOte TErminology (AAENOTE) hierarchy competency questions (CQs), SPARQL queries, and results

B.1. AAENOTE hierarchy

The AAENOTE classes and their subclasses can be found in the following figures:

1. Fig. 8 `Observation`
2. Fig. 9 `Anatomical location`
3. Fig. 10 `Medical device`
4. Fig. 11 `Procedure`
5. Fig. 12 `Identifier`
6. Fig. 13 `Person`
7. Fig. 14 `Annotated document`

Table 6
Ontology development methods versus this paper

	Uschold and King [62]	Grüninger and Fox [20]	METHONTOLOGY [14]	On-To-Knowledge [60]	NeOn [59]	This Paper
Pre-development				Feasibility study phase to identify problems.	Initiation phase to specify ontology requirements, intended use, users, and formal language in the form of competency questions and glossary of terms.	Use clinician identified research question to form competency questions. Form terms for annotation, perform data screening and pre-annotation, develop annotation guideline, schedule annotation, implement annotation sessions, and evaluate annotations.
Specification	Identify ontology purpose, intended use, and users.	Define ontology requirements as competency questions based on a scenario or user provided problem.	Specification phase to produce ontology specifications.	Kickoff phase to document ontology requirements and semi-formal ontology.	Part of pre-development.	Part of pre-development.
Conceptualization	Identify terms and relationships of interest.	Define terminology of ontology.	Conceptualization phase to build glossary of terms.	Refinement phase to refine semi-formal ontology and formalize ontology iteratively based on domain expert interviews.	Design phase to produce an informal and formal model to meet requirements.	Annotation and annotation guideline revision phases determine the terms and concepts that are documented in data.
Formalization	Codify ontology into formal language.	Specify terminology definitions using first order logic.	Part of conceptualization.	Part of conceptualization.	Implementation phase to implement the formal model into an ontology language.	Codify terminology and ontology in Protégé using OWL.
Implementation	Part of formalization.	Represent terminology in a formal language.	Implementation phase to codify ontology into formal ontology.			Part of formalization.
Management				Apply the ontology and manage it's evolution and maintenance.	Use the ontology to detect errors or missing knowledge for design phase of the next ontology version.	Terminology can be updated to a new annotation guideline version and its corresponding annotated corpus. Ontology can include new clinical knowledge.

Table 6
(Continued)

	Uschold and King [62]	Grüninger and Fox [20]	METHONTOLOGY [14]	On-To-Knowledge [60]	NeOn [59]	This Paper
Knowledge Acquisition			Interview experts and/or analyze text.	Knowledge creation, capture, retrieval and access, and use.	Knowledge is introduced by domain experts and ontology practitioners at different development phases.	Knowledge was acquired iteratively with users and domain experts. Discussions with users and domain experts resulted in annotation guideline revisions and discovering the knowledge needed for ontology reasoning and inferences. Annotations by domain expert annotators to capture knowledge from text.
Evaluation	Adapt what has been done in the knowledge based systems field for ontologies.	Evaluate the ontology by proving completeness theorems to answer competency questions.	Evaluate the ontologies, software environment, and documenation with the requirement specification document during and between each phase. Document how the ontology was evaluated, errors detected, and knowledge sources for evaluation.	Proposed technology-focused (i.e., the development tool's evaluation of syntax and semantics of the ontology and the evaluation of tools and applications for interoperability and scalability), user-focused (i.e., user satisfaction with the application and comparing the ontology based application to pre-existing ones), and formal evaluation.	Evaluate the ontology using 5 different tasks [54] which includes selecting individual ontology network components for evaluation, selecting the evaluation goal and approach, identifying the reference and evaluation metric, applying the selected evaluation, and presenting the combined results of the individual components.	Ability to answer competency questions and if the ontology is useful from the perspective of the user.
Documentation	Document ontology type and purpose		Document ideally all phases, knowledge acquisition, and evaluation.	Document kickoff phase ontology requirements.	Document ontology requirement specifications, ontology description, and evaluation.	Annotation guidelines record changes, annotations document knowledge acquisition, and evaluation documentation answers competency questions in a format understandable to the users and developers.

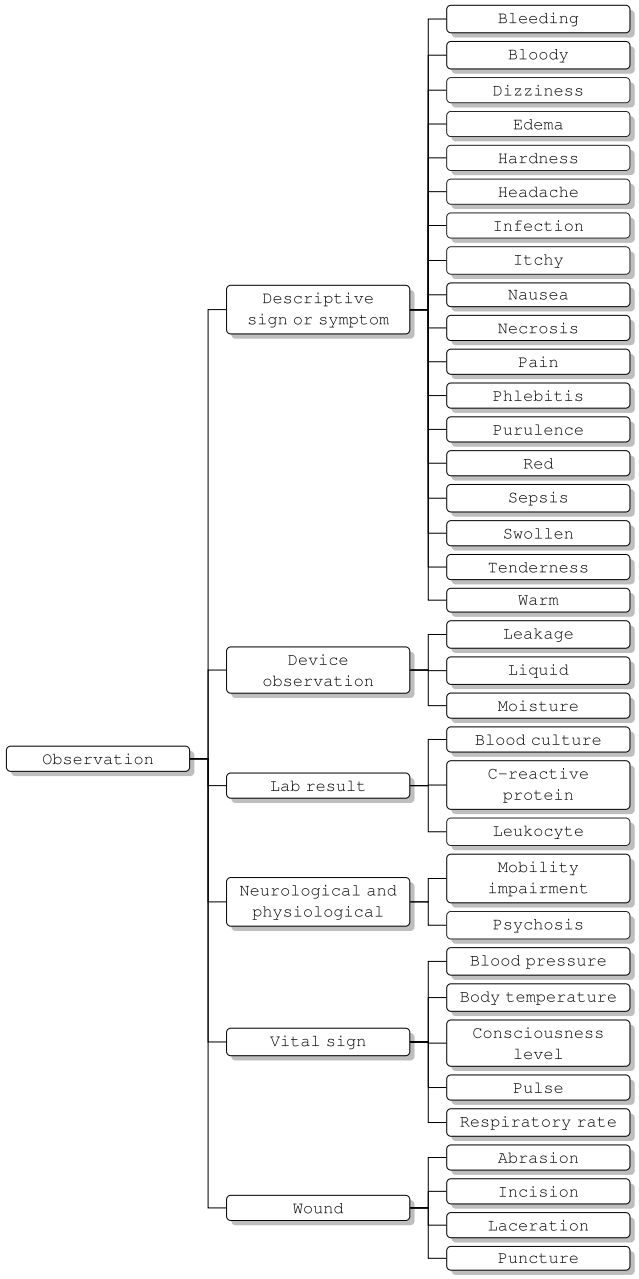


Fig. 8. AAENOTE observation class hierarchy.

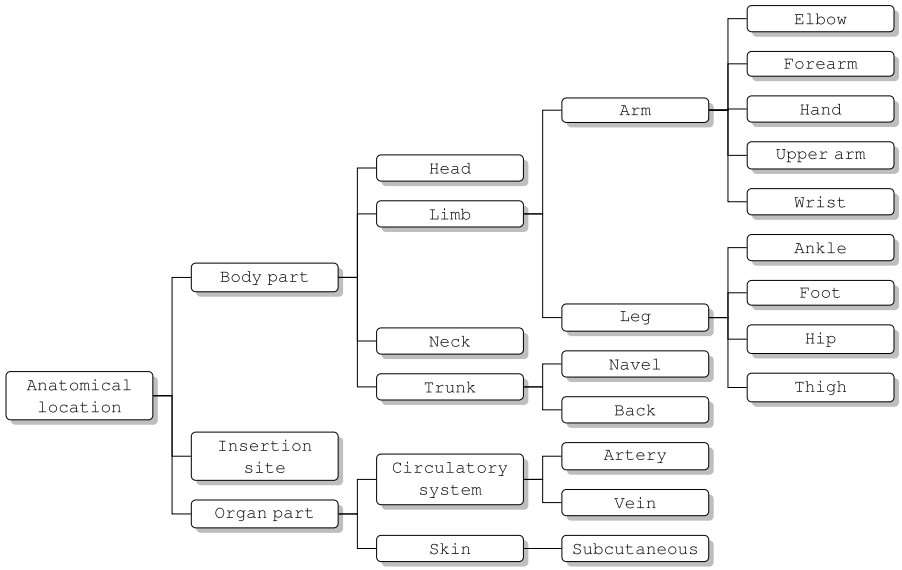


Fig. 9. AAENOTE anatomical location class hierarchy.

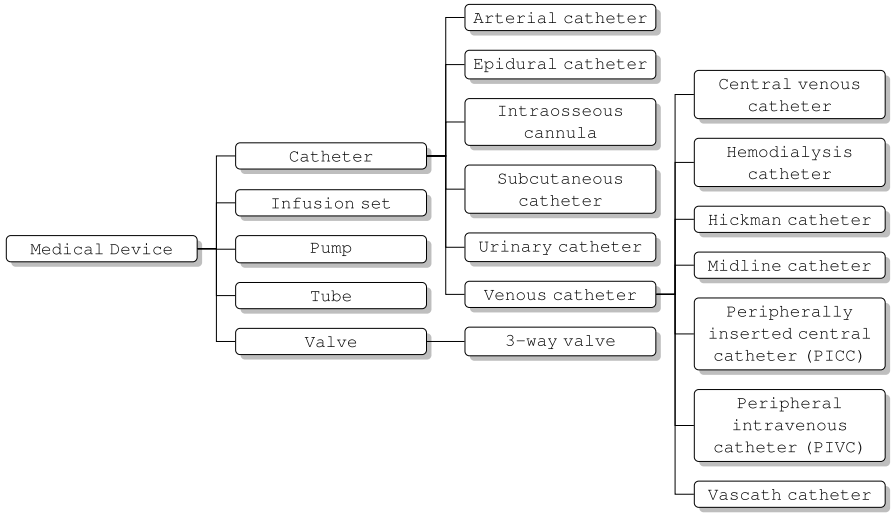


Fig. 10. AAENOTE medical device class hierarchy.

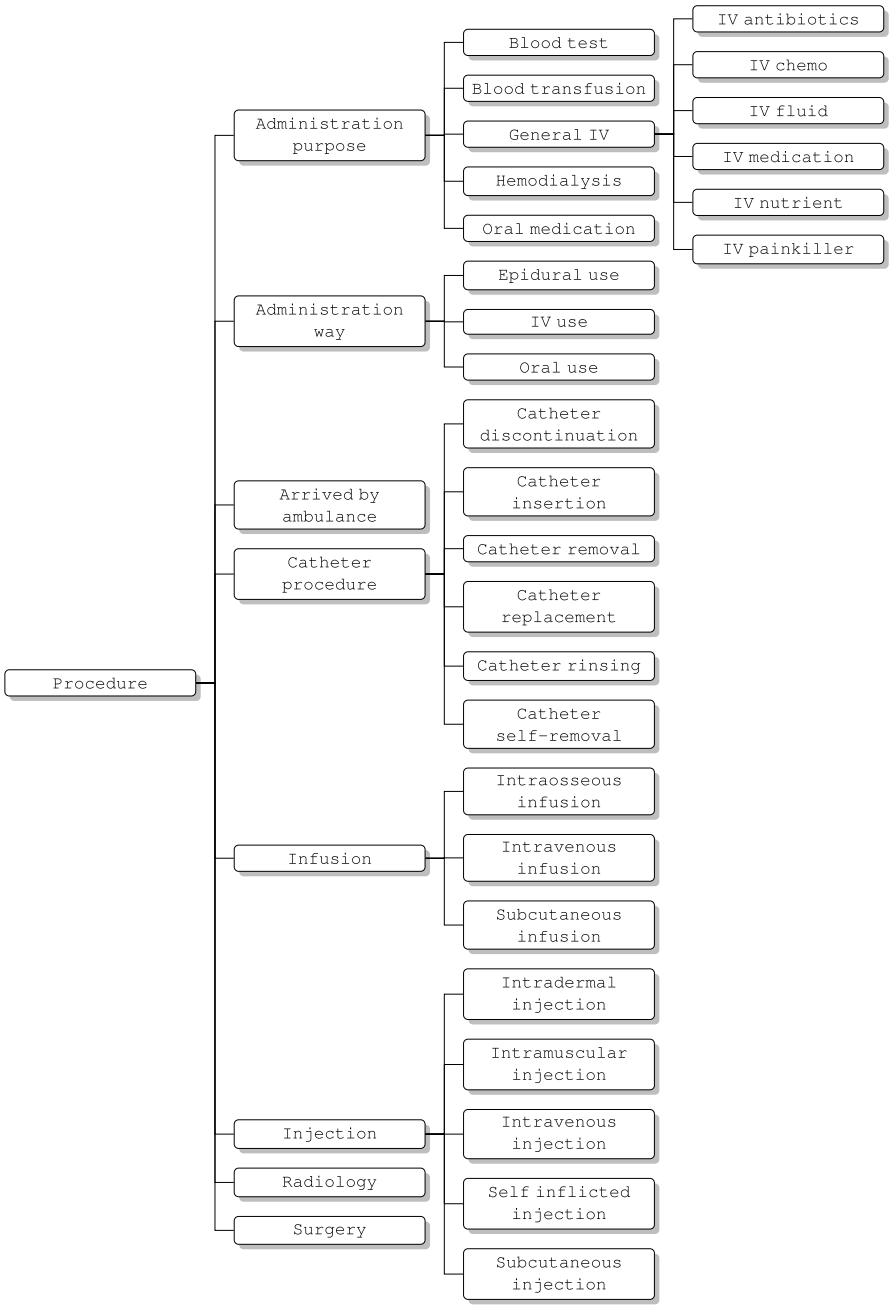


Fig. 11. AAENOTE procedure class hierarchy.

