



Article

Towards the Use of Standardized Terms in Clinical Case Studies for Process Mining in Healthcare [†]

Emmanuel Helm ^{1,2,*} , Anna M. Lin ¹ , David Baumgartner ¹ , Alvin C. Lin ^{3,*} and Josef Küng ²

¹ Research Department Advanced Information Systems and Technology, University of Applied Sciences Upper Austria, 4232 Hagenberg, Austria; anna.lin@fh-hagenberg.at (A.M.L.); david.baumgartner@fh-hagenberg.at (D.B.)

² Institute for Applied Knowledge Processing, Johannes Kepler University, 4040 Linz, Austria; josef.kueng@jku.at

³ Faculty of Medicine, University of Toronto, Toronto, ON M5S 1A8, Canada

* Correspondence: emmanuel.helm@fh-hagenberg.at (E.H.); alvin.lin@mail.utoronto.ca (A.C.L.)

[†] This paper is an extended version of our paper published in International Workshop on Process-Oriented Data Science for Healthcare 2019 (PODS4H19), Vienna, Austria. September 2, 2019 called “Adopting Standard Clinical Descriptors for Process Mining Case Studies in Healthcare”.

Received: 5 January 2020; Accepted: 14 February 2020; Published: 19 February 2020



Abstract: Process mining can provide greater insight into medical treatment processes and organizational processes in healthcare. To enhance comparability between processes, the quality of the labelled-data is essential. A literature review of the clinical case studies by Rojas et al. in 2016 identified several common aspects for comparison, which include methodologies, algorithms or techniques, medical fields, and healthcare specialty. However, clinical aspects are not reported in a uniform way and do not follow a standard clinical coding scheme. Further, technical aspects such as details of the event log data are not always described. In this paper, we identified 38 clinically-relevant case studies of process mining in healthcare published from 2016 to 2018 that described the tools, algorithms and techniques utilized, and details on the event log data. We then correlated the clinical aspects of patient encounter environment, clinical specialty and medical diagnoses using the standard clinical coding schemes SNOMED CT and ICD-10. The potential outcomes of adopting a standard approach for describing event log data and classifying medical terminology using standard clinical coding schemes are further discussed. A checklist template for the reporting of case studies is provided in the Appendix A to the article.

Keywords: process mining; healthcare; terminology; ICD; SNOMED

1. Introduction

Process mining is a discipline that allows for greater understanding into real-life processes of recorded systems behaviour. Through process mining techniques, numerous case studies and successful companies have demonstrated further quality improvement, compliance, and optimization of processes.

In healthcare, recent review papers have provided an overview of process mining across clinical case studies. Rojas et al in 2016 identified eleven common aspects across 74 clinical case studies [1]. These aspects include methodologies, techniques or algorithms, medical fields and healthcare specialty. In 2018, Erdogan and Tarhan conducted a systematic mapping of 172 case studies with mostly the same metrics and aspects [2]. These papers are very specific as to *how* these case studies were conducted, which enhances comparison between different process mining techniques in different settings. However, from a medical perspective, the terms and categories listed under *medical fields*

and *healthcare specialty* are not structured in a uniform way, and do not follow a standardized clinical coding scheme. Further, basic characteristics of the event log data (timeframe, number of cases or patients, healthcare facility/organization) are not always clearly reported.

The number of case studies on process mining in healthcare continues to increase steadily. As such, a standard approach of reporting event log data, clinical specialties and medical diagnoses would provide greater clarity and enhance comparability between treatments of specific diseases across different healthcare settings.

In this article, further to the studies examined by Rojas et al., we conducted a forward search of processing mining case studies in healthcare for the three-year period from January 2016 to December 2018. We identified case studies that described basic characteristics of the event log data, and where information on the patient encounter environment, clinical specialty and medical diagnoses could be assigned under a standard clinical coding scheme. Section 2 describes how the forward search was conducted and which criteria we applied to filter the results. In addition, the methods describe standard clinical coding systems and terminologies that were used. In Section 3, the results of our analysis are presented. Section 4 discusses the benefits and gives an outlook on the potential clinical insights gained by reporting and classifying clinical terms, clinical specialties and medical diagnoses using a standard clinical coding scheme.

This article is an extension to the paper presented at the workshop on Process-Oriented Data Science for Healthcare 2019 (PODS4H19), held in conjunction with the BPM2019 conference in Vienna, Austria [3]. It presents further details to the results of our analysis and it provides an outline for a reporting template in Appendix A. This template can be used as a checklist for the reporting of case studies on process mining in healthcare.

2. Materials and Methods

Our paper focused on answering three questions: (1) Which clinically-relevant case studies of process mining in healthcare will be selected for this study? (2) What were the technical aspects identified? (3) How can we improve the clarity and comparability of the clinical terms and aspects described?

2.1. Selection of Clinically-Relevant Case Studies

Our starting point was the review paper by Rojas et al. [1] which identified 74 case studies where process mining tools, techniques or algorithms were applied in the healthcare domain. We then performed a forward search using Google Scholar, in reference to the 74 identified articles and the review paper itself. The inclusion criteria (IC) were applied at once in the Google Scholar search and the exclusion criteria (EC) were applied manually afterwards (see Figure 1).

