Journal of Biomedical Informatics 130 (2022) 104076

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Original Research

Are we ready for conformance checking in healthcare? Measuring adherence to clinical guidelines: A scoping systematic literature review

Eimy Oliart [a], Eric Rojas [b], Daniel Capurro [c,*]

[a] School of Computing and Information Systems, University of Melbourne, Melbourne, Australia
[b] Department of Clinical Laboratories, School of Medicine, Pontificia Universidad Catolica de Chile, Santiago, Chile
[c] School of Computing and Information Systems, Centre for the Digital Transformation of Health, University of Melbourne, Melbourne, Australia

ARTICLE INFO

Keywords:
Clinical guidelines
Electronic medical records
Declarative constraints
Medical information systems
Process mining

ABSTRACT

Clinical guidelines are recommendations of how to diagnose, treat, and manage a patient's medical condition. Health organizations must measure adherence to clinical guidelines to enhance the quality of service, but due to the complexity of the medical environment, there is no simple way of measuring adherence to clinical guidelines. This scoping review will systematically assess the criteria used to measure adherence to clinical guidelines in the past 20 years and explore the suitability of using process mining techniques. We will use a workflow protocol based on declarative and temporal constraints to translate the narrative text rules in the publications into a high-level process model. This approach will enable us to explore the main patterns and gaps identified when measuring adherence to clinical guidelines and how they affect the adoption of process mining techniques. The main contributions of this paper are a) a comprehensive analysis of the criteria used for measuring adherence, considering a diverse set of medical conditions b) a framework that will classify the level of complexity of the rules used to measure adherence based on declarative and temporal constraints c) list of key trends and gaps identified in the literature and how they relate to the use of process mining techniques in healthcare.

1. Introduction

Adherence to clinical guidelines (CGs) is considered a key quality indicator in health care, but due to the complexity of clinical environments, reliable and sustainable quality improvements are difficult to implement [1]. According to the Institute of Medicine [2, p. 27], clinical practice guidelines are defined as "systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances". CG's main objective is to provide recommendations on how to diagnose, treat and manage clinical conditions. They can also provide appropriate knowledge for health care providers, guide the allocation of resources, reduce the risk of liability for negligence in the duty of care, and enable the assessment and assurance of quality in healthcare [2].

1.1. Background and problem statement

Process mining (PM) is considered a promising technology to support health organizations assess adherence to CGs and increase the quality of their processes [3]. The main goal of PM is to discover, monitor, and enhance processes by analyzing data extracted from event logs [4,5]. Organizations can discover their processes as they occur in real life, verify compliance to regulations or practices, and obtain insights regarding resource utilization, performance issues, and bottlenecks in the workflow [6]. However, implementing PM techniques in the health industry has specific challenges due to the complexity of health processes. Combi et al. [7] describe several issues that contribute to the complexity of health processes: high level of variability; time-critical behavior of activities, especially when considering emergency situations; and the constant evolution of research which trigger changes in knowledge and available technologies. Although process mining provides an interesting perspective to analyze healthcare processes, there has been limited adoption of this technique when measuring adherence to CGs. The exploration and review of the literature regarding measuring adherence to CGs can provide a better insight on how to proceed to use PM as a measurement tool.

1.2. Related work

In the related work, the topic of how process mining and

* Corresponding author at: 700 Swanston St, Melbourne 3010, Australia.
E-mail address: dcapurro@unimelb.edu.au (D. Capurro).

https://doi.org/10.1016/j.jbi.2022.104076
Received 30 November 2021; Received in revised form 13 April 2022; Accepted 15 April 2022
Available online 4 May 2022