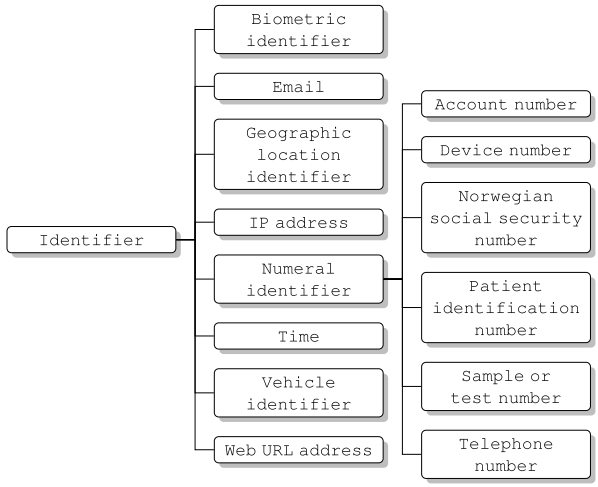


Fig. 12. AAENOTE identifier class hierarchy.

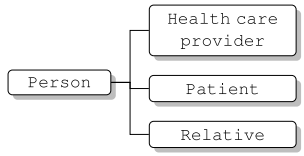


Fig. 13. AAENOTE person class hierarchy.

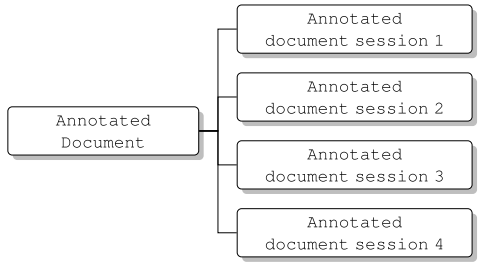


Fig. 14. AAENOTE annotated document class hierarchy.

B.2. Details about converting annotation guideline to terminology

To further differentiate observations, `Descriptive sign or symptom`, `Vital sign`, `Neurological and physiological`, `Wound`, `Lab result`, and `Device observation` subclasses were introduced. In addition, `Insertion site` was introduced because some documents document a catheter's insertion site without mentioning a specific body part. Similarly, `Organ parts`, such as `Skin` and `Circulatory system`, were also introduced because they are documented instead of body part. For `Medical device`, additional catheters were included (i.e., `Intraosseous cannula` and `Subcutaneous catheter`) for specificity. The category for device parts was removed because the terminology only covers catheter and catheter parts; thus, it is a part of a medical device if it is not under a catheter. For `Procedure`, `Administration way` was introduced to describe how a substance was administered into the patient, `Catheter procedure` subclasses all include "catheter" to indicate the action is for catheters only, and different `Infusion` types were included to differentiate from `Injection` types.

A hierarchy for certain data properties was introduced to organize the anatomical location descriptors, contents within a document, and document identity. For instance, document identity-related data properties were included for identification, such as annotator ID, filename, and annotation session. Unlike the data property values for body temperature (i.e., hyperthermia, normal, and hypothermia) and severity level (i.e., high, normal, and low), `Conciousness level` required a separate data property with values of alert, confusion, painfully responsive, unresponsive, and verbally responsive). The "observation is diagnosed by" and "observation said by" data properties were added to distinguish signs from symptoms because signs are what a clinician observes and symptoms are what a patient says.

B.3. AAENOTE competency questions and terminology usage

Table 7
AAENOTE competency questions and terminology usage

Competency Question	Terminology Classes and Relationships to Find Instances*
1. Does patientA have phlebitis, and was it infectious phlebitis, chemical phlebitis, or mechanical phlebitis?	Explicit: `Patient` $\xrightarrow{\text{Person has}}$ `Phlebitis`. Implicit: `Phlebitis` $\xrightarrow{\text{Located nearby/on/at/in}}$ `Anatomical location`. †`Anatomical location` is part of a person, and typically a patient given the context is an AE note. ‡Need clinical knowledge to determine if the `Phlebitis` is infectious phlebitis, chemical phlebitis, or mechanical phlebitis.
2. Does patientA have an infection?	Explicit: `Patient` $\xrightarrow{\text{Person has}}$ `Infection`. `Patient` $\xrightarrow{\text{Person has}}$ `Infection` and other `Observation(s)`. Implicit: `Infection` $\xrightarrow{\text{Located nearby/on/at/in}}$ `Anatomical location`. `Infection` and other `Observation(s)` $\xrightarrow{\text{Located nearby/on/at/in}}$ `Anatomical location`. †`Anatomical location` is part of a person, and typically a patient given the context is an AE note. ‡Need clinical knowledge to determine if the `Observations` combined indicate infection without `Infection` explicitly included.

Table 7
(Continued)

Competency Question	Terminology Classes and Relationships to Find Instances*
3. Does patientA have a BSI?	§ Cannot determine if there is a BSI without a microbiology laboratory result of a positive blood culture and the cultured bacteria name.
4. How many patients have an infection or BSI?	Same as Competency Question 1 and 2.
5. Which patients have sepsis?	Explicit: `Patient` —Person has→ `Sepsis`. `Patient` —Person has→ `Sepsis` and other `Observation(s)`. Implicit: ‡Need clinical knowledge to determine if the `Observations` combined indicate `Sepsis`.
6. Does patientB have a catheter?	Explicit: `Patient` —Person has→ `Catheter`. Implicit: `Catheter` —Located nearby/on/at/in→ `Anatomical location`. †`Anatomical location` is part of a person, and typically a patient given the context is an AE note. ‡Need clinical knowledge to determine if the `Observations`, `Anatomical locations`, and/or `Procedures` combined indicate a `Catheter` is present.
7. Does patientB have a PIVC?	Explicit: `Patient` —Person has→ `PIVC`. Implicit: `PIVC` —Located nearby/on/at/in→ `Anatomical location`. †`Anatomical location` is part of a person, and typically a patient given the context is an AE note. ‡Need clinical knowledge to determine if the `Observations`, `Anatomical locations`, and/or `Procedures` combined indicate a `PIVC` is present.
8a. How many catheters does patientB have?	Same as Competency Question 6.
8b. Where are the catheters in patientB?	Explicit: `Patient` —Person has→ `Anatomical location` (X). `Patient` —Person has→ `Catheter` (Y). `Catheter` (Y) —Located nearby/on/at/in→ `Anatomical location` (X). Implicit: `Catheter` —Located nearby/on/at/in→ `Anatomical location`. †`Anatomical location` is part of a person, and typically a patient given the context is an AE note.

Table 7
(Continued)

Competency Question	Terminology Classes and Relationships to Find Instances*
8c. Why does patientB need the catheter(s)?	Explicit: (a) Patient has a catheter and the catheter is used in the procedure. `Patient` (X) —Person has→ `Catheter` (Y). `Procedure` —Procedure uses→ `Catheter` (Y). (b) Patient has a catheter, patient has a procedure, and that procedure uses that catheter. `Patient` (X) —Person has→ `Catheter` (Y). `Patient` (X) —Person has→ `Procedure` (Z). `Procedure` (Z) —Procedure uses→ `Catheter` (Y). Implicit: (a) Patient has a catheter and patient has a procedure. `Patient` (X) —Person has→ `Catheter`. `Patient` (X) —Person has→ `Procedure`. ‡Need clinical knowledge to determine which `Procedure` is likely to use or involve a specific type of `Catheter`.
9a. Does patientC have an infection and catheter?	Explicit: `Patient` (X) —Person has→ `Infection`. `Patient` (X) —Person has→ `Catheter`. Implicit: ‡Need clinical knowledge to determine if the `Observations` combined indicate `Infection`. ‡Need clinical knowledge to determine if the `Observations`, `Anatomical locations`, and/or `Procedures` combined indicate a `Catheter` is present.
9b. Was patientC's infection associated with a catheter?	§ Cannot determine if an infection is associated with a catheter unless that catheter is tested in the microbiology lab.

B.4. AAENOTE CQ 1: Does patientA have phlebitis, and was it infectious phlebitis, chemical phlebitis, or mechanical phlebitis?

The patients who have phlebitis are listed in Table 8 using Listing 1. And anatomical locations with phlebitis in Table 9 were queried using Listing 2.

B.5. AAENOTE CQ 2: Does patientA have an infection?

B.5.1. AAENOTE CQ 2 explicit

The patients who have infection are listed in Table 10 using Listing 3. Whereas, patients who have infection and/or other observations are listed in Table 11 using Listing 4.

B.5.2. AAENOTE CQ 2 implicit

By using anatomical locations it is possible to implicitly identify infection within a patient because the anatomical locations refer to a place on a human and in the context of AE notes anatomical locations commonly refer to a place

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://www.semanticweb.org/2022/04/aaenote#>
SELECT DISTINCT ?indv_patient ?class_phlebitis
WHERE {
        ?indv_patient rdf:type :patient ;
                rdf:type ?class_patient .
        FILTER(?class_patient != owl:NamedIndividual) .
        ?indv_phlebitis rdf:type* :phlebitis ;
                rdf:type ?class_phlebitis .
        FILTER(?class_phlebitis != owl:NamedIndividual) .
        FILTER(REGEX(str(?class_phlebitis), "phlebitis")) .
        ?indv_patient :person_has ?indv_phlebitis .
}
ORDER BY ?indv_patient
```

Listing 1: Patient has phlebitis

Table 8

AAENOTE SPARQL query result: list patients that have phlebitis

Query Result	Patient	Phlebitis
1	patient.T1.2.lo22a.SD_0003	✓
2	patient.T1.3.do13a.PO_0005	✓
3	patient.T2.2.do13a.SP_0008	✓
4	patient.T2.2.lo12a.SD_0003	✓
5	patient.T2.2.lo22a.PO_0005	✓
6	patient.T2.3.do23a.PO_0005	✓
7	patient.T2.3.po24a.SD_0003	✓
8	patient.T2.4.lo22a.SP_0008	✓
9	patient.T3.3.po14a.SD_0003	✓
10	patient.T4.1.po14a.PO_0005	✓

Query result 1–10: `Patient` $\xrightarrow{\text{Person has}}$ `Phlebitis`.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://www.semanticweb.org/2022/04/aaenote#>
SELECT DISTINCT ?indv_location ?class_phlebitis
WHERE {
        ?indv_phlebitis rdf:type :phlebitis ;
                rdf:type ?class_phlebitis .
        FILTER(?class_phlebitis != owl:NamedIndividual) .
        ?indv_location rdf:type/rdfs:subClassOf* :anatomical_location ;
                rdf:type ?class_location .
        FILTER(?class_location != owl:NamedIndividual) .
        ?indv_phlebitis :located_nearby_on_at_in ?indv_location .
}
ORDER BY ?indv_location
```

Listing 2: List phlebitis located nearby/on/at/in an anatomical location

on a patient. As shown in Table 12, only 1 instance where infection is located at an anatomical location was found using either Listing 5 or Listing 6.

B.6. AAENOTE CQ 3: Does patientA have a BSI?

Cannot determine if there is a BSI without a microbiology laboratory result of a positive blood culture and the cultured bacteria name.

Table 9
AAENOTE SPARQL query result: list phlebitis at an anatomical location

Query Result	Anatomical Location	Phlebitis
1	elbow.T1.4.lo22a.DO_0010	✓
2	elbow.T2.1.do13a.DO_0010	✓
3	elbow.T2.4.lo12a.DO_0010	✓
4	elbow.T3.3.so11a.DO_0010	✓
5	elbow.T3.3.so21a.DO_0010	✓
6	hand.T3.4.lo12a.SP_0008	✓
7	hand.T4.3.do13a.PO_0005	✓
8	hand.T7.2.lo22a.PO_0005	✓
9	wrist.T4.2.do13a.SP_0008	✓
10	patient.T4.1.po14a.PO_0005	✓

Query result 1–10: Listing 2 SPARQL query: `Phlebitis` —Located nearby/on/at/in→ `Anatomical location`.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://www.semanticweb.org/2022/04/aaenote#>
SELECT DISTINCT ?indv_patient ?class_infection
WHERE {
        ?indv_patient rdf:type :patient ;
                rdf:type ?class_patient .
        FILTER(?class_patient != owl:NamedIndividual) .
        ?indv_infection rdf:type* :infection ;
                rdf:type ?class_infection .
        FILTER(?class_infection != owl:NamedIndividual) .
        FILTER(REGEX(str(?class_infection), "infection")) .
        ?indv_patient :person_has ?indv_infection .
}
ORDER BY ?indv_patient
```

Listing 3: List patients who have infection

B.7. AAENOTE CQ 4: How many patients have an infection or BSI?

Explicitly, Table 4 shows the number of patients with infection and the number of patients with infection and/or other observations (Listing 7). Implicitly, there was only 1 anatomical location that had infection (Appendix B.5.2), so the results are not shown for the query in Listing 8. If there was clinical knowledge, it could provide the insight required to determine if combinations of observations at certain anatomical locations are indications of an infection (Listing 9 and Table 13). As previously stated in Appendix B.6, BSI cannot be determined because this requires microbiology laboratory blood test results.

B.8. AAENOTE CQ 5: Which patients have sepsis?

Explicitly, 5 patients have sepsis (Table 14, Listing 10) and 3 of the 5 patients with sepsis have sepsis and another observation (Table 15, Listing 11). Additional clinical knowledge is needed to determine if other observations combined without the sepsis class are indications of sepsis.

B.9. AAENOTE CQ 6: Does patientB have a catheter?

The patient and the type of catheter a patient has can be found explicitly using Listing 12. A subset of the results are in Table 16, where each patient individual is listed with a type of catheter and how many catheters of that specific type are present. Thus, the same patient can be listed multiple times as seen in Table 16's Query Result 28–30 where the same patient is listed 3 times because the patient has 3 catheters of different types. Implicitly, an anatomical

Table 10
AAENOTE SPARQL query result: list patients that have infection

Query Result	Patient	Infection
1	patient.T1.2.lo22a.SD_0006	✓
2	patient.T1.4.lo22a.SP_0002	✓
3	patient.T10.2.do23a.SP_0007	✓
4	patient.T12.4.lo12a.SP_0007	✓
5	patient.T2.2.do23a.SP_0006	✓
6	patient.T2.2.lo12a.SD_0006	✓
7	patient.T2.3.do23a.PO_0010	✓
8	patient.T2.3.po14a.SD_0006	✓
9	patient.T3.2.do13a.SP_0007	✓
10	patient.T4.4.lo22a.DO_0008	✓
11	patient.T4.4.so11a.DP_0010	✓
12	patient.T5.2.po24a.DO_0008	✓
13	patient.T5.4.lo22a.SP_0007	✓
14	patient.T7.3.po24a.SD_0006	✓

Query result 1–14: `Patient` $\xrightarrow{\text{Person has}}$ `Infection`.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://www.semanticweb.org/2022/04/aaenote#>
SELECT DISTINCT ?indv_patient ?observations
WHERE {
        {
                # Group ?class_observations alphabetically by ?indv_patient
                SELECT ?indv_patient ?class_patient
                    (GROUP_CONCAT (DISTINCT str(?class_observation); separator=", " ) as ?observations)
                WHERE {
                        # Organize alphabetically ?class_observations
                        SELECT ?indv_patient ?class_patient ?class_observation
                        WHERE {
                                ?indv_patient rdf:type :patient ;
                                        rdf:type ?class_patient .
                                FILTER(?class_patient != owl:NamedIndividual) .
                                FILTER(REGEX(str(?class_patient), "patient")) .
                                ?indv_observation rdf:type/rdfs:subClassOf* :observation ;
                                        rdf:type ?class_observation .
                                FILTER(?class_observation != owl:NamedIndividual) .
                                ?indv_patient :person_has ?indv_observation .
                        }
                        GROUP BY ?indv_patient ?class_patient ?class_observation
                        ORDER BY ?class_observation
                }
                GROUP BY ?indv_patient ?class_patient
        }
        FILTER REGEX(str(?observations), "infection")
}
ORDER BY ?indv_patient
```

Listing 4: List patients who have infection and/or other observations

location with a catheter indicates a person has a catheter (Listing 13, Table 17). Additional clinical knowledge is needed to determine if other observations combined indicate a catheter is present.