- IC1: All articles that reference either the review paper by Rojas et al. [1] or any of the 74 articles identified in their review were included.
- IC2: All articles published between 01.01.2016 and 31.12.2018 were included.
- IC3: All articles published in English were included.
- EC1: Articles that do not include evidence of a clinically-relevant case study of process mining in healthcare were excluded.
- EC2: Articles that present a case study based on data that was already used for an earlier case study were excluded.
- EC3: Articles that do not describe the characteristics of the event log data (e.g., timeframe, number of cases or patients, healthcare facility) or do not describe which process mining technique or algorithm was applied were excluded.
- EC4: Articles that did not describe any clinical context (i.e., clinical specialty or medical diagnosis) were excluded.

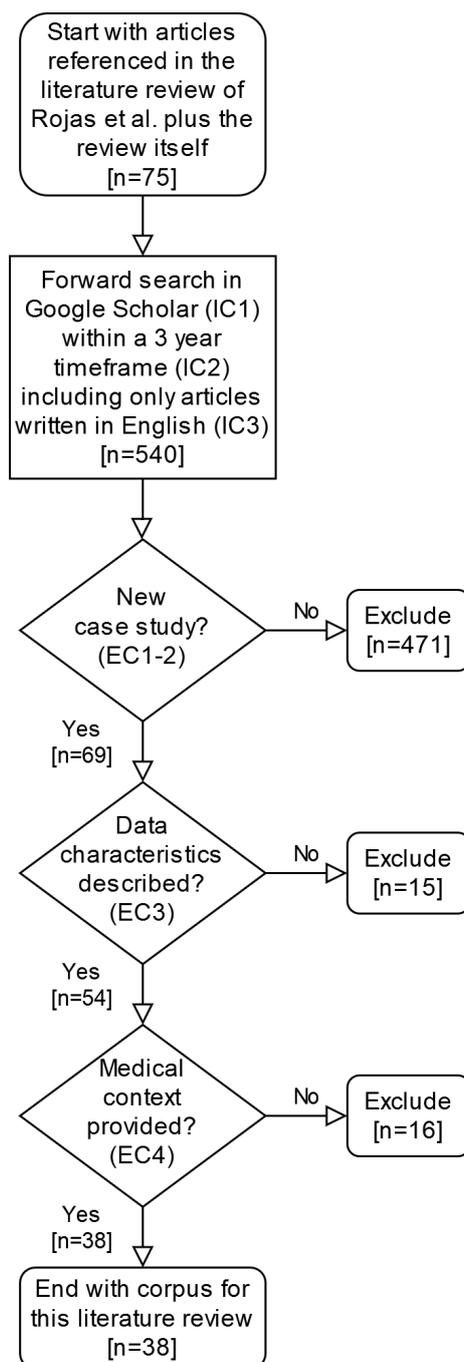


Figure 1. Flowchart on the case study selection strategy.

2.2. Technical Aspects

A detailed account of the tools, techniques or algorithms used in process mining case studies in healthcare have been previously described [1]. Also, other technical descriptors such as the data type and geographical analysis have been used to describe the event log data [4]. For the technical scope of our paper, our focus was on (1) the *tools* used in the case studies, (2) the *techniques or algorithms* used, and (3) the *process mining perspectives*.

2.3. Clinical Aspects and Standard Coding Schemes

Medical language is full of homonyms, synonyms, eponyms, acronyms and abbreviations; and each healthcare specialty comes with their own sub-terminology [5]. To improve the clarity and

comparability of the clinical aspects described in our selected papers, we adopted the use of standard clinical coding schemes of SNOMED CT and ICD-10. Namely, the clinical terms were matched to their best corresponding standard clinical descriptor, with respect to three clinical categories: (1) the type of patient encounter environment (2) clinical speciality and (3) medical diagnosis (i.e., disease or health problem).

2.3.1. SNOMED CT

The Systematized Nomenclature of Medicine – Clinical Terms is an internationally recognized standard that classifies clinically-relevant terminology and concepts, along with their synonyms and relationships, into numeric coded values. Available in multiple languages and maintained by SNOMED International, there are currently over 340,000 numerically coded concepts that can be combined grammatically to create an expression. We used SNOMED CT international browser (<https://browser.ihtsdotools.org/>) in version v20190131 for clinical descriptors on the *Patient encounter environment* and *Clinical speciality*.

2.3.2. ICD-10

For classification of clinical diagnoses and health problems, the commonly accepted system is the International Classification of Diseases or ICD, which is maintained by the World Health Organization (WHO). The most current version is ICD-10 and it utilizes an alphanumeric coding scheme with more than 14,000 single clinical codes of medical terms organized hierarchically into 22 chapters. We used the WHO ICD-10 browser in the 2016 version (<https://icd.who.int/browse10/2016/en>) for clinical descriptors on *medical diagnoses*.

3. Results

3.1. Selection of Clinically-Relevant Case Studies

Our forward search yielded initially a total of 540 papers, and after our inclusion and exclusion criteria were applied, 38 articles were selected (cf. Figure 1). For all 38 papers, basic characteristics of the event log data were retrieved (e.g., origin of data, number of cases or patients, healthcare facility, timeframe of the study). The results of the technical and clinical aspects are described below.