conformance checking has been used to check adherence to CGS, is explored. Two main approaches are mentioned: the use of declarative constraints and the use of temporal constraints. One approach has been the use of declarative constraints as the main perspective. Grando et al. [8] explored the use of declarative language to translate clinical recommendations into process models and assess compliance to computer interpretable guidelines. Implementing conformance checking techniques based on declarative models allows for greater flexibility while still maintaining compact workflows [6] and are more suitable to describe processes with high variability [9]. Bottrighi et al. [10] describe that CGs contain a mix of procedural knowledge—when physicians follow a sequence of predefined tasks and declarative knowledge—when specific constraints inform the execution of the process without restricting how the physician will behave to satisfy them. Borrego and Barba [11] and Burattin et al. [9] also used a declarative constraint approach, based on DECLARE language [12], to enhance the performance of conformance checking techniques. In addition to declarative constraints, it is important to consider temporal constraints when measuring adherence to CGs. Barba et al. [13] describe two clinical situations where a time condition must be measured: a) therapeutic procedures, where the duration of an activity and time intervals between tasks must be followed; b) drug administration, where tasks are repeated frequently. Maggi and Westergaard [14] discuss that temporal constraints aim to guarantee the correct execution of time conditions like latencies and deadlines. Overall, temporal constraints perform a key role when describing the duration, delay, and repetition of an activity, not only for providing a time formalism but also for allowing implicit temporal reasoning within CGs [15].

1.3. Motivation

Despite the growing interest in using conformance checking techniques to measure adherence to CGs, there are not extensive studies in the health domain using declarative and temporal constraints. There is no standard metric defined to easily measure adherence to CG that provides the best combination of result and efficiency. Health organizations struggle to find a balance between variability and predicted outcomes. This scoping review seeks to explore the suitability of PM techniques to assess adherence to clinical guidelines using declarative and temporal constraints. To achieve this, we systematically reviewed biomedical publications that measured adherence to CGs and applied a workflow protocol based on declarative and temporal constraints to describe and evaluate the complexity of the criteria used for measuring adherence. The definition of a new conformance checking language is out of scope for the study, the main objective is to describe, in a consistent way, the use of declarative and time constraints to measure adherence to CGs.

1.4. Research questions

This study will address the following questions: 1) How do researchers measure adherence to CGs using declarative and time constraints? Are there any patterns or trends? 2) What are the existing gaps when measuring adherence to CGs and how they affect the use of PM techniques?.

2. Methods

2.1. Inclusion and exclusion criteria

We included studies that met the following inclusion criteria: 1) observational studies that measured adherence to CGs using data from medical records; 2) clearly described the rules considered for measuring adherence and provided quantitative metrics; 3) studies published after 1st of January of 2000 and written in English language. We excluded studies with 1) interventions to modify adherence to CGs; 2) that did not provide a specific guideline or best practices; and 3) no clear definition of the rules to determine adherence to CGs.

2.2. Search strategy and study selection

We conducted our search on PubMed using the following search terms to comprehensively cover different aspects of the CGs: medication, surgical, clinical, treatment and primary care guidelines. The following terms were used to conduct the search:

```
(("clinical guideline*"[Title]) OR ("treatment guideline*"[Title])
OR ("medication guideline*"[Title]) OR ("primary care guideline*"
[Title]) OR ("surgical guideline*"[Title])) AND ("guideline adherence"[MeSH Terms]) AND measur*
```

The study screening and selection was conducted in duplicate by the authors (EO, ER, DC) and any disagreement was solved by consensus. This review was conducted in accordance to the PRISMA extension for scoping reviews [16] guideline and used Covidence [17] to support the study selection process.

2.3. Data extraction and analysis

Two authors (EO, ER) participated in the data extraction process using an iteratively tested extraction form. Criteria used for measuring adherence was extracted exclusively from the methodology section presented by each study. We did not assess the risk of bias of each individual study since our focus was on systematically collecting guideline adherence rules and not the conclusions of each primary study. Data analysis consisted of descriptive statistics and qualitative synthesis. First, we proceeded to do an interpretive review [18] of the data collected and defined a protocol for translating the text-based rules into a high-level process, by applying declarative and temporal constraints (see Section 2.4). Secondly, we used a framework approach to classify the complexity of the results obtained from our data interpretation. Both translation protocol and complexity framework were iteratively developed, tested, and reviewed. Additionally, we classified each CG based on the stage of the clinical process the study covered: 1) Screening and Prevention; b) Diagnostic; c) Treatment and Management. Risk of bias was not assessed for individual studies since this review focused on the methods to assess adherence and not the results of such assessments.

2.4. Translation protocol

The first step to translate the text-based rules into a high-level process was to identify which rules measured adherence to the CGs. After identifying the text describing the rules, we proceeded to identifying the tasks in