B.10. AAENOTE CQ 7: Does patientB have a PIVC?

This competency question (CQ) can be answered similarily to Appendix B.9. Explicitly using Listing 14 (Table 18) and implicitly using Listing 15 (Table 19). Likewise, additional clinical knowledge is needed to determine if other observations combined indicate a PIVC is present.

Table 11
AAENOTE SPARQL query result: list patients that have infection and/or another observation

	Person	Observation											
Query Result	Patient	blood pressure	body temp	c reactive protein	conciousness level	infection	neurological and physiological	observation	pulse	red	respiratory rate	sepsis	swollen
1	patient.T1.2.lo22a.SD_0006	✓	✓	✓		✓			✓				
2	patient.T1.4.lo22a.SP_0002		✓			✓					✓		
3	patient.T10.2.do23a.SP_0007				✓	✓							
4	patient.T12.4.lo12a.SP_0007				✓	✓		✓					
5	patient.T2.2.do23a.SP_0006					✓							
6	patient.T2.2.lo12a.SD_0006		✓	✓		✓							
7	patient.T2.3.do23a.PO_0010					✓				✓			
8	patient.T2.3.po14a.SD_0006	✓	✓	✓		✓	✓	✓	✓		✓		
9	patient.T3.2.do13a.SP_0007			✓	✓	✓		✓					
10	patient.T4.4.lo22a.DO_0008					✓				✓			✓
11	patient.T4.4.so11a.DP_0010					✓					✓		
12	patient.T5.2.po24a.DO_0008					✓				✓			✓
13	patient.T5.4.lo22a.SP_0007			✓		✓	✓						
14	patient.T7.3.po24a.SD_0006	✓	✓	✓		✓			✓			✓	

Query result 1–14: `Patient` —Person has→ `Infection` and other `Observation(s)`.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://www.semanticweb.org/2022/04/aaenote#>
SELECT DISTINCT ?indv_location ?class_infection
WHERE {
        ?indv_infection rdf:type :infection ;
                rdf:type ?class_infection .
        FILTER(?class_infection != owl:NamedIndividual) .
        ?indv_location rdf:type/rdfs:subClassOf* :anatomical_location ;
                rdf:type ?class_location .
        FILTER(?class_location != owl:NamedIndividual) .
        ?indv_infection :located_nearby_on_at_in ?indv_location .
}
ORDER BY ?indv_location
```

Listing 5: List infection located nearby/on/at/in an anatomical location

B.11. AAENOTE CQ 8a: How many catheters does patientB have?

Listing 16 is used to explicitly query the number and types of catheters a patient has as shown in Table 20. Sometimes there is a direct relationship between a person and multiple catheters such as in Query Result 28, 30, 45, and 52 of Table 20. Whereas, typically there is only one direct relationship between one catheter and an anatomical location when using the implicit query Listing 17 (Table 21). Similarly to Appendix B.7 where with clinical knowledge it could be possible to identify infection indications based on a combination of observations at a certain anatomical location, it could also be possible to determine if combinations of observations at certain anatomical locations are indications of a catheter.

B.12. AAENOTE CQ 8b: Where are the catheters in patientB?

Explicitly Listing 18 and Table 22. Implicitly, it is the same as Appendix B.9's Listing 13 and Table 17 where catheter located at an anatomical location indicates a person has the catheter. Here also, additional clinical knowledge is needed to determine if other observations combined indicate a catheter is present.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://www.semanticweb.org/2022/04/aaenote#>
SELECT DISTINCT ?indv_location ?observations
WHERE {
        {
                SELECT ?indv_location ?class_location
                    (GROUP_CONCAT (DISTINCT ?class_observation; separator=", " ) as ?observations)
                WHERE {
                        {
                                SELECT ?indv_location ?class_location ?class_observation
                                WHERE {
                                        ?indv_observation rdf:type/rdfs:subClassOf* :observation ;
                                                rdf:type ?class_observation .
                                        FILTER(?class_observation != owl:NamedIndividual) .
                                        ?indv_location rdf:type/rdfs:subClassOf* :anatomical_location ;
                                                rdf:type ?class_location .
                                        FILTER(?class_location != owl:NamedIndividual) .
                                        ?indv_observation :located_nearby_on_at_in ?indv_location .
                                }
                                GROUP BY ?indv_location ?class_location ?class_observation
                                ORDER BY ?class_observation
                        }
                }
                GROUP BY ?indv_location ?class_location
                ORDER BY ?class_location
        }
        FILTER REGEX(str(?observations), "infection")
}
ORDER BY ?indv_location
```

Listing 6: List infection and other observations located nearby/on/at/in an anatomical location

Table 12

AAENOTE SPARQL query result: list infections and/or observations at an anatomical location

Query Result	Anatomical Location	Infection
1	skin.T4.2.po14a.SL_0006	✓

Same results for Listing 5 and Listing 6 SPARQL query.

Query result 1 from Listing 5 SPARQL query: `Infection` $\xrightarrow{\text{Located nearby/on/at/in}}$ `Anatomical location`.

Query result 1 from Listing 6 SPARQL query: `Infection` and other `observation(s)` $\xrightarrow{\text{Located nearby/on/at/in}}$ `Anatomical location`.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://www.semanticweb.org/2022/04/aaenote#>
SELECT ?class_patient ?observations ?n_patients
WHERE {
        {
                SELECT ?class_patient
                    (GROUP_CONCAT (DISTINCT str(?class_observation); separator=", " ) as ?observations)
                    (COUNT (DISTINCT ?indv_patient ) as ?n_patients)
                WHERE {
                        ?indv_patient rdf:type :patient ;
                                rdf:type ?class_patient .
                        FILTER(?class_patient != owl:NamedIndividual) .
                        ?indv_infection rdf:type :infection ;
                                rdf:type ?class_observation .
                        FILTER(?class_observation != owl:NamedIndividual) .
                        ?indv_patient :person_has ?indv_infection .
                }
                GROUP BY ?class_patient ?class_observation
                ORDER BY ?class_observation
        }
        UNION
        {

                SELECT ?class_patient ?observations (COUNT (DISTINCT ?indv_patient ) as ?n_patients)
                WHERE {
                        {
                                SELECT ?indv_patient ?class_patient
                                    (GROUP_CONCAT (
                                        DISTINCT str(?class_observation); separator=", " ) as ?observations)
                                WHERE {
                                        SELECT ?indv_patient ?class_patient ?class_observation
                                        WHERE {
                                                ?indv_patient rdf:type :patient ;
                                                        rdf:type ?class_patient .
                                                FILTER(?class_patient != owl:NamedIndividual) .
                                                ?indv_observation rdf:type/rdfs:subClassOf* :observation ;
                                                        rdf:type ?class_observation .
                                                FILTER(?class_observation != owl:NamedIndividual) .
                                                ?indv_patient :person_has ?indv_observation .
                                        }
                                        GROUP BY ?indv_patient ?class_patient ?class_observation
                                        ORDER BY ?class_observation
                                }
                                GROUP BY ?indv_patient ?class_patient
                        }
                        FILTER REGEX(str(?observations), "infection")
                }
                GROUP BY ?class_patient ?observations
        }
}
```

Listing 7: Count the number of patients with infection and number patients with infection and other observations

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://www.semanticweb.org/2022/04/aaenote#>
SELECT ?class_location ?observations (COUNT (DISTINCT ?indv_location) as ?n_instances)
WHERE {
        {
                SELECT ?indv_location ?class_location
                    (GROUP_CONCAT (DISTINCT ?class_observation; separator=", " ) as ?observations)
                WHERE {
                        {
                                SELECT ?indv_location ?class_location ?class_observation
                                WHERE {
                                        ?indv_observation rdf:type/rdfs:subClassOf* :observation ;
                                                rdf:type ?class_observation .
                                        FILTER(?class_observation != owl:NamedIndividual) .
                                        ?indv_location rdf:type/rdfs:subClassOf* :anatomical_location ;
                                                rdf:type ?class_location .
                                        FILTER(?class_location != owl:NamedIndividual) .
                                        ?indv_observation :located_nearby_on_at_in ?indv_location .
                                }
                                GROUP BY ?indv_location ?class_location ?class_observation
                                ORDER BY ?class_observation
                        }
                }
                GROUP BY ?indv_location ?class_location
                ORDER BY ?class_location
        }
        FILTER REGEX(str(?observations), "infection")
}
GROUP BY ?class_location ?observations
ORDER BY ?class_location
```

Listing 8: Count the number of anatomical locations where infection and other observations are located nearby/on/at/in that anatomical location

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://www.semanticweb.org/2022/04/aaenote#>
SELECT ?class_location ?observations (COUNT (DISTINCT ?indv_location) as ?n_instances)
WHERE {
        {
                SELECT ?indv_location ?class_location
                    (GROUP_CONCAT (DISTINCT ?class_observation; separator=", " ) as ?observations)
                WHERE {
                        {
                                SELECT ?indv_location ?class_location ?class_observation
                                WHERE {
                                        ?indv_observation rdf:type/rdfs:subClassOf* :observation ;
                                                rdf:type ?class_observation .
                                        FILTER(?class_observation != owl:NamedIndividual) .
                                        ?indv_location rdf:type/rdfs:subClassOf* :anatomical_location ;
                                                rdf:type ?class_location .
                                        FILTER(?class_location != owl:NamedIndividual) .
                                        ?indv_observation :located_nearby_on_at_in ?indv_location .
                                }
                                GROUP BY ?indv_location ?class_location ?class_observation
                                ORDER BY ?class_observation
                        }
                }
                GROUP BY ?indv_location ?class_location
                ORDER BY ?class_location
        }
}
GROUP BY ?class_location ?observations
ORDER BY ?class_location
```

Listing 9: Count the number of anatomical locations with the same observations located nearby/on/at/in that anatomical location

Table 13
AAENOTE SPARQL results for anatomical location with combinations of observations

		Anatomical Location				Observation										
Query Result	Instances	arm	elbow	hand	skin	body temp	edema	infection	mobility impairment	pain	phlebitis	purulence	red	swollen	warm	wound
1	12	✓								✓						
2	7	✓												✓		
3	6	✓											✓	✓		
4	4	✓								✓				✓		
5	3	✓								✓			✓	✓		
6	2	✓								✓			✓			
7	2	✓											✓			
8	1	✓							✓	✓						
9	1	✓							✓	✓		✓				
10	1	✓				✓			✓	✓		✓				
11	1	✓										✓				
12	1	✓											✓		✓	
13	1	✓					✓						✓			
14	1	✓											✓	✓	✓	
15	7		✓												✓	
16	7		✓										✓			
17	7		✓										✓	✓	✓	
18	6		✓							✓			✓	✓		
19	5		✓							✓				✓		
20	4		✓								✓					
21	3		✓											✓		
22	2		✓										✓		✓	
23	1		✓							✓						
24	1		✓									✓				
25	4			✓										✓		
26	3			✓							✓					
27	2			✓									✓	✓		✓
28	1			✓						✓						
29	1			✓						✓				✓		
30	1			✓									✓	✓		
31	3				✓					✓						
32	3				✓									✓		
33	2				✓					✓				✓		
34	2				✓					✓			✓	✓		
35	2				✓								✓			
36	2				✓								✓	✓		
37	1				✓										✓	
38	1				✓			✓								
39	1				✓								✓		✓	
40	1				✓		✓						✓	✓		

Query result 1–40: `Observation(s)` $\xrightarrow{\text{Located nearby/on/at/in}}$ `Anatomical location`.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://www.semanticweb.org/2022/04/aaenote#>
SELECT ?indv_patient ?class_sepsis
WHERE {
        ?indv_patient rdf:type :patient .
        ?indv_sepsis rdf:type :sepsis ;
                rdf:type ?class_sepsis .
       FILTER(?class_sepsis != owl:NamedIndividual) .
       ?indv_patient :person_has ?indv_sepsis .
}
ORDER BY ?indv_patient
```

Listing 10: List patients who have sepsis

Table 14

AAENOTE SPARQL query result: list patients that have sepsis

Query Result	Patient	Sepsis
1	patient.T1.2.do13a.SP_0006	✓
2	patient.T1.4.lo12a.SP_0006	✓
3	patient.T6.3.lo22a.DP_0010	✓
4	patient.T7.2.po24a.SL_0004	✓
5	patient.T7.3.po24a.SD_0006	✓