3.2. Technical Aspects

3.2.1. Tools

Table 1 summarizes our findings of the most commonly used Tools to enable process mining techniques and algorithms. ProM (<https://www.promtools.org>) was the most frequent, found in 18 of the selected case studies, and was also found of the same in [1]. Nowadays Disco (<https://fluxicon.com/disco>) is becoming more prevalent, and we found 11 cases as well. To complete our analysis, PALIA was used twice; and in both cases, this tool was used in combination with another tool or technique.

There are a wide variety of other Tools used, often together with ProM, which are listed together in the table for a total of 13 papers. Six case studies introduced self-developed tools.

Table 1. Papers with their corresponding Tools most commonly used (non-disjoint). In some papers, other tools were developed or used, resulting in two categories of *Others* and *Self-developed*.

Tool	ProM	Disco	PALIA	pMineR	Others	Self-Developed
Papers	[6–23]	[4,7–9,18,24–29]	[30,31]	[32]	[6,12,19,25,29–31,33–38]	[16,27,33,39–41]

3.2.2. Techniques or Algorithms

Table 2 describes the four most used techniques and algorithms amongst the selected case studies. Our analysis revealed that *Fuzzy miner* (as implemented in Disco) was most frequently used, appearing in 11 of the case studies. Of note, several papers that utilized ProM also presented self-developed approaches that were case-specific based on the ProM environment. Further, the *Inductive visual miner* is one of the more recent built-in miners in ProM, and is now more frequently used and reported as such. Five case studies used the Trace Clustering technique. Other types like BPMN, ANOVA and machine learning were sometimes used but not on a frequent basis. While the Heuristic miner algorithm was frequented as per previous reviews, [1,2], it was only used in two of our 38 selected papers.

Table 2. Papers with their corresponding Techniques or Algorithms that were mainly used.

Techniques/Algorithms	Fuzzy Miner	Self-Developed	Clustering	Heuristic Miner
Papers	[4,7–9,18,24–29]	[10,11,13,15,16,27,37,39–42]	[6,26,32–34]	[10,21]

3.2.3. Process Mining Perspectives

Our analysis showed that the majority of the case studies (30 in the total) mainly aimed for the *Control Flow* perspective in their dataset (see Table 3). Of those remaining, five papers analyzed the *Conformance* perspective, two focused on *Organizational*, and one on *Performance*.

Table 3. Papers with their corresponding process mining perspectives.

Perspectives	Control Flow	Conformance	Organizational	Performance
Papers	[4,6,7,10–17,19,20,22,23,25–30,33–39,41,42]	[9,18,21,32,40]	[24,31]	[8]

3.3. Clinical Aspects using Standard Clinical Descriptors

3.3.1. Encounter Environment

From the patient's perspective, we considered five clinical settings or encounter environments: (1) Inpatient, (2) Outpatient, (3) Accident and Emergency department or AED, (4) General practitioner or GP practice site, and (5) Pharmacy. All five encounter environments could be coded by SNOMED CT. For each paper, at least one of these five encounter environments was retrieved. Most of the papers examined events within the Inpatient environment, followed by AED environment (cf. Table 4).

Table 4. Papers with their corresponding SNOMED CT encounter environment.

SNOMED CT	Environment	Papers
440654001	Inpatient	[4,6,8,11–16,18–23,25–29,33–42]
440655000	Outpatient	[7,9,17,32,38]
225728007	AED	[10,11,13,17,19,22,24,27,28,30,35,36]
394761003	GP practice site	[31]
264372000	Pharmacy	[4]

3.3.2. Clinical Specialty

SNOMED CT offers the code of 394658006 for *Clinical specialty*, which further contains 18 high-level specialties. Table 5 shows 11 of the 18 high-level clinical specialties were identified in our selected papers. The most identified clinical specialty was *Medical specialty*, followed by *Surgical specialty* and *Emergency medicine*. Of note, some of 18 high-level specialties in SNOMED CT are further divided into sub-specialties of greater clinical specificity. For example, *Medical specialty* has 44 sub-specialties that include e.g., *Dermatology*, *Neurology* and *Cardiology*. In this paper, we identified and assigned sub-specialties to their corresponding high-level *Clinical specialty*. Also, for example, if

several different medical sub-specialities were described in one paper, we counted these sub-specialities together as *Medical specialty*.

Table 5. Papers with their corresponding SNOMED CT clinical specialty.

SNOMED CT	Clinical Specialty	Papers
394592004	Clinical oncology	[14,37,38]
394581000	Community medicine	[31]
722163006	Dentistry	[7,9]
722164000	Dietetics and nutrition	[31]
773568002	Emergency medicine	[10–13,17,19,22,24,28,30,35,36]
394814009	General practice	[7,9,31]
408446006	Gynecological oncology	[41]
394733009	Medical specialty	[4,6,9,12,15,16,18,21,22,25,29,32–35,39,40,42]
722165004	Nursing	[9,24,31]
394585009	Obstetrics and gynecology	[16,41]
394732004	Surgical specialty	[8,9,11,12,14,20,23,26,32,37,41]

3.3.3. Medical Diagnosis

For each paper, we focused on identifying the medical diagnosis (i.e., disease or health problem) or description of a medical diagnosis. We then assigned these terms to their corresponding highest chapter or block category in ICD-10. Table 6 shows a total of 15 out of the 22 ICD-10 chapter categories for disease and health related problems were covered amongst the papers. The category with the most papers listed was *Diseases of the circulatory system* followed by *Neoplasms*. Two papers [16,21] were not included in Table 6, since several hundred diseases and health problems were cited and classified using ICD-9. Of the remaining 36 case studies, ICD-10 was already used in 8 papers to code the diagnosis [12,14,21,22,33,34,38,40].

Table 6. Papers with their corresponding ICD-10 medical diagnosis.

ICD-10	Diagnosis	Papers
A00 - B99	Certain Infectious and parasitic diseases	[4,12,13,19,22,36]
C00 - D48	Neoplasms	[12,14,18,20,25,37,38,41]
E00 - E90	Endocrine, nutritional and metabolic diseases	[9,12,15,31,32,42]
F00 - F99	Mental and behavioural disorders	[12,40]
G00 - G99	Diseases of the nervous system	[12]
H60 - H95	Diseases of the ear and mastoid process	[12]
I00 - I99	Diseases of the circulatory system	[6,8,12,15,25,30,33–35,39,42]
J00 - J99	Diseases of the respiratory system	[12,24]
K00 - K93	Diseases of the digestive system	[7,12,24,28]
M00 - M99	Diseases of the musculoskeletal system and connective tissue	[12,24,40]
N00 - N99	Diseases of the genitourinary system	[12]
000 - 099	Pregnancy, childbirth and the puerperium	[12]
R00 - R99	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified	[10,12,24]
S00 - T98	Injury, poisoning and certain other consequences of external causes	[10–12,17,24,27,29]
Z00 - Z99	Factors influencing health status and contact with health services	[12,23,26,32,37]

4. Discussion

Whether for process discovery, conformance checking, or enhancement, process mining case studies are influenced by the quality of the labeled data. The benefits of high-quality, labeled data include improved accuracy, efficiency and predictability of processes, not only for the study itself but also for comparability across studies. Further, high-quality, labeled data can make other kinds of future analyses and even machine learning techniques (e.g., supervised learning, trend estimation, clustering) easier and more efficient to achieve. In process mining case studies in healthcare, labeled data often encompasses clinical aspects and terms. As such, our aim was to examine clinically-relevant case studies since Rojas et al. [4] and determine how to improve upon the clarity and comparability of clinical aspects and terms described.

4.1. Reporting Basic Characteristics of the Event Log Data

For our analysis, we selected papers that described basic characteristics of the event log data. These characteristics included the origin or source of the data, the healthcare facility, the number of cases or patients, and the timeframe of the study. For example, in Rinner et al. [18], event logs were extracted for a total of 1023 patients starting melanoma surveillance between January 2010 to June 2017, from a local melanoma registry in a medical university and Hospital Information System (HIS) in Austria. In papers where these characteristics were not clearly reported, the retrieval process was time-consuming. Several papers provided additional details (e.g., patient age, data from private insurance or public health records). Presumably for reasons of privacy and anonymity, specifics on the healthcare facility (e.g., hospital name) were not always provided, however, the country of origin was always reported. While variations exist in the style of reporting, we recommend case studies include these aforementioned basic characteristics when reporting the event log data.

4.2. Adopting the Use of Standard Clinical Descriptors

4.2.1. Encounter Environment

A patient can have vastly different experiences within the healthcare system depending on the clinical setting or encounter environment. For example, a patient with heart failure who presents to the AED may require admission as a hospital inpatient, follow-up at their GP practice site or outpatient clinic, and prescription drugs at a pharmacy. As such, in our analysis of the selected papers, we focused on five patient encounter environments: Inpatient, Outpatient, AED, GP practice site, and Pharmacy. All five encounter types can be coded by SNOMED CT. While further details can be provided (e.g., Outpatient Clinic for Thyroid Disease [32]), we recommend case studies report at least the patient encounter environment using standard clinical codes e.g., SNOMED CT.

4.2.2. Clinical Specialty

Different clinical specialties are often involved in the care of a patient. For example, for a patient diagnosed with cancer, a multidisciplinary care plan can encompass input from a medical specialty, a surgical specialty and clinical oncology. As each specialty offers their own unique set of knowledge and expertise, it is important to identify which clinical specialty is involved.

For each of our selected papers, we identified at least one of the 18 high-level clinical specialties coded by SNOMED CT. For greater specificity, SNOMED CT offers further standard clinical codes for sub-specialities. In fact, Baek et al. list multiple sub-specialities along with their corresponding SNOMED CT codes in their study [12]. Also, instead of *Clinical specialty*, another category of clinical descriptors such as the type of medical practitioner or occupation could have been considered (e.g., mapping to surgeon instead of surgical specialty).