Query result 1–5: `Patient` $\xrightarrow{\text{Person has}}$ `Sepsis`.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://www.semanticweb.org/2022/04/aaenote#>
SELECT ?indv_patient ?observations
WHERE {
        {
                SELECT ?indv_patient ?class_patient
                    (GROUP_CONCAT (DISTINCT str(?class_observation); separator=", " ) as ?observations)
                WHERE {
                        {
                                SELECT ?indv_patient ?class_patient ?class_observation
                                WHERE {
                                        ?indv_patient rdf:type :patient ;
                                                rdf:type ?class_patient .
                                        FILTER(?class_patient != owl:NamedIndividual) .
                                        ?indv_observation rdf:type/rdfs:subClassOf* :observation ;
                                                rdf:type ?class_observation .
                                        FILTER(?class_observation != owl:NamedIndividual) .
                                        ?indv_patient :person_has ?indv_observation .
                                }
                                GROUP BY ?indv_patient ?class_patient ?class_observation
                                ORDER BY ?class_observation
                        }
                }
                GROUP BY ?indv_patient ?class_patient
        }
       FILTER REGEX(str(?observations), "sepsis")
}
ORDER BY ?indv_patient
```

Listing 11: List patients that have sepsis and/or other observations

Table 15
AAENOTE SPARQL query result: list patients that have sepsis and/or another observation

	Person	Observation								
Query Result	Patient	blood pressure	body temp	c reactive protein	infection	pulse	red	respiratory rate	sepsis	swollen
1	patient.T1.2.do13a.SP_0006								✓	
2	patient.T1.4.lo12a.SP_0006								✓	
3	patient.T6.3.lo22a.DP_0010							✓	✓	
4	patient.T7.2.po24a.SL_0004		✓				✓		✓	✓
5	patient.T7.3.po24a.SD_0006	✓	✓	✓	✓	✓			✓	

Query result 1–5: `Patient` $\xrightarrow{\text{Person has}}$ `Sepsis` and other `Observation(s)`.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://www.semanticweb.org/2022/04/aaenote#>
SELECT ?indv_patient ?class_catheter (COUNT (DISTINCT ?indv_catheter ) as ?n_catheters)
WHERE {
        ?indv_patient rdf:type :patient ;
                rdf:type ?class_patient .
        FILTER(?class_patient != owl:NamedIndividual) .
        ?indv_catheter rdf:type/rdfs:subClassOf* :catheter ;
                rdf:type ?class_catheter .
        FILTER(?class_catheter != owl:NamedIndividual) .
        ?indv_patient :person_has ?indv_catheter .
}
GROUP BY ?indv_patient ?class_catheter
ORDER BY ?indv_patient ?class_catheter
```

Listing 12: List patients that have a catheter, the catheter's type, and the number of that catheter type

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://www.semanticweb.org/2022/04/aaenote#>
SELECT ?indv_location ?class_catheter
WHERE {
        ?indv_catheter rdf:type/rdfs:subClassOf* :catheter ;
                rdf:type ?class_catheter .
        FILTER(?class_catheter != owl:NamedIndividual) .
        ?indv_location rdf:type/rdfs:subClassOf* :anatomical_location ;
                rdf:type ?class_location .
        FILTER(?class_location != owl:NamedIndividual) .
        ?indv_catheter :located_nearby_on_at_in ?indv_location .
}
ORDER BY ?indv_location
```

Listing 13: List the anatomical location and the type of catheter located nearby/on/at/in there

Table 16
AAENOTE SPARQL query result: list of patients with a catheter

		Person	Catheter								
Query Result	Number of specific catheters	Patient	catheter	central venous catheter	epidural catheter	hickman catheter	hemodialysis catheter	peripheral intravenous catheter	urinary catheter	vas catheter	venous catheter
1	1	patient.T1.1.lo12a.LO_0007						✓			
2	1	patient.T1.1.lo22a.LO_0007						✓			
3	1	patient.T1.1.lo22a.LO_0009					✓				
4	1	patient.T1.2.do13a.LO_0001						✓			
5	1	patient.T1.2.do13a.SP_0001							✓		
6	1	patient.T1.2.do13a.SP_0005						✓			
7	1	patient.T1.2.do13a.SP_0006						✓			
8	1	patient.T1.2.do13a.SP_0010						✓			
9	1	patient.T1.2.lo12a.PO_0005						✓			
10	1	patient.T1.2.lo12a.SD_0007								✓	
11	1	patient.T1.2.lo22a.SD_0002						✓			
12	1	patient.T1.2.lo22a.SD_0007								✓	
13	1	patient.T1.2.po24a.DO_0001			✓						
				...							
22	1	patient.T1.4.lo12a.SP_0003		✓							
23	1	patient.T1.4.lo22a.DO_0003									✓
24	1	patient.T1.4.lo22a.SP_0005						✓			
25	1	patient.T1.4.lo22a.SP_0006						✓			
26	1	patient.T1.4.lo22a.SP_0010						✓			
27	1	patient.T1.4.po24a.LD_0004				✓					
28	1	patient.T1.4.po24a.LD_0009	✓								
29	1	patient.T1.4.po24a.LD_0009		✓							
30	1	patient.T1.4.po24a.LD_0009			✓						
				...							
93	1	patient.T3.4.so11a.LO_0003									✓
94	1	patient.T3.4.so21a.LO_0004						✓			
95	2	patient.T4.1.lo22a.LO_0003						✓			
				...							
138	1	patient.T9.2.so11a.LD_0003						✓			
139	1	patient.T9.3.po24a.SD_0005						✓			

Query result 1–139: `Patient` $\xrightarrow{\text{Person has}}$ `Catheter`. And provide the specific type of `Catheter`.
Query result 28–30, same patient but different catheters.
Query result 95, the patient has 2 PIVCs.

Table 17
AAENOTE SPARQL query result: list anatomical locations with catheters

	Anatomical Location	Catheter		
Query Result	Individuals under class	catheter	venous catheter	peripheral intravenous catheter
1	arm.T14.3.po24a.SD_0002			✓
2	arm.T2.3.po24a.LO_0002			✓
3	arm.T2.3.po24a.SD_0001			✓
4	arm.T3.3.po24a.LO_0005			✓
5	arm.T5.3.po24a.LO_0004			✓
6	arm.T5.3.po24a.SD_0003			✓
7	arm.T5.3.so11a.DO_0002			✓
8	arm.T5.4.lo12a.DO_0003			✓
9	arm.T5.4.lo22a.DO_0003		✓	
	...			
14	arm.T9.4.lo12a.DO_0003			✓
15	body_part.T8.3.po14a.LO_0004			✓
16	elbow.T2.2.lo12a.SD_0001			✓
	...			
69	hand.T9.2.so21a.LD_0003			✓
70	navel.T5.3.so11a.LP_0010	✓		
71	navel.T5.4.do23a.LP_0010	✓		
72	skin.T11.3.po24a.LO_0001			✓
73	skin.T5.2.do23a.LO_0005			✓
74	subcutaneous.T12.2.do23a.LO_0004			✓
75	wrist.T2.2.do13a.LO_0005			✓
76	wrist.T3.2.do23a.LO_0005			✓
77	wrist.T3.4.so11a.LO_0005			✓
78	wrist.T7.3.po14a.LO_0005			✓

Query result 1–78: `Catheter` $\xrightarrow{\text{Located nearby/on/at/in}}$ `Anatomical location`. And provide the specific `Anatomical location` and type of `Catheter`.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://www.semanticweb.org/2022/04/aaenote#>
SELECT ?indv_patient ?class_pivc (COUNT (DISTINCT ?indv_pivc) as ?n_catheters)
WHERE {
        ?indv_patient rdf:type :patient ;
                rdf:type ?class_patient .
        FILTER(?class_patient != owl:NamedIndividual) .
        ?indv_pivc rdf:type/rdfs:subClassOf* :peripheral_intravenous_catheter ;
                rdf:type ?class_pivc .
        FILTER(?class_pivc != owl:NamedIndividual) .
        ?indv_patient :person_has ?indv_pivc .
}
GROUP BY ?indv_patient ?class_pivc
ORDER BY ?indv_patient
```

Listing 14: List patients that have a PIVC and number of PIVCs

Table 18
AAENOTE SPARQL query result: list patients that have a PIVC

Query Result	Number of PIVCs	Patient	peripheral intravenous catheter
1	1	patient.T1.1.lo12a.LO_0007	✓
2	1	patient.T1.1.lo22a.LO_0007	✓
3	1	patient.T1.2.do13a.LO_0001	✓
4	1	patient.T1.2.do13a.SP_0005	✓
5	1	patient.T1.2.do13a.SP_0006	✓
		...	
31	1	patient.T2.1.lo22a.LO_0001	✓
32	2	patient.T2.1.lo22a.LO_0003	✓
33	1	patient.T2.2.do23a.LO_0001	✓
		...	
91	1	patient.T8.4.lo22a.SP_0010	✓
92	1	patient.T9.2.so11a.LD_0003	✓
93	1	patient.T9.3.po24a.SD_0005	✓

Query result 1–93: `Patient` $\xrightarrow{\text{Person has}}$ `PIVC`.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://www.semanticweb.org/2022/04/aaenote#>
SELECT ?indv_location ?class_pivc
WHERE {
        ?indv_pivc rdf:type/rdfs:subClassOf* :peripheral_intravenous_catheter ;
                rdf:type ?class_pivc .
        FILTER(?class_pivc != owl:NamedIndividual) .
        ?indv_location rdf:type/rdfs:subClassOf* :anatomical_location ;
                rdf:type ?class_location .
        FILTER(?class_location != owl:NamedIndividual) .
        ?indv_pivc :located_nearby_on_at_in ?indv_location .
}
ORDER BY ?indv_location
```

Listing 15: List anatomical locations with a PIVC

Table 19
AAENOTE SPARQL query result: list anatomical locations with a PIVC

	Anatomical Location	Catheter
Query Result	Individuals under class	peripheral intravenous catheter
1	arm.T14.3.po24a.SD_0002	✓
2	arm.T2.3.po24a.LO_0002	✓
3	arm.T2.3.po24a.SD_0001	✓
4	arm.T3.3.po24a.LO_0005	✓
5	arm.T5.3.po24a.LO_0004	✓
	...	
14	body_part.T8.3.po14a.LO_0004	✓
15	elbow.T2.2.lo12a.SD_0001	✓
16	elbow.T2.2.lo22a.SD_0001	✓
	...	
67	hand.T7.2.lo22a.PO_0005	✓
68	hand.T9.2.so21a.LD_0003	✓
69	skin.T11.3.po24a.LO_0001	✓
70	skin.T5.2.do23a.LO_0005	✓
71	subcutaneous.T12.2.do23a.LO_0004	✓
72	wrist.T2.2.do13a.LO_0005	✓
73	wrist.T3.2.do23a.LO_0005	✓
74	wrist.T3.4.so11a.LO_0005	✓
75	wrist.T7.3.po14a.LO_0005	✓

Query result 1–75: `PIVC` $\xrightarrow{\text{Located nearby/on/at/in}}$ `Anatomical location`. And provide the specific `Anatomical location`.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://www.semanticweb.org/2022/04/aaenote#>
# Group ?class_catheters alphabetically by ?indv_patient
SELECT ?indv_patient (GROUP_CONCAT (DISTINCT str(?class_catheter); separator=", " ) as ?catheters)
    (COUNT (DISTINCT ?indv_catheter ) as ?n_catheters)
WHERE {
        {
                # Organize alphabetically ?class_catheter
                SELECT ?indv_patient ?indv_catheter ?class_catheter
                WHERE {
                        ?indv_patient rdf:type :patient ;
                                rdf:type ?class_patient .
                        FILTER(?class_patient != owl:NamedIndividual) .
                        ?indv_catheter rdf:type/rdfs:subClassOf* :catheter ;
                                rdf:type ?class_catheter .
                        FILTER(?class_catheter != owl:NamedIndividual) .
                        ?indv_patient :person_has ?indv_catheter .
                }
                GROUP BY ?indv_patient ?indv_catheter ?class_catheter
                ORDER BY ?class_catheter
        }
}
GROUP BY ?indv_patient
ORDER BY ?indv_patient
```

Listing 16: Count the number of catheters a patient has and provide the types of catheters

Table 20
AAENOTE SPARQL query result: list how many catheters and type of catheter a patient has

		Person	Catheter								
Query Results	Number of catheters	Patient	catheter	central venous catheter	epidural catheter	hickman catheter	hemodialysis catheter	peripheral intravenous catheter	urinary catheter	vas catheter	venous catheter
1	1	patient.T1.1.lo12a.LO_0007						✓			
2	1	patient.T1.1.lo22a.LO_0007						✓			
3	1	patient.T1.1.lo22a.LO_0009					✓				
4	1	patient.T1.2.do13a.LO_0001						✓			
5	1	patient.T1.2.do13a.SP_0001							✓		
6	1	patient.T1.2.do13a.SP_0005						✓			
7	1	patient.T1.2.do13a.SP_0006						✓			
8	1	patient.T1.2.do13a.SP_0010						✓			
9	1	patient.T1.2.lo12a.PO_0005						✓			
10	1	patient.T1.2.lo12a.SD_0007								✓	
11	1	patient.T1.2.lo22a.SD_0002						✓			
12	1	patient.T1.2.lo22a.SD_0007								✓	
13	1	patient.T1.2.po24a.DO_0001			✓						
				…							
26	1	patient.T1.4.lo22a.SP_0010						✓			
27	1	patient.T1.4.po24a.LD_0004						✓			
28	4	patient.T1.4.po24a.LD_0009	✓	✓	✓				✓		
29	1	patient.T1.4.po24a.SO_0009							✓		
30	2	patient.T1.4.so11a.DP_0002						✓	✓		
				…							
45	2	patient.T2.1.lo22a.LO_0003						✓			
46	1	patient.T2.2.do13a.SP_0003		✓							
47	1	patient.T2.2.do23a.LO_0001						✓			
48	1	patient.T2.2.do23a.LO_0007						✓			
49	1	patient.T2.2.do23a.SP_0006						✓			
50	1	patient.T2.2.lo22a.PO_0005						✓			
51	1	patient.T2.2.po14a.DO_0002						✓			
52	3	patient.T2.2.so11a.LD_0009	✓	✓					✓		
53	1	patient.T2.2.so11a.PO_0005						✓			
				…							
70	1	patient.T3.2.do23a.SP_0003		✓							
71	1	patient.T3.2.po14a.DO_0001			✓						
72	1	patient.T3.2.po14a.DO_0003									✓
73	1	patient.T3.2.so11a.LD_0003						✓			
74	1	patient.T3.2.so11a.LD_0004				✓					
75	1	patient.T3.3.do13a.SL_0002						✓			
				…							
131	1	patient.T9.2.so11a.LD_0003						✓			
132	1	patient.T9.3.po24a.SD_0005						✓			

Query result 1–132: `Patient` $\xrightarrow{\text{Person has}}$ `Catheter`. And count how many `Catheter(s)` of a specific type the `Patient` has.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://www.semanticweb.org/2022/04/aaenote#>
# Group ?class_catheters alphabetically by ?indv_location
SELECT ?indv_location (GROUP_CONCAT (DISTINCT str(?class_catheter); separator=", " ) as ?catheters)
    (COUNT (DISTINCT ?indv_catheter ) as ?n_catheters)
WHERE {
        {
                # Organize alphabetically ?class_catheter
                SELECT ?indv_location ?indv_catheter ?class_catheter
                WHERE {
                        ?indv_catheter rdf:type/rdfs:subClassOf* :catheter ;
                                rdf:type ?class_catheter .
                        FILTER(?class_catheter != owl:NamedIndividual) .
                        ?indv_location rdf:type/rdfs:subClassOf* :anatomical_location ;
                                rdf:type ?class_location .
                        FILTER(?class_location != owl:NamedIndividual) .
                        ?indv_catheter :located_nearby_on_at_in ?indv_location .
                }
                GROUP BY ?indv_location ?indv_catheter ?class_catheter
                ORDER BY ?class_catheter
        }
}
GROUP BY ?indv_location
ORDER BY ?indv_location
```

Listing 17: Count the number of catheters located nearby/on/at/in an anatomical location and provide the types of catheters

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://www.semanticweb.org/2022/04/aaenote#>
SELECT ?indv_patient ?indv_catheter ?indv_location
WHERE {
        ?indv_patient rdf:type :patient ;
                rdf:type ?class_patient .
        FILTER(?class_patient != owl:NamedIndividual) .
        ?indv_catheter rdf:type/rdfs:subClassOf* :catheter ;
                rdf:type ?class_catheter .
        FILTER(?class_catheter != owl:NamedIndividual) .
        ?indv_location rdf:type/rdfs:subClassOf* :anatomical_location ;
                rdf:type ?class_location .
        FILTER(?class_location != owl:NamedIndividual) .
        ?indv_patient :person_has ?indv_catheter ;
            :person_has ?indv_location .
        ?indv_catheter :located_nearby_on_at_in ?indv_location .
}
ORDER BY ?indv_patient
```

Listing 18: List patients that have an anatomical location, a catheter, and the patient's catheter is located nearby/on/at/in the patient's anatomical location

Table 21

AAENOTE SPARQL query result: list how many catheters and type of catheter an anatomical location has

		Anatomical Location	Catheter		
Query Results	Number of catheters	Individuals under class	catheter	peripheral intravenous catheter	venous catheter
1	1	arm.T14.3.po24a.SD_0002		✓	
2	1	arm.T2.3.po24a.LO_0002		✓	
3	1	arm.T2.3.po24a.SD_0001		✓	
4	1	arm.T3.3.po24a.LO_0005		✓	
5	1	arm.T5.3.po24a.LO_0004		✓	
6	1	arm.T5.3.po24a.SD_0003		✓	
7	1	arm.T5.3.so11a.DO_0002		✓	
8	1	arm.T5.4.lo12a.DO_0003		✓	
9	1	arm.T5.4.lo22a.DO_0003			✓
		...			
14	1	arm.T9.4.lo12a.DO_0003		✓	
15	1	body_part.T8.3.po14a.LO_0004		✓	
16	1	elbow.T2.2.lo12a.SD_0001		✓	
		...			
67	1	hand.T6.3.do23a.PO_0005		✓	
68	1	hand.T7.2.lo22a.PO_0005		✓	
69	1	hand.T9.2.so21a.LD_0003		✓	
70	1	navel.T5.3.so11a.LP_0010	✓		
71	1	navel.T5.4.do23a.LP_0010	✓		
72	1	skin.T11.3.po24a.LO_0001		✓	
73	1	skin.T5.2.do23a.LO_0005		✓	
74	1	subcutaneous.T12.2.do23a.LO_0004		✓	
75	1	wrist.T2.2.do13a.LO_0005		✓	
76	1	wrist.T3.2.do23a.LO_0005		✓	
77	1	wrist.T3.4.so11a.LO_0005		✓	
78	1	wrist.T7.3.po14a.LO_0005		✓	

Query result 1–78: `Catheter` $\xrightarrow{\text{Located nearby/on/at/in}}$ `Anatomical location`. And count how many `Catheter(s)` of a specific type the specific `Anatomical location` has.

Table 22

AAENOTE SPARQL query result: patient has an anatomical location, patient has a catheter, and that catheter is at the patient's anatomical location

Query Result	Patient	Catheter	Anatomical Location
1	patient.T1.4.lo22a.DO_0003	venous_catheter.T6.4.lo22a.DO_0003	arm.T5.4.lo22a.DO_0003
2	patient.T4.2.lo22a.SD_0003	peripheral_intravenous_catheter.T5.2.lo22a.SD_0003	hand.T6.2.lo22a.SD_0003

Query result 1–2: `Patient` (X) $\xrightarrow{\text{Person has}}$ `Anatomical location` (Y), `Patient` (X) $\xrightarrow{\text{Person has}}$ `Catheter` (Z), and `Catheter` (Z) $\xrightarrow{\text{Located nearby/on/at/in}}$ `Anatomical location` (Y).

B.13. AAENOTE CQ 8c: Why does patientB need the catheter(s)?

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://www.semanticweb.org/2022/04/aaenote#>
SELECT ?indv_patient ?class_catheter
    (GROUP_CONCAT (DISTINCT str(?class_procedure); separator=", " ) as ?procedures)
WHERE {
        {
                SELECT ?indv_patient ?class_catheter ?class_procedure
                WHERE {
                        ?indv_patient rdf:type :patient ;
                                rdf:type ?class_patient .
                        FILTER(?class_patient != owl:NamedIndividual) .
                        ?indv_catheter rdf:type/rdfs:subClassOf* :catheter ;
                                rdf:type ?class_catheter .
                        FILTER(?class_catheter != owl:NamedIndividual) .
                        ?indv_procedure rdf:type/rdfs:subClassOf* :procedure ;
                                rdf:type ?class_procedure .
                        FILTER(?class_procedure != owl:NamedIndividual) .
                        ?indv_patient :person_has ?indv_catheter .
                        ?indv_procedure :procedure_uses ?indv_catheter .
                }
                GROUP BY ?indv_patient ?class_catheter ?class_procedure
                ORDER BY ?class_procedure
        }
}
GROUP BY ?indv_patient ?class_catheter
ORDER BY ?indv_patient
```

Listing 19: List patients that have a catheter and the procedures which use that catheter

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://www.semanticweb.org/2022/04/aaenote#>
SELECT ?indv_patient ?class_catheter
    (GROUP_CONCAT (DISTINCT str(?class_procedure); separator=", " ) as ?procedures)
WHERE {
        {
                SELECT ?indv_patient ?class_catheter ?class_procedure
                WHERE {
                        ?indv_patient rdf:type :patient ;
                                rdf:type ?class_patient .
                        FILTER(?class_patient != owl:NamedIndividual) .
                        ?indv_catheter rdf:type/rdfs:subClassOf* :catheter ;
                                rdf:type ?class_catheter .
                        FILTER(?class_catheter != owl:NamedIndividual) .
                        ?indv_procedure rdf:type/rdfs:subClassOf* :procedure ;
                                rdf:type ?class_procedure .
                        FILTER(?class_procedure != owl:NamedIndividual) .
                        ?indv_patient :person_has ?indv_catheter ;
                            :person_has ?indv_procedure .
                        ?indv_procedure :procedure_uses ?indv_catheter .
                }
                GROUP BY ?indv_patient ?class_catheter ?class_procedure
                ORDER BY ?class_procedure
        }
}
GROUP BY ?indv_patient ?class_catheter
ORDER BY ?indv_patient
```

Listing 20: List patients that have a catheter, a procedure, and the patient's procedure uses the patient's catheter

Table 23
AAENOTE SPARQL query result: list patients that have a catheter which was used for a procedure

	Person	Catheter					Procedure										
Query Results	Patient	central venous catheter	epidural catheter	peripheral intravenous catheter	venous catheter	urinary catheter	administration purpose	catheter discontinue use	catheter rinse	catheter self removal	general iv	iv antibiotics	iv chemo	iv fluid	iv medication	iv nutrient	subcutaneous injection
1	patient.T1.1.lo12a.LO_0007			✓								✓					
2	patient.T1.1.lo22a.LO_0007			✓								✓		✓			
3	patient.T1.2.do13a.LO_0001			✓								✓					
4	patient.T1.4.lo12a.DO_0003			✓				✓									
5	patient.T1.4.lo12a.SP_0003	✓													✓		
6	patient.T1.4.lo22a.DO_0003				✓			✓									
7	patient.T1.4.lo22a.SP_0010			✓								✓			✓		
8	patient.T1.4.so11a.DP_0002			✓						✓							
9	patient.T1.4.so11a.DP_0002					✓				✓							
10	patient.T1.4.so11a.LO_0001			✓								✓					
11	patient.T11.1.lo22a.LO_0004			✓									✓				
12	patient.T12.4.so11a.LO_0003				✓				✓								
13	patient.T2.2.do23a.LO_0007			✓								✓		✓			
14	patient.T2.3.so11a.LP_0007			✓						✓							
15	patient.T2.3.so21a.LP_0004			✓										✓			
16	patient.T2.4.lo22a.DO_0001		✓				✓										
17	patient.T2.4.so11a.DP_0001			✓												✓	
18	patient.T2.4.so11a.LO_0007			✓								✓					✓
19	patient.T3.2.so11a.LD_0003			✓										✓			
20	patient.T3.4.po24a.SO_0002			✓						✓							
21	patient.T3.4.so11a.LO_0003				✓				✓								
22	patient.T4.3.po14a.SD_0005			✓						✓							
23	patient.T4.3.so11a.LP_0004			✓										✓			
24	patient.T5.3.po24a.LO_0003			✓					✓					✓			
25	patient.T5.4.lo12a.SP_0010			✓								✓					
26	patient.T6.1.lo22a.LO_0007			✓								✓		✓			
27	patient.T6.3.lo22a.DP_0001	✓									✓						
28	patient.T7.2.do23a.LO_0004			✓									✓				
29	patient.T7.3.po24a.SD_0007				✓				✓								
30	patient.T7.4.so11a.LO_0007			✓								✓					✓
31	patient.T8.4.lo22a.SP_0010			✓								✓					
32	patient.T9.3.po24a.SD_0005			✓						✓							

Query result 1–32: `Patient` $\xrightarrow{\text{Person has}}$ `Catheter` (Y) and `Procedure` $\xrightarrow{\text{Procedure uses}}$ `Catheter` (Y).

Table 24

AAENOTE SPARQL query result: list patients that have a catheter, have a procedure, and where the catheter was used for that procedure

	Person	Catheter	Procedure		
Query Result	Patient	peripheral intravenous catheter	iv antibiotics	iv fluid	iv medication
1	patient.T1.1.lo22a.LO_0007	✓	✓	✓	
2	patient.T1.4.lo22a.SP_0010	✓	✓		✓
3	patient.T2.2.do23a.LO_0007	✓	✓	✓	
4	patient.T6.1.lo22a.LO_0007	✓	✓	✓	

Query result 1–14: `Patient` (X) $\xrightarrow{\text{Person has}}$ `Catheter` (Y), `Patient` (X) $\xrightarrow{\text{Person has}}$ `Procedure` (Z), and `Procedure` (Z) $\xrightarrow{\text{Procedure uses}}$ `Catheter` (Y).