In any event, the task of identifying and assigning such standard clinical codes is time consuming, and beyond the scope of this paper. For future case studies, we recommend reporting the clinical specialty (or similar clinical descriptor such as medical practitioner) by adopting standard clinical codes e.g., SNOMED CT.

4.2.3. Medical Diagnosis

There are literally thousands of medical diagnoses, and each diagnosis comes with its own treatment and management plan. ICD-10 is a standard coding scheme in healthcare that provides specific clinical descriptors and codes for diseases and health conditions. In our analysis, we were able to identify at least one medical diagnosis or description of a medical diagnosis in each paper, which we could map to the corresponding ICD-10 code. Further, over 25% (10 out of 38) of our selected papers utilized either ICD-9 or ICD-10 codes in their study. For broader comparison across studies, we assigned the selected papers to one or more of the 22 ICD-10 chapters or block categories. In Table 6 we only listed the ICD-10 chapters that were covered in the case studies.

It is important to distinguish the difference between a medical diagnosis (i.e., the process of identifying the disease or medical condition that explains a patient's signs and symptoms) versus a patient's signs (e.g., rash) or symptoms (e.g., cough). While the majority of ICD-10 chapters describe a group of medical diagnoses, some cover other clinical descriptors, such as signs and symptoms (R00-R99), external causes of morbidity and mortality (V01-V98), and codes for special purposes (U00-99). ICD-10 also allows for the coding of location, severity, cause, manifestation and type of health problem [43].

Taken together, we recommend adopting use of a standard coding scheme e.g., ICD-10 for clinical terms and aspects relating to medical diagnosis in process mining case studies in healthcare. Recently developed, ICD-11 is not adopted yet but provides backward compatibility, i.e., ICD-10 coded case studies will be comparable to newer ICD-11 coded ones, once the new coding scheme will be taken on by the information system vendors.

4.3. Conclusions and Future Perspectives

In summary, we propose adopting a standard for describing event log data and reporting medical terminology using standard clinical descriptors and coding schemes. In doing so, the goal is to improve accuracy and comparability across future clinically-relevant process mining case studies in health care. As such, we provide a sample checklist template of standard criteria for the reporting of such case studies, in Appendix A.

In scientific research, the idea of having a set of guidelines, criteria, or standards for peer-reviewed publications is not novel. In fact, journals such as *Nature* are taking initiatives by creating mandatory reporting summary templates (<https://www.nature.com/documents/nr-reporting-summary-flat.pdf>), in order to improve comparability, transparency, and reproducibility of the work they publish [44]. Other journals and disciplines, including biomedical informatics, are following suit [45]. Thus, as data sets become more transparent and available, consistency in reporting characteristics of the event log data (e.g., origin of data, number of patients or cases, healthcare facility, timeframe of the study) will aid in improving comparability and reproducibility across future studies.

Further to the work by Rojas et al. [28], we identified and described the clinical terms and aspects in our selected papers with respect to three categories: the patient encounter environment, clinical specialty, and medical diagnosis. We then correlated the clinical terms and aspects to their respective standard clinical descriptors and codes found in SNOMED CT and ICD-10. For studies where a higher granularity for patient encounter environments is needed, SNOMED CT offers more codes and the compositional grammar could be useful. Similarly, for *Clinical specialty* in SNOMED CT, reporting of sub-specialties under e.g., *Medical speciality* will provide increased specificity for clarity and comparison.

As aforementioned, several case studies have already adopted the use of a standard clinical coding scheme to describe medical diagnoses. However, our consideration of SNOMED CT and ICD-10 serves only as a starting point. In fact, SNOMED CT also provides standard codes for medical diagnoses, which can provide further specificity and clarity. For example, instead of ICD-10, the one of Systematized Nomenclature for Dentistry or SNODENT CT (which is incorporated into SNOMED CT) could have been used to code for the clinical descriptors of missing and filled teeth in one of our selected papers [7].

Finally, when adopting the use of standard clinical descriptors, we recognize other fundamental clinical categories to consider are medical investigations and procedures. As such, the use of standard clinical descriptors is becoming increasingly relevant, not only for clarity and comparability, but efficiency in outcome measurements such as length of stay (LOS) and financial cost. For example, in their paper, Baek et al. utilized process mining techniques and statistical methods to identify the factors associated with LOS in a South Korean hospital [12]. This study is just one use case for a more detailed description of the medical context where process mining case studies could allow for future

meta-studies, e.g., benchmarking LOS in different hospitals or countries, based on diagnoses while also considering other important factors like the patient encounter environment.

Author Contributions: Data curation, A.M.L. and D.B.; Investigation, E.H., A.M.L., D.B. and A.C.L.; Project administration, E.H.; Supervision, J.K. All authors have read and agreed to the published version of the manuscript.

Funding: The work presented in this paper was conducted in the context of the first author's PhD project and financed from internal funds of the Research Department Advanced Information Systems and Technology and the Institute for Applied Knowledge Processing.

Acknowledgments: Supported by the Process-Oriented Data Science for Healthcare Alliance (PODS4H Alliance). Open Access Funding by the University for Continuing Education Krems, the University of Applied Sciences BFI Vienna and the University of Applied Sciences Upper Austria.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Reporting Template Outline

The following sections can be used as a checklist template when writing a case study on process mining in healthcare and are intended to help address key base data, clinical and technical aspects.