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://www.semanticweb.org/2022/04/aaenote#>
SELECT ?indv_patient (GROUP_CONCAT (DISTINCT str(?class_catheter); separator=", " ) as ?catheters)
    (GROUP_CONCAT (DISTINCT str(?class_procedure); separator=", " ) as ?procedures)
WHERE {
        {
                SELECT ?indv_patient ?class_catheter ?class_procedure
                WHERE {
                        ?indv_patient rdf:type :patient ;
                                rdf:type ?class_patient .
                        FILTER(?class_patient != owl:NamedIndividual) .
                        ?indv_catheter rdf:type/rdfs:subClassOf* :catheter ;
                                rdf:type ?class_catheter .
                        FILTER(?class_catheter != owl:NamedIndividual) .
                        ?indv_procedure rdf:type/rdfs:subClassOf* :procedure ;
                                rdf:type ?class_procedure .
                        FILTER(?class_procedure != owl:NamedIndividual) .
                        ?indv_patient :person_has ?indv_catheter ;
                            :person_has ?indv_procedure .
                }
                GROUP BY ?indv_patient ?class_catheter ?class_procedure
                ORDER BY ?class_catheter ?class_procedure
        }
}
GROUP BY ?indv_patient
ORDER BY ?indv_patient
```

Listing 21: List patients that have a catheter and a procedure, and all catheters and all procedures the patient has

Table 25
AAENOTE SPARQL query result: list patients that have a catheter and a procedure

	Person	Catheter					Procedure								
Query Result	Patient	catheter	central venous catheter	hemodialysis catheter	peripheral intravenous catheter	urinary catheter	administration purpose	blood test	general iv	iv antibiotics	iv chemo	iv fluid	iv medication	procedure	surgery
1	patient.T1.1.lo22a.LO_0007				✓					✓					
2	patient.T1.1.lo22a.LO_0009			✓				✓							
3	patient.T1.2.do13a.SP_0001					✓									✓
4	patient.T1.2.do13a.SP_0010				✓					✓				✓	
5	patient.T1.2.lo22a.SD_0002				✓				✓				✓		
6	patient.T1.4.do13a.LP_0004				✓							✓			
7	patient.T1.4.lo22a.SP_0006				✓			✓							
8	patient.T1.4.lo22a.SP_0010				✓			✓		✓			✓		
9	patient.T12.2.lo22a.SD_0002				✓				✓						
10	patient.T15.3.po24a.SD_0002				✓				✓						
11	patient.T2.1.lo22a.LO_0001				✓					✓					
12	patient.T2.2.do13a.SP_0003		✓										✓		
13	patient.T2.2.do23a.LO_0001				✓					✓					
14	patient.T2.2.do23a.LO_0007				✓					✓		✓			
15	patient.T2.2.lo22a.PO_0005				✓										✓
16	patient.T2.3.do23a.PO_0005				✓										✓
17	patient.T3.2.do13a.LO_0007				✓					✓					
18	patient.T3.2.do23a.SP_0003		✓										✓		
19	patient.T3.3.lo22a.SO_0008	✓											✓		
20	patient.T3.3.po14a.SD_0002				✓								✓		
21	patient.T3.4.lo22a.SP_0003		✓										✓		
22	patient.T4.2.po24a.DO_0002				✓		✓								
23	patient.T5.2.do13a.LO_0003				✓							✓			
24	patient.T5.3.po24a.LO_0001				✓					✓					
25	patient.T6.1.lo22a.LO_0007				✓					✓		✓			
26	patient.T6.2.do13a.LO_0004				✓						✓				
27	patient.T6.4.do13a.LP_0002				✓		✓								
28	patient.T6.4.do23a.LP_0004				✓							✓			
29	patient.T8.4.lo22a.SP_0010				✓			✓							

Query result 1–29: `Patient` (X) $\xrightarrow{\text{Person has}}$ `Catheter` and `Patient` (X) $\xrightarrow{\text{Person has}}$ `Procedure`.

B.14. AAENOTE CQ 9a: Does patientC have an infection and catheter?

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://www.semanticweb.org/2022/04/aaenote#>
SELECT ?indv_patient ?class_catheter ?class_infection
WHERE {
        ?indv_patient rdf:type :patient ;
                rdf:type ?class_patient .
        FILTER(?class_patient != owl:NamedIndividual) .
        ?indv_catheter rdf:type/rdfs:subClassOf* :catheter ;
                rdf:type ?class_catheter .
        FILTER(?class_catheter != owl:NamedIndividual) .
        ?indv_infection rdf:type :infection ;
                rdf:type ?class_infection .
        FILTER(?class_infection != owl:NamedIndividual) .
        ?indv_patient :person_has ?indv_catheter ;
                :person_has ?indv_infection .
}
ORDER BY ?indv_patient
```

Listing 22: List patients that have an infection and a catheter

Table 26
AAENOTE SPARQL query result: list patients that have a infection and a catheter

Query Result	Patient	peripheral intravenous catheter	infection
1	patient.T2.2.do23a.SP_0006	✓	✓

Query result 1: `Patient` (X) $\xrightarrow{\text{Person has}}$ `Infection` and `Patient` (X) $\xrightarrow{\text{Person has}}$ `Catheter`.

B.15. AAENOTE CQ 9b: Was patientC's infection associated with a catheter?

Cannot determine if a patient's infection is associated with a catheter unless that catheter is tested in the microbiology lab.

Appendix C. Catheter infection indications ontology (CIIO) hierarchy, assumptions, indications, and competency questions

C.1. CIIO hierarchy

The CIIO classes and their subclasses can be found in the following figures:

1. Fig. 15 `Observation`
2. Fig. 16 `Anatomical location`
3. Fig. 17 `Medical device`
4. Fig. 18 `Procedure`
5. Fig. 19 `Person`
6. Fig. 20 `Document`
7. Fig. 21 `Sentence`

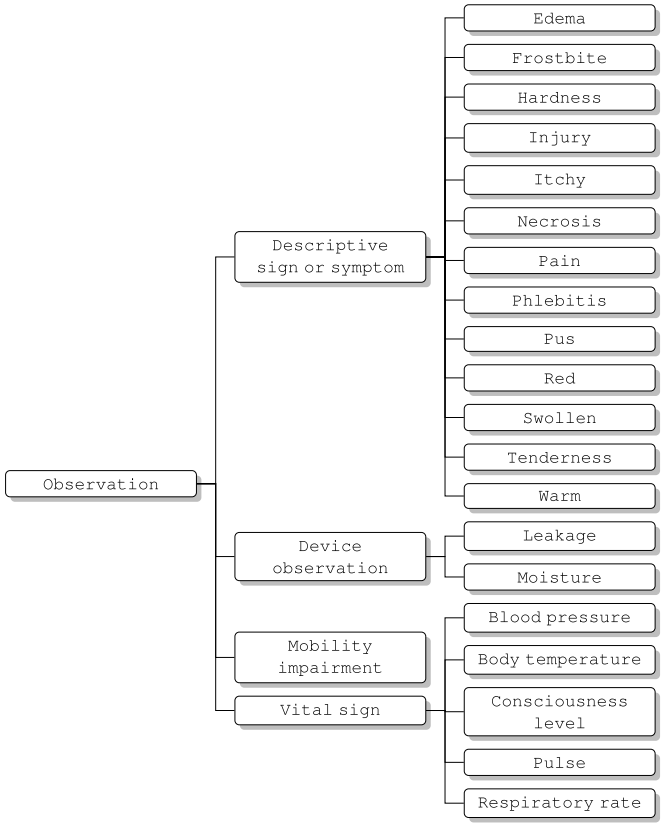


Fig. 15. CIIO observation class hierarchy.

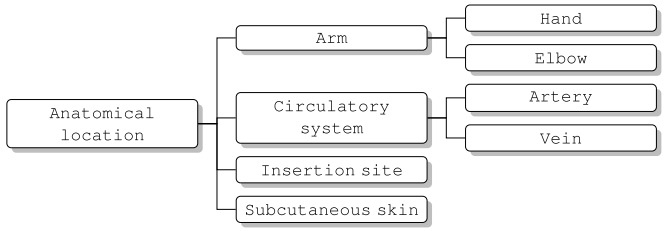


Fig. 16. CIIO anatomical location class hierarchy.

C.2. CIIO assumptions

To address the competency questions (CQs) which ask about patients and not documents and alleviate the problem of patients being implicitly mentioned, 1 AE document represents 1 patient. In the actual electronic incident reporting system database, if the reported AE is related to a patient, there will be a patient ID. This allows users to

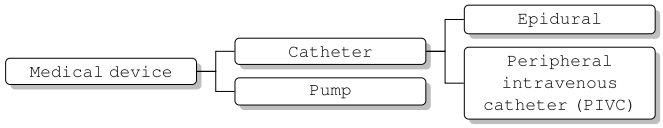


Fig. 17. CIIO medical device class hierarchy.

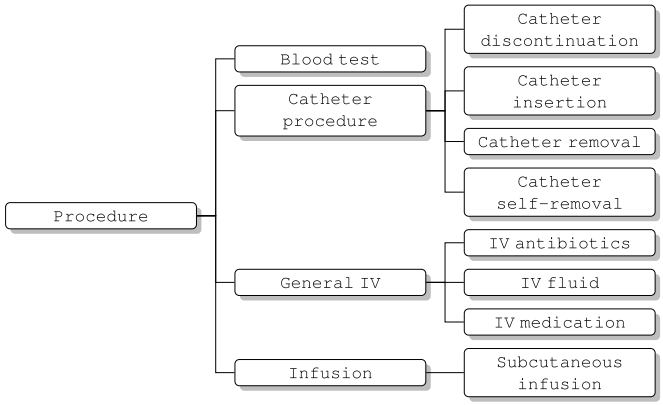


Fig. 18. CIIO procedure class hierarchy.

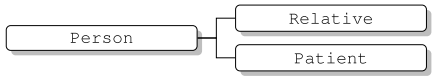


Fig. 19. CIIO person class hierarchy.

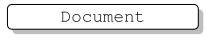


Fig. 20. CIIO document class hierarchy.

Fig. 21. CIIO sentence class hierarchy.

know if the AE is about 1 patient, more than 1 patient, or no patients. Additionally, concepts that are documented within different sentences of the same document are likely describing concepts that occurred in the same event for the same patient. Furthermore, concepts documented in the same sentence that are linked together by a relationship are directly related. Certain relationships can provide the reason for why a concept was or is needed (i.e., procedureA $\xrightarrow{\text{Procedure uses}}$ deviceX, therefore deviceX was needed to perform procedureA).

C.3. Catheter and infection indications

C.3.1. Catheter indications

1. A patient has a specific catheter documented.

2. Any IV usage or infusion indicates some type of catheter is used. IV usage includes general IV, IV medication, IV fluid, and IV antibiotics. Infusion includes infusion, intraosseous infusion, intravenous infusion, and subcutaneous infusion. Based on the type of IV usage or infusion alone, it is not enough to determine what type of catheter was used.
3. Catheter procedures indicate that a catheter is or was present because they require a catheter. Catheter procedures include catheter insertion, catheter discontinued use, catheter removal, catheter replacement, and catheter self-removal.
4. Infusion phlebitis indication indicates that some type of catheter is used.

C.3.2. Peripheral intravenous catheter (PIVC) indications

1. PIVCs are rarely documented, so any PIVC explicitly documented indicates a PIVC was used or in use.
2. Leaking IV or an infusion at the arm, elbow, or hand indicates a PIVC is used. IV includes general IV, IV medication, IV fluid, and IV antibiotics. Infusion includes infusion, intraosseous infusion, intravenous infusion, and subcutaneous infusion. Central venous catheters (CVCs) are deep, so there should not have leakage on the skin. PIVC leakage typically occurs because the catheter dressing is not properly secured or the PIVC is placed near a movable joint (i.e., elbow) and becomes dislodged.

C.3.3. Epidural indication

1. Epidural usage will explicitly be documented.
2. A epidural catheter is the only catheter located nearby/on/at/in the back (i.e., spinal cord).

C.3.4. Infusion phlebitis

As previously stated, infusion phlebitis can be mechanical, chemical, or infectious [23]. Regardless of cause, it is documented similarly and can either be a catheter-related infection or complication.

1. Early stage of infusion phlebitis is indicated by an insertion site or infusion with 2 of the following signs: (i) pain or tenderness, (ii) red, (iii) swollen or edema, or (iv) warm.
2. Medium stage of infusion phlebitis is indicated by (1) a vein with pain and (2) an insertion site or infusion with 2 signs: (i) red, (ii) swollen or edema, or (iii) warm.
3. Advanced stage of infusion phlebitis is indicated by (1) a vein with pain and hardness and (2) an insertion site or infusion with 2 signs: (i) red, (ii) swollen or edema, or (iii) warm.

C.3.5. Infection

1. Pus at an insertion site indicates an infection. Because pus present is a sure sign of infection.

C.3.6. Bloodstream infection (BSI)

1. A bloodstream infection is indicated by a blood test with a positive test result and/or the name of the cultured bacteria.

C.3.7. Sepsis

1. Infection combined with mobility impairment, high body temperature, and frostbite indicates sepsis.
2. Meeting the Quick Sequential Organ Failure Assessment Score (qSOFA) [56] sepsis criteria is indicated if there is an infection indication and at least 2 of the following: (1) high respiratory rate, (2) low blood pressure, or (3) a consciousness level that is either confusion, verbally responsive, painfully responsive, or unresponsive.
3. Sepsis is indicated if there is an infection indication and the National Early Warning Score 2 (NEWS2) [52] criteria for clinical deterioration is met by a combination of (1) high respiratory rate, (2) low blood pressure, (3) high pulse, (4) low body temperature or high body temperature, and (5) consciousness level = confusion, verbally responsive, painfully responsive, or unresponsive.

C.4. CIIO competency questions and ontology usage

SPARQL queries for answering the CIIO competency questions are available on GitHub at https://github.com/melissayan/aaenote_and_ciio/wiki/Ontology-SPARQL-Queries.

Table 27
CIIO competency questions and ontology usage

Competency Question	Ontology Classes and Relationships to Find Instances
1. Does patientA have phlebitis, and was it infectious phlebitis, chemical phlebitis, or mechanical phlebitis?	Early stage infusion phlebitis (a or b): (a) (`Pain` or `Tenderness`), `Red`, (`Swollen` or `Edema`), `Warm` $\xrightarrow{\text{Located nearby/on/at/in}}$ `Injection site`. (b) `Infusion` $\xrightarrow{\text{Is observered with}}$ (`Pain` or `Tenderness`), `Red`, (`Swollen` or `Edema`), `Warm`. Medium stage infusion phlebitis (a) and (b or c): (a) (`Pain` or `Tenderness`) $\xrightarrow{\text{Located nearby/on/at/in}}$ `Vein`. (b) `Red`, (`Swollen` or `Edema`), `Warm` $\xrightarrow{\text{Located nearby/on/at/in}}$ `Injection site`. (c) `Infusion` $\xrightarrow{\text{Is observered with}}$ `Red`, (`Swollen` or `Edema`), `Warm`. Advanced stage infusion phlebitis: (a) `Hardness` and (`Pain` or `Tenderness`) $\xrightarrow{\text{Located nearby/on/at/in}}$ `Vein`. (b) `Red`, (`Swollen` or `Edema`), `Warm` $\xrightarrow{\text{Located nearby/on/at/in}}$ `Injection site`. (c) `Infusion` $\xrightarrow{\text{Is observered with}}$ `Red`, (`Swollen` or `Edema`), `Warm`. ‡Need additional documentation or pathophysiology studies at the body's cellular or genetic-level to determine if the `Phlebitis` is infectious phlebitis, chemical phlebitis, or mechanical phlebitis.
2. Does patientA have an infection?	Pus at an insertion site indicates infection. `Sentence` $\xrightarrow{\text{Contains}}$ `Pus` and `Insertion site`. `Pus` $\xrightarrow{\text{Located nearby/on/at/in}}$ `Insertion site`.
3. Does patientA have a BSI?	Documentation with a blood test with (a) a positive test result and/or (b) name of cultured bacteria. `Sentence` $\xrightarrow{\text{Contains}}$ `Blood test`. (a) `Blood test` $\xRightarrow{\text{Has blood test result}}$ Positive. (b) `Blood test` $\xRightarrow{\text{Has cultured bacteria}}$ bacteria name (i.e., Streptococcus, Staphylococcus, *S. aureus*, etc.).
4. How many patients have an infection or BSI?	Same as Competency Question 1 and 2.