Appendix A.1. Base Data Aspects

Table A1. Aspects that describe the basic characteristics of the data.

Aspect	Description / Example
Data Source	e.g., Administrative system, Clinical support system, Medical devices
Descriptive Statistics	Statistics of the base data; e.g., the number of cases and patients
Timeframe	The period during which the underlying data was collected
Geographical Area	Country or region where the data was collected

Appendix A.2. Clinical Aspects

Table A2. Clinical aspects of the mined healthcare process.

Aspect	Coding Scheme	Listing / Example
Process Type	-	Organizational or medical treatment process; following the definition of Lenz and Reichert [46]
Encounter Type	-	Elective or non-elective care
Encounter Environment	SNOMED CT	see Table 4; e.g., Inpatient (440654001)
Clinical Specialty	SNOMED CT	see Table 5; e.g., Dentistry (722163006)
Diagnosis	ICD 10	e.g., J10.0 Influenza with pneumonia, seasonal influenza virus identified
Investigations/Procedures	-	e.g., Complete blood count, X-ray imaging, Colonoscopy, Appendectomy

Appendix A.3. Technical Aspects

Table A3. Technical aspects about the process mining techniques.

Aspect	Listing / Example
Type	Discovery, Conformance or Enhancement
Perspective	Control-flow, Organizational, Case, Time
Tools (Version)	e.g., ProM 6.9 or Disco 2.2.1
Implementation Strategy	Direct, Semi-Automated, Integrated Suite
Analysis Strategy	Basic, New implementation, Extended Analysis
Methodology	e.g., L* life-cycle model
Techniques/Algorithms	e.g., Genetic Process Mining, Inductive Mining

References

1. Rojas, E.; Munoz-Gama, J.; Sepúlveda, M.; Capurro, D. Process mining in healthcare: A literature review. *J. Biomed. Inform.* **2016**, *61*, 224–236. [[CrossRef](#)] [[PubMed](#)]
2. Erdogan, T.G.; Tarhan, A. Systematic Mapping of Process Mining Studies in Healthcare. *IEEE Access* **2018**, *6*, 24543–24567. [[CrossRef](#)]
3. Helm, E.; Lin, A.M.; Baumgartner, D.; Lin, A.C.; Küng, J. Adopting Standard Clinical Descriptors for Process Mining Case Studies in Healthcare. In Proceedings of the International Conference on Business Process Management, Vienna, Austria, 1–6 September 2019; Springer: Cham, Switzerland, 2019.
4. Rojas, E.; Capurro, D. Characterization of drug use patterns using process mining and temporal abstraction digital phenotyping. In Proceedings of the International Conference on Business Process Management, Sydney, Australia, 9–14 September 2018; Springer: Cham, Switzerland, 2018; pp. 187–198.
5. Benson, T.; Grieve, G. *Principles of Health Interoperability: SNOMED CT, HL7 and FHIR*; Springer: Berlin/Heidelberg, Germany, 2016.
6. Funkner, A.A.; Yakovlev, A.N.; Kovalchuk, S.V. Data-driven modeling of clinical pathways using electronic health records. *Procedia Comput. Sci.* **2017**, *121*, 835–842. [[CrossRef](#)]
7. Fox, F.; Aggarwal, V.R.; Whelton, H.; Johnson, O. A data quality framework for process mining of electronic health record data. In Proceedings of the IEEE International Conference on Healthcare Informatics (ICHI), New York, NY, USA, 4–7 June 2018; pp. 12–21.
8. Erdogan, T.G.; Tarhan, A. A Goal-Driven Evaluation Method Based On Process Mining for Healthcare Processes. *Appl. Sci.* **2018**, *8*, 894. [[CrossRef](#)]
9. Lismont, J.; Janssens, A.S.; Odnoletkova, I.; v en Broucke, S.; Caron, F.; Vanthienen, J. A guide for the application of analytics on healthcare processes: A dynamic view on patient pathways. *Comput. Biol. Med.* **2016**, *77*, 125–134. [[CrossRef](#)]
10. Duma, D.; Aringhieri, R. An ad hoc process mining approach to discover patient paths of an Emergency Department. *Flex. Serv. Manuf. J.* **2018**, 1–29. [[CrossRef](#)]
11. Yang, S.; Sarcevic, A.; Farneth, R.A.; Chen, S.; Ahmed, O.Z.; Marsic, I.; Burd, R.S. An approach to automatic process deviation detection in a time-critical clinical process. *J. Biomed. Inform.* **2018**, *85*, 155–167. [[CrossRef](#)]
12. Baek, H.; Cho, M.; Kim, S.; Hwang, H.; Song, M.; Yoo, S. Analysis of length of hospital stay using electronic health records: A statistical and data mining approach. *PLoS ONE* **2018**, *13*. [[CrossRef](#)]
13. Mannhardt, F.; Blinde, D. Analyzing the trajectories of patients with sepsis using process mining. *CEUR Workshop Proc.* **2017**, *1859*, 72–80.
14. Tóth, K.; Machalik, K.; Fogarassy, G.; Vathy-Fogarassy, Á. Applicability of process mining in the exploration of healthcare sequences. In Proceedings of the 2017 IEEE 30th Neumann Colloquium (NC), Budapest, Hungary, 24–25 November 2017; pp. 151–156.
15. Alharbi, A.; Bulpitt, A.; Johnson, O. Improving Pattern Detection in Healthcare Process Mining Using an Interval-Based Event Selection Method. In Proceedings of the International Conference on Business Process Management, Barcelona, Spain, 10–15 September 2017; Springer: Cham, Switzerland, 2017; pp. 88–105.
16. Chen, Y.; Kho, A.N.; Liebovitz, D.; Ivory, C.; Osmundson, S.; Bian, J.; Malin, B.A. Learning bundled care opportunities from electronic medical records. *J. Biomed. Inform.* **2018**, *77*, 1–10. [[CrossRef](#)]
17. Andrews, R.; Wynn, M.T.; Vallmuur, K.; ter Hofstede, A.H.M.; Bosley, E.; Elcock, M.; Rashford, S. Pre-hospital retrieval and transport of road trauma patients in Queensland: A process mining analysis. In Proceedings of the International Conference on Business Process Management, Sydney, Australia, 9–14 September 2018.
18. Rinner, C.; Helm, E.; Dunkl, R.; Kittler, H.; Rinderle-Ma, S. Process Mining and Conformance Checking of Long Running Processes in the Context of Melanoma Surveillance. *Int. J. Environ. Res. Public Health* **2018**, *15*, 2809. [[CrossRef](#)] [[PubMed](#)]
19. Mannhardt, F.; Toussaint, P.J. Revealing Work Practices in Hospitals Using Process Mining. In *Studies in Health Technology and Informatics*; IOS Press: Gothenburg, Sweden, 2018; pp. 281–285.
20. Stefanini, A.; Aloini, D.; Dulmin, R.; Mininno, V. Service Reconfiguration in Healthcare Systems: The Case of a New Focused Hospital Unit. In Proceedings of the International Conference on Health Care Systems Engineering, Florence, Italy, 29–31 May 2017; Springer: Cham, Switzerland, 2017; pp. 179–188.