Table 27
(Continued)

Competency Question	Ontology Classes and Relationships to Find Instances
5. Which patients have sepsis?	Sepsis is indicated by (a) an infection indication combined with (b): (1) mobility impairment, (2) high body temperature (i.e., hyperthermia), and (3) frostbite: (a) `Sentence` $\xrightarrow{\text{Contains}}$ `pus` and `Insertion site`. `Pus` $\xrightarrow{\text{Located nearby/on/at/in}}$ `Insertion site`. (b) `Sentence` $\xrightarrow{\text{Contains}}$ `Mobility impairment`, `Body temperature`, and `Frostbite`. `Body temperature` $\xRightarrow{\text{Has body temperature range}}$ hyperthermia. Meeting the Quick Sequential Organ Failure Assessment Score (qSOFA) criteria of (a) an infection indication and (b) at least 2 of the following: (1) high respiratory rate, (2) low blood pressure, or (3) a consciousness level that is either confusion, verbally responsive, painfully responsive, or unresponsive: (a) `Sentence` $\xrightarrow{\text{Contains}}$ `pus` and `Insertion site`. `Pus` $\xrightarrow{\text{Located nearby/on/at/in}}$ `Insertion site`. (b) `Sentence` $\xrightarrow{\text{Contains}}$ `Respiratory rate`, `Blood pressure`, and/or `Consciousness level`. `Respiratory rate` $\xRightarrow{\text{Has severity}}$ high. `Blood pressure` $\xRightarrow{\text{Has severity}}$ low. `Consciousness level` $\xRightarrow{\text{Has consciousness state}}$ confusion, verbally responsive, painfully responsive, or unresponsive. Meeting (a) an infection indication and (b) the National Early Warning Score 2 (NEWS2) criteria for clincial deterioration by having a combination of: (1) high respiratory rate, (2) low blood pressure, (3) high pulse, (4) low body temperature or high body temperature (i.e., hypothermia or hyperthermia), and (5) consciousness level = confusion, verbally responsive, painfully responsive, or unresponsive: (a) `Sentence` $\xrightarrow{\text{Contains}}$ `Pus` and `Insertion site`. `Pus` $\xrightarrow{\text{Located nearby/on/at/in}}$ `Insertion site`. (b) `Sentence` $\xrightarrow{\text{Contains}}$ `Respiratory rate`, `Blood pressure`, `Pulse`, `Body temperature`, and/or `Consciousness level`. `Respiratory rate` $\xRightarrow{\text{Has severity}}$ high. `Blood pressure` $\xRightarrow{\text{Has severity}}$ low. `Pulse` $\xRightarrow{\text{Has severity}}$ high. `Body temperature` $\xRightarrow{\text{Has body temperature range}}$ hypothermia or hyperthermia. `Consciousness level` $\xRightarrow{\text{Has consciousness state}}$ confusion, verbally responsive, painfully responsive, or unresponsive.

Table 27
(Continued)

Competency Question	Ontology Classes and Relationships to Find Instances
6. Does patientB have a catheter?	Documentation of a specific catheter, IV usage, infusion, or catheter-related procedure: `Sentence` $\xrightarrow{\text{Contains}}$ `Medical device(s)` or `IV`-related, `infusion`-related, or `Catheter procedure`-related procedures. Early stage infusion phlebitis is indicated by (a) an injection site or (b) infusion with 2 of the following signs: (i) pain or tenderness, (ii) red, (iii) swollen or edema, or (iv) warm: (a) `Sentence` $\xrightarrow{\text{Contains}}$ `Injection site` and `Pain` or `Tenderness`, `Red`, `Swollen` or `Edema`, and/or `Warm`. `Pain` or `tenderness`, `Red`, `Swollen` or `Edema`, or `Warm` $\xrightarrow{\text{Located nearby/on/at/in}}$ `Injection site`. (b) `Sentence` $\xrightarrow{\text{Contains}}$ `Infusion` and `Pain` or `Tenderness`, `Red`, `Swollen` or `Edema`, and/or `Warm`. `Pain` or `Tenderness`, `Red`, `Swollen` or `Edema`, or `Warm` $\xrightarrow{\text{Is observered with}}$ `Infusion`. Medium stage infusion phlebitis is indicated by (a) a vein with pain or tenderness, and (b) an injection site or (c) infusion with 2 of the following signs: (i) red, (ii) swollen or edema, or (iii) warm: (a) `Sentence` $\xrightarrow{\text{Contains}}$ `Vein` and `Pain` or `Tenderness`. `Pain` or `Tenderness` $\xrightarrow{\text{Located nearby/on/at/in}}$ `Vein`. (b) `Sentence` $\xrightarrow{\text{Contains}}$ `Injection site` and `Red`, `Swollen` or `Edema`, and/or `Warm`. `Red`, `Swollen` or `Edema`, or `Warm` $\xrightarrow{\text{Located nearby/on/at/in}}$ `Injection site`. (c) `Sentence` $\xrightarrow{\text{Contains}}$ `Infusion` and `Pain` or `Tenderness`, `Red`, `Swollen` or `Edema`, and/or `Warm`. `Red`, `Swollen` or `Edema`, or `Warm` $\xrightarrow{\text{Is observered with}}$ `Infusion`. Advanced stage infusion phlebitis is indicated by (a) a vein with (i) hardness and (ii) pain or tenderness, and (b) an injection site or (c) infusion with 2 of the following signs: (i) red, (ii) swollen or edema, or (iii) warm: (a) `Sentence` $\xrightarrow{\text{Contains}}$ `Vein`, `Hardness`, and `Pain` or `Tenderness`. `Hardness` and `Pain` or `Tenderness` $\xrightarrow{\text{Located nearby/on/at/in}}$ `Vein`. (b) `Sentence` $\xrightarrow{\text{Contains}}$ `Injection site` and `Red`, `Swollen` or `Edema`, and/or `Warm`. `Red`, `Swollen` or `Edema`, or `Warm` $\xrightarrow{\text{Located nearby/on/at/in}}$ `Injection site`. (c) `Sentence` $\xrightarrow{\text{Contains}}$ `Infusion` and `Pain` or `Tenderness`, `red`, `Swollen` or `Edema`, and/or `Warm`. `Red`, `Swollen` or `Edema`, or `Warm` $\xrightarrow{\text{Is observered with}}$ `Infusion`.

Table 27
(Continued)

Competency Question	Ontology Classes and Relationships to Find Instances
7. Does patientB have a PIVC?	PIVCs are rarely documented, so any PIVC explicitly documented indicates a PIVC was used or in use: `Sentence` $\xrightarrow{\text{Contains}}$ `PIVC`. Leaking IV or infusion at the arm, elbow or hand indicates PIVC usage: `Sentence` $\xrightarrow{\text{Contains}}$ `Leakage` and `IV`-related or `Infusion`-related procedures. `Leakage` $\xrightarrow{\text{Is observed with}}$ `IV`-related or `Infusion`-related procedures.
8a. How many catheters does patientB have?	Same as Competency Question 5's documentation of a specific catheter, IV usage, infusion, or catheter-related procedure: `Sentence` $\xrightarrow{\text{Contains}}$ `Medical device(s)` or `IV`-related, `Infusion`-related, or `Catheter procedure`-related procedures. * The exact number of catheters per document cannot be counted because multiple sentences within the document could be describing the same catheter and documented procedures can use the same catheter.
8b. Where are the catheters in patientB?	Medical device located nearby/on/at/in an anatomical location: `Sentence` $\xrightarrow{\text{Contains}}$ `Anatomical location` and `Medical device`. `Medical device` $\xrightarrow{\text{Located nearby/on/at/in}}$ `Anatomical location`. IV usage, infusion, or catheter-related procedure located nearby/on/at/in an anatomical location: `Sentence` $\xrightarrow{\text{Contains}}$ `Anatomical location` and `IV`-related, `Infusion`-related, or `Catheter procedure`-related procedures. `IV`-related, `Infusion`-related, or `Catheter procedure`-related procedures $\xrightarrow{\text{Located nearby/on/at/in}}$ `Anatomical location`. * The exact anatomical location of catheters per document cannot be determined because multiple sentences within a document could be describing the same catheter at the same location but with more general terms (i.e., arm instead of hand), the location's position was not documented (e.g., If a sentence contains elbowA, right handB, and right armC, then handB is part of armC, but elbowA might or might not be part of armC), multiple procedures can be performed at the same location, and an additional anatomical ontology is needed to infer the location based on catheter type.

Table 27
(Continued)

Competency Question	Ontology Classes and Relationships to Find Instances
8c. Why does patientB need the catheter(s)?	The medical device is needed and used in a specific procedure. `Sentence` $\xrightarrow{\text{Contains}}$ `Medical device` and `Procedure`. `Procedure` $\xrightarrow{\text{Procedure uses}}$ `Medical device`. Document has sentences with medical devices and/or procedures: `Sentence` $\xrightarrow{\text{Contains}}$ `Medical device` and/or `Procedure`. * The exact reason cannot be determined unless the $\xrightarrow{\text{Procedure uses}}$ object property links `procedure` and `medical device` because the indications do not provide a list of reasons for why a specific catheter can be used. However, a clinician can view the retrieved list of devices and procedures to determine if the devices in a document could be used for the procedures documented.
9a. Does patientC have an infection and catheter?	Document with (a) infection indication and (b or c) catheter indication: (a) `Sentence` $\xrightarrow{\text{Contains}}$ `Pus` and `Insertion site`. `Pus` $\xrightarrow{\text{Located nearby/on/at/in}}$ `Insertion site`. (b) `Sentence` $\xrightarrow{\text{Contains}}$ `Medical device`. (c) `Sentence` $\xrightarrow{\text{Contains}}$ `Anatomical location` and `IV`-related, `Infusion`-related, or `Catheter procedure`-related procedures.
9b. Was patientC's infection associated with a catheter?	§ Cannot determine if an infection is associated with a catheter unless that catheter is tested in the microbiology lab.

References

[1] E. Alexandrou, G. Ray-Barruel, P.J. Carr, S.A. Frost, S. Inwood, N.S. Higgins, F. Lin, L. Alberto, L.A. Mermel and C.M. Rickard, Use of Short Peripheral Intravenous Catheters: Characteristics, Management, and Outcomes Worldwide, *Journal of Hospital Medicine* **13**(5) (2018). doi:10.12788/jhm.3039.

[2] L. Amos, D. Anderson, S. Brody, A. Ripple and B.L. Humphreys, UMLS users and uses: A current overview, *Journal of the American Medical Informatics Association* (2020). doi:10.1093/jamia/ocaa084.

[3] A.R. Aronson, Effective mapping of biomedical text to the UMLS Metathesaurus: The MetaMap program, *Proceedings. AMIA Symposium* (2001), 17–21.

[4] M. Bada, M. Eckert, M. Palmer and L. Hunter, An overview of the CRAFT concept annotation guidelines, in: *Proceedings of the Fourth Linguistic Annotation Workshop, LAW 2010*, Uppsala, Sweden, July 15–16, 2010, Association for Computational Linguistics, 2010, pp. 207–211.

[5] M. Bada and L. Hunter, Desiderata for ontologies to be used in semantic annotation of biomedical documents, *Journal of Biomedical Informatics* **44**(1) (2011), 94–101. doi:10.1016/j.jbi.2010.10.002.

[6] M. Bada, N. Vasilevsky, W.A. Baumgartner, M. Haendel and L.E. Hunter, Gold-standard ontology-based anatomical annotation in the CRAFT Corpus, *Database* **2017** (2017). doi:10.1093/database/bax087.

[7] Biological Spatial Ontology, 2020. http://obofoundry.org/ontology/bspo.

[8] O. Bodenreider, The Unified Medical Language System (UMLS): Integrating biomedical terminology, *Nucleic Acids Research 32(Database issue)* (2004), D267–70. doi:10.1093/nar/gkh061.

[9] H.K. Butcher, G.M. Bulechek, J.M. Dochterman and C.M. Wagner, *Nursing Interventions Classification (NIC)*, Elsevier, 2018. ISBN 9780323497701.

[10] O. Corcho, M. Fernández-López and A. Gómez-Pérez, Methodologies, tools and languages for building ontologies. Where is their meeting point?, *Data & Knowledge Engineering* **46**(1) (2003), 41–64. doi:10.1016/S0169-023X(02)00195-7.

[11] A. Degbelo, A snapshot of ontology evaluation criteria and strategies, in: *Proceedings of the 13th International Conference on Semantic Systems, SEMANTiCS 2017*, Amsterdam, the Netherlands, September 11–14, 2017, ACM, 2017. ISBN 9781450352963. doi:10.1145/3132218.3132219.

[12] C. Dendle, R.L. Stuart and D. Egerton-Warburton, Think before you insert an intravenous catheter, *The Medical journal of Australia* **199**(10) (2013), 655. doi:10.5694/mja13.10733.

[13] eHealth & International Classification for Nursing Practice (ICNP™), 2021. https://www.icn.ch/what-we-do/projects/ehealth-icnptm.
[14] M. Fernández-López, A. Gómez-Pérez and N. Juristo, METHONTOLOGY: From ontological art towards ontological engineering, in: *Proceedings of the Ontological Engineering AAAI-97*, 24–26 March 1997, Spring Symposium Series, AAAI Press, Stanford, CA, USA, 1997.
[15] Foundational Model of Anatomy, 2008. http://si.washington.edu/projects/fma.
[16] D. Garijo, WIDOCO: A wizard for documenting ontologies, in: *Proceedings, Part II*, The Semantic Web – ISWC 2017 – 16th International Semantic Web Conference, Vienna, Austria, October 21–25, 2017, Springer, 2017, pp. 94–102. doi:10.1007/978-3-319-68204-4_9.
[17] A. Goldfain, B. Smith, S. Arabandi, M. Brochhausen and W.R. Hogan, Vital sign ontology, in: *Proceedings of the Workshop on Bio-Ontologies, Intelligent Systems in Molecular Biology (ISMB)*, Vienna, Austria, July 17–19, 2011, 2011.
[18] A. Gómez-Pérez, Towards a framework to verify knowledge sharing technology, *Expert Systems with Applications* **11**(4) (1996), 519–529. doi:10.1016/S0957-4174(96)00067-X.
[19] A. Gómez-Pérez, Ontology evaluation, in: *Handbook on Ontologies*, S. Staab and R. Studer, eds, Springer, Berlin Heidelberg, Berlin, Heidelberg, 2004, pp. 251–273. ISBN 9783540247500. doi:10.1007/978-3-540-24750-0_13.
[20] M. Grüninger and M.S. Fox, Methodology for the design and evaluation of ontologies, in: *Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing, International Joint Conference on Artificial Intelligence (IJCAI-95)*, Montreal, Canada, 1995, ISSN 1045-0823.
[21] Y. He, S. Sarntivijai, Y. Lin, Z. Xiang, A. Guo, S. Zhang, D. Jagannathan, L. Toldo, C. Tao and B. Smith, OAE: The ontology of adverse events, *Journal of Biomedical Semantics* **5** (2014), 29. doi:10.1186/2041-1480-5-29.
[22] T. Heather Herdman, S. Kamitsuru and C.T. Lopes, *NANDA International Nursing Diagnoses: Definitions & Classification, 2021–2023*, Thieme Medical Publishers, Incorporated, 2021. ISBN 9781684204540.
[23] R. Higginson and A. Parry, Phlebitis: Treatment, care and prevention, *Nursing Times* **107**(36) (2011), 18–21.
[24] L.H. Høvik, K.H. Gjeilo, S. Lydersen, C.M. Rickard, B. Røtvold, J.K. Damås, E. Solligård and L.T. Gustad, Monitoring quality of care for peripheral intravenous catheters; feasibility and reliability of the peripheral intravenous catheters mini questionnaire (PIVC-miniQ), *BMC Health Services Research* **19**(1) (2019), 636. doi:10.1186/s12913-019-4497-z.
[25] E. Hovy, M. Marcus, M. Palmer, L. Ramshaw and R. Weischedel, in: *OntoNotes: The 90% Solution, in: Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings*, New York, New York, USA, June 4–9, 2006, Association for Computational Linguistics, 2006, pp. 57–60.
[26] L.E. Huerta and T.W. Rice, Pathologic difference between sepsis and bloodstream infections, *The Journal of Applied Laboratory Medicine* **3**(4) (2019), 654–663. doi:10.1373/jalm.2018.026245.
[27] Infectious Disease Ontology, 2017, https://bioportal.bioontology.org/ontologies/IDO.
[28] Infusion Nurses Society, 2021 Infusion Therapy Standards of Practice Updates, *Journal of Infusion Nursing* **44**(4) (2021), 189–190. doi:10.1097/NAN.0000000000000396.
[29] K.N. Iskander, M.F. Osuchowski, D.J. Stearns-Kurosawa, S. Kurosawa, D. Stepien, C. Valentine and D.G. Remick, Sepsis: Multiple abnormalities, heterogeneous responses, and evolving understanding, *Physiological Reviews* **93**(3) (2013), 1247–1288. doi:10.1152/physrev.00037.2012.
[30] A. Jackson, Infection control – a battle in vein: Infusion phlebitis, *Nursing Times* **94**(4) (1998), 68–71.
[31] J. Kolikof, K. Peterson and A.M. Baker, Central venous catheter, in: *StatPearls, StatPearls Publishing, Treasure*, Island (FL), 2021.
[32] L.A. Mermel, Short-term peripheral venous catheter-related bloodstream infections: A systematic review, *Clinical Infectious Diseases* **65**(10) (2017), 1757–1762. doi:10.1093/cid/cix562.
[33] S.M. Meystre, G.K. Savova, K.C. Kipper-Schuler and J.F. Hurdle, Extracting information from textual documents in the electronic health record: A review of recent research, *Yearbook of Medical Informatics* (2008), 128–144. doi:10.1055/s-0038-1638592.
[34] C.J. Mungall, C. Torniai, G.V. Gkoutos, S.E. Lewis and M.A. Haendel, Uberon, an integrative multi-species anatomy ontology, *Genome Biology* **13**(1) (2012), R5. doi:10.1186/gb-2012-13-1-r5.
[35] National Cancer Institute Thesaurus. https://bioportal.bioontology.org/ontologies/NCIT.
[36] NCI Thesaurus, https://ncithesaurus.nci.nih.gov/ncitbrowser/.
[37] V. Ng, Machine learning for entity coreference resolution: A retrospective look at two decades of research, in: *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017*, ISSN, Vol. 31, AAAI Press, San Francisco, California, USA, 2017, pp. 4877–4884. doi:10.1609/aaai.v31i1.11149.
[38] N.F. Noy and C.D. Hafner, The state of the art in ontology design: A survey and comparative review, *AI Magazine* **18**(3) (1997), 53–53. doi:10.1609/aimag.v18i3.1306.
[39] Ontology for General Medical Science, 2021, http://www.obofoundry.org/ontology/ogms.
[40] Ontology of Adverse Events, 2021. http://www.obofoundry.org/ontology/oae.
[41] J. Paulsen, A. Mehl, Å. Askim, E. Solligård, B.O. Åsvold and J.K. Damås, Epidemiology and outcome of Staphylococcus aureus bloodstream infection and sepsis in a Norwegian county 1996–2011: An observational study, *BMC Infectious Diseases* **15** (2015), 116. doi:10.1186/s12879-015-0849-4.
[42] A. Philpot, E. Hovy and P. Pantel, The omega ontology, in: *Proceedings of OntoLex 2005 – Ontologies and Lexical Resources*, Jeju Island, Korea, October 15, 2005, 2005.
[43] H.S. Pinto and J.P. Martins, Ontologies: How can they be built?, *Knowledge and information systems* **6**(4) (2004), 441–464. doi:10.1007/s10115-003-0138-1.

[44] S. Pradhan, A. Moschitti, N. Xue, O. Uryupina and Y. Zhang, CoNLL-2012 shared task: Modeling multilingual unrestricted coreference in OntoNotes, in: *Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning – Proceedings of the Shared Task: Modeling Multilingual Unrestricted Coreference in OntoNotes, EMNLP-CoNLL 2012*, Jeju Island, Korea, July 13, 2012, Association for Computational Linguistics, 2012, pp. 1–40.
[45] S. Pradhan, L. Ramshaw, M. Marcus, M. Palmer, R. Weischedel and N. Xue, CoNLL-2011 shared task: Modeling unrestricted coreference in OntoNotes, in: *Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task, CoNLL 2011*, Portland, Oregon, USA, June 23–24, 2011, Association for Computational Linguistics, 2011, pp. 1–27.
[46] S.S. Pradhan, E. Hovy, M. Marcus, M. Palmer, L. Ramshaw and R. Weischedel, OntoNotes: A unified relational semantic representation, in: *Proceedings of the First IEEE International Conference on Semantic Computing (ICSC 2007)*, Irvine, California, USA, September 17–19, 2007, 2007, pp. 517–526. doi:10.1109/ICSC.2007.83.
[47] G. Ray-Barruel, D.F. Polit, J.E. Murfield and C.M. Rickard, Infusion phlebitis assessment measures: A systematic review, *Journal of Evaluation in Clinical Practice* **20**(2) (2014), 191–202. doi:10.1111/jep.12107.
[48] Relation Ontology, 2021. http://obofoundry.org/ontology/ro.
[49] J.M. Rohde, A.J. Odden, C. Bonham, L. Kuhn, P.N. Malani, L.M. Chen, S.A. Flanders and T.J. Iwashyna, The epidemiology of acute organ system dysfunction from severe sepsis outside of the intensive care unit, *Journal of Hospital Medicine* **8**(5) (2013), 243–247. doi:10.1002/jhm.2012.
[50] C. Rosse and J.L.V. Mejino, The foundational model of anatomy ontology, in: *Anatomy Ontologies for Bioinformatics: Principles and Practice*, A. Burger, D. Davidson and R. Baldock, eds, Springer, London, London, 2008, pp. 59–117. ISBN 9781846288852. doi:10.1007/978-1-84628-885-2_4.
[51] C. Rosse and J.L.V. Mejino Jr., A reference ontology for biomedical informatics: The foundational model of anatomy, *Journal of Biomedical Informatics* **36**(6) (2003), 478–500. doi:10.1016/j.jbi.2003.11.007.
[52] Royal College of Physicians, National Early Warning Score (NEWS) 2: Standardising the assessment of acute-illness severity in the NHS. Updated report of a working party, Technical Report, Royal College of Physicians, 2017.
[53] K.E. Rudd, N. Kissoon, D. Limmathurotsakul, S. Bory, B. Mutahunga, C.W. Seymour, D.C. Angus and T.E. West, The global burden of sepsis: Barriers and potential solutions, *Critical Care* **22**(1) (2018), 232. doi:10.1186/s13054-018-2157-z.
[54] M. Sabou and M. Fernandez, Ontology (network) evaluation, in: *Ontology Engineering in a Networked World*, M.C. Suárez-Figueroa, A. Gómez-Pérez, E. Motta and A. Gangemi, eds, Springer, Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 193–212. ISBN 9783642247941. doi:10.1007/978-3-642-24794-1_9.
[55] R. Salomão, B.L. Ferreira, M.C. Salomão, S.S. Santos, L.C.P. Azevedo and M.K.C. Brunialti, Sepsis: Evolving concepts and challenges, *Brazilian Journal of Medical and Biological Research* **52**(4) (2019), e8595. doi:10.1590/1414-431X20198595.
[56] C.W. Seymour, V.X. Liu, T.J. Iwashyna, F.M. Brunkhorst, T.D. Rea, A. Scherag, G. Rubenfeld, J.M. Kahn, M. Shankar-Hari, M. Singer, C.S. Deutschman, G.J. Escobar and D.C. Angus, Assessment of clinical criteria for sepsis: For the third international consensus definitions for sepsis and septic shock (sepsis-3), *JAMA* **315**(8) (2016), 762–774. doi:10.1001/jama.2016.0288.
[57] B. Smith, W. Ceusters, B. Klagges, J. Köhler, A. Kumar, J. Lomax, C. Mungall, F. Neuhaus, A.L. Rector and C. Rosse, Relations in biomedical ontologies, *Genome Biology* **6**(5) (2005), R46. doi:10.1186/gb-2005-6-5-r46.
[58] P. Stenetorp, S. Pyysalo, G. Topić, T. Ohta, S. Ananiadou and J. Tsujii, BRAT: A web-based tool for NLP-assisted text annotation, in: *EACL 2012, 13th Conference of the European Chapter of the Association for Computational Linguistics*, Avignon, France, April 23–27, 2012, 2012, pp. 102–107.
[59] M.C. Suárez-Figueroa, A. Gómez-Pérez and M. Fernández-López, The NeOn methodology for ontology engineering, in: *Ontology Engineering in a Networked World*, M.C. Suárez-Figueroa, A. Gómez-Pérez, E. Motta and A. Gangemi, eds, Springer, Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 9–34. ISBN 9783642247941. doi:10.1007/978-3-642-24794-1_2.
[60] Y. Sure, S. Staab and R. Studer, On-to-knowledge methodology (OTKM), in: *Handbook on Ontologies*, S. Staab and R. Studer, eds, Springer, Berlin Heidelberg, Berlin, Heidelberg, 2004, pp. 117–132. ISBN 9783540247500. doi:10.1007/978-3-540-24750-0_6.
[61] R.S. Travillian, T. Adamusiak, T. Burdett, M. Gruenberger, J. Hancock, A.-M. Mallon, J. Malone, P. Schofield and H. Parkinson, Anatomy ontologies and potential users: Bridging the gap, *Journal of Biomedical Semantics* **2**(Suppl 4) (2011), S3. doi:10.1186/2041-1480-2-S4-S3.
[62] M. Uschold and M. King, Towards a methodology for building ontologies, in: *Workshop on Basic Ontological Issues in Knowledge Sharing, Held in Conjuction with IJCAI-95*, Artificial Intelligence Applications Institute, Montreal, Canada, Edinburgh, 1995, pp. 1–13.
[63] M. Villegas, A. Intxaurrondo, A. Gonzalez-Agirre, M. Marimon and M. Krallinger, The MeSpEN resource for English-Spanish medical machine translation and terminologies: Census of parallel corpora, glossaries and term translations, in: *Proceedings of the LREC 2018 Workshop "MultilingualBIO: Multilingual Biomedical Text Processing", European Language Resources Association (ELRA)*, 2018.
[64] Vital Sign Ontology, 2012. https://bioportal.bioontology.org/ontologies/VSO.
[65] H.J. White and M.P. Soos, Anatomy, thorax, superior vena cava, in: *StatPearls, StatPearls Publishing, Treasure*, Island (FL), 2021.
[66] M.Y. Yan, L.H. Høvik, L.T. Gustad and Ø. Nytrø, Understanding and reasoning about early signs of sepsis: From annotation guideline to ontology, in: *IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2021*, Houston, TX, USA, December 9–12, 2021, IEEE, 2021, pp. 1906–1911. doi:10.1109/BIBM52615.2021.9669311.
[67] M.Y. Yan, L.H. Høvik, A. Pedersen, L.T. Gustad and Ø. Nytrø, Preliminary processing and analysis of an adverse event dataset for detecting sepsis-related events, in: *IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2021*, Houston, TX, USA, December 9–12, 2021, IEEE, 2021, pp. 1605–1610. doi:10.1109/BIBM52615.2021.9669410.
[68] L.-C. Yu, C.-H. Wu, A. Philpot and E. Hovy, OntoNotes: Sense pool verification using Google N-Gram and statistical tests, in: *Proceedings of the OntoLex Workshop at the 6th International Semantic Web Conference (ISWC 2007)*, 2007.

[69] L. Zhang, S. Cao, N. Marsh, G. Ray-Barruel, J. Flynn, E. Larsen and C.M. Rickard, Infection risks associated with peripheral vascular catheters, *Journal of Infection Prevention* **17**(5) (2016), 207–213. doi:10.1177/1757177416655472.

ISBN 978-82-326-7518-0 (printed ver.)
ISBN 978-82-326-7517-3 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

NTNU
Norwegian University of
Science and Technology