21. Kurniati, A.P.; Rojas, E.; Hogg, D.; Hall, G.; Johnson, O. The assessment of data quality issues for process mining in healthcare using Medical Information Mart for Intensive Care III, a freely available e-health record database. *Health Inform. J.* **2019**, *25*, 1878–1893. [[CrossRef](#)] [[PubMed](#)]
22. De Vries, G.J.; Neira, R.A.Q.; Geleijnse, G.; Dixit, P.; Mazza, B.F. Towards Process Mining of EMR Data. In Proceedings of the International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC), Porto, Portugal, 21–23 February 2017.
23. Kirchner, K.; Marković, P. Unveiling Hidden Patterns in Flexible Medical Treatment Processes – A Process Mining Case Study. In Proceedings of the International Conference on Decision Support System Technology, Heraklion, Greece, 22–25 May 2018; Springer: Cham, Switzerland, 2018; pp. 169–180.
24. Alvarez, C.; Rojas, E.; Arias, M.; Munoz-Gama, J.; Sepúlveda, M.; Herskovic, V.; Capurro, D. Discovering role interaction models in the Emergency Room using Process Mining. *J. Biomed. Inform.* **2018**, *78*, 60–77. [[CrossRef](#)]
25. Metsker, O.; Yakovlev, A.; Bolgova, E.; Vasin, A.; Koval-chuk, S. Identification of Pathophysiological Subclinical Variances During Complex Treatment Process of Cardiovascular Patients. *Procedia Comput. Sci.* **2018**, *138*, 161–168. [[CrossRef](#)]
26. Kirchner, K.; Marković, P.; Delias, P. Automatic Creation of Clinical Pathways - A Case Study. *Data Sci. Bus. Intell.* **2016**, *179*, 188.
27. Yang, S.; Zhou, M.; Chen, S.; Dong, X.; Ahmed, O.; Burd, R.S.; Marsic, I. Medical Workflow Modeling Using Alignment-Guided State-Splitting HMM. In Proceedings of the IEEE International Conference on Healthcare Informatics (ICHI), Park City, UT, USA, 23–26 August 2017; pp. 144–153.
28. Rojas, E.; Sepúlveda, M.; Munoz-Gama, J.; Capurro, D.; Traver, V.; Fernandez-Llatas, C. Question-driven methodology for analyzing emergency room processes using process mining. *Appl. Sci.* **2017**, *7*, 302. [[CrossRef](#)]
29. Stell, A.; Piper, I.; Moss, L. Automated Measurement of Adherence to Traumatic Brain Injury (TBI) Guidelines using Neurological ICU Data. In Proceedings of the International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC), Madeira, Portugal, 19–21 January 2018; SciTePress: Groningen, The Netherlands, 2018.
30. Fernandez-Llatas, C.; Ibanez-Sanchez, G.; Celda, A.; Mandingorra, J.; Aparici-Tortajada, L.; Martinez-Millana, A.; Munoz-Gama, J.; Sepúlveda, M.; Rojas, E.; Gálvez, V.; et al. Analyzing Medical Emergency Processes with Process Mining : The Stroke Case. In Proceedings of the International Conference on Business Process Management, Sydney, Australia, 9–14 September 2018; Springer: Cham, Switzerland, 2018; pp. 214–225.
31. Conca, T.; Saint-Pierre, C.; Herskovic, V.; Sepúlveda, M.; Capurro, D.; Prieto, F.; Fernandez-Llatas, C. Multidisciplinary Collaboration in the Treatment of Patients With Type 2 Diabetes in Primary Care: Analysis Using Process Mining. *J. Med. Internet Res.* **2018**, *20*. [[CrossRef](#)]
32. Gatta, R.; Vallati, M.; Lenkowicz, J.; Casa, C.; Cellini, F.; Damiani, A.; Valentini, V. A Framework for Event Log Generation and Knowledge Representation for Process Mining in Healthcare. In Proceedings of the IEEE International Conference on Tools with Artificial Intelligence (ICTAI), Volos, Greece, 17 December 2018; pp. 647–654.
33. Najjar, A.; Reinharz, D.; Girouard, C.; Gagné, C. A two-step approach for mining patient treatment pathways in administrative healthcare databases. *Artif. Intell. Med.* **2018**, *87*, 34–48. [[CrossRef](#)]
34. Chen, J.; Sun, L.; Guo, C.; Wei, W.; Xie, Y. A data-driven framework of typical treatment process extraction and evaluation. *J. Biomed. Inform.* **2018**, *83*, 178–195. [[CrossRef](#)]
35. Yan, H.; Van Gorp, P.; Kaymak, U.; Lu, X.; Ji, L.; Chiau, C.C.; Korsten, H.H.; Duan, H. Aligning event logs to task-time matrix clinical pathways in BPMN for variance analysis. *J. Biomed. Health Inform.* **2018**, *22*, 311–317. [[CrossRef](#)]
36. Neira, R.A.Q.; de Vries, G.J.; Caffarel, J.; Stretton, E. Extraction of Data from a Hospital Information System to Perform Process Mining. In Proceedings of the World Congress on Medical and Health Informatics MedInfo, Xiamen, China, 21–25 August 2017; pp. 554–558.
37. Dagliati, A.; Sacchi, L.; Zambelli, A.; Tibollo, V.; Pavesi, L.; Holmes, J.H.; Bellazzi, R. Temporal electronic phenotyping by mining careflows of breast cancer patients. *J. Biomed. Inform.* **2017**, *66*, 136–147. [[CrossRef](#)] [[PubMed](#)]

38. Baker, K.; Dunwoodie, E.; Jones, R.G.; Newsham, A.; Johnson, O.; Price, C.P.; Wolstenholme, J.; Leal, J.; McGinley, P.; Twelves, C.; et al. Process mining routinely collected electronic health records to define real-life clinical pathways during chemotherapy. *Int. J. Med. Inform.* **2017**, *103*, 32–41. [[CrossRef](#)] [[PubMed](#)]
39. Huang, Z.; Ge, Z.; Dong, W.; He, K.; Duan, H. Probabilistic modeling personalized treatment pathways using electronic health records. *J. Biomed. Inform.* **2018**, *86*, 33–48. [[CrossRef](#)]
40. Johnson, O.; Dhafari, T.B.; Kurniati, A.; Fox, F.; Rojas, E. The ClearPath Method for Care Pathway Process Mining and Simulation. In Proceedings of the International Conference on Business Process Management, Sydney, Australia, 9–14 September 2018; Springer: Cham, Switzerland, 2018; pp. 239–250.
41. Jimenez-Ramirez, A.; Barba, I.; Reichert, M.; Weber, B.; Del Valle, C. Clinical Processes-The Killer Application for Constraint-Based Process Interactions? In Proceedings of the International Conference on Advanced Information Systems Engineering, Tallinn, Estonia, 11–15 June 2018; Springer: Cham, Switzerland, 2018; pp. 374–390.
42. Huang, Z.; Dong, W.; Ji, L.; He, C.; Duan, H. Incorporating comorbidities into latent treatment pattern mining for clinical pathways. *J. Biomed. Inform.* **2016**, *59*, 227–239. [[CrossRef](#)]
43. World Health Organization. *International Statistical Classification of Diseases and Related Health Problems*; World Health Organization: Geneva, Switzerland, 2004; Volume 2,
44. Munafò, M.R.; Nosek, B.A.; Bishop, D.V.M.; Button, K.S.; Chambers, C.D.; Du Sert, N.P.; Simonsohn, U.; Wagenmakers, E.J.; Ware, J.J.; Ioannidis, J.P. A manifesto for reproducible science. *Nat. Hum. Behav.* **2017**, *1*, 21. [[CrossRef](#)]
45. Bakken, S. The journey to transparency, reproducibility, and replicability. *J. Am. Med. Informatics Assoc.* **2019**, *26*, 185–187. [[CrossRef](#)] [[PubMed](#)]
46. Lenz, R.; Reichert, M. IT support for healthcare processes—premises, challenges, perspectives. *Data Knowl. Eng.* **2007**, *61*, 39–58. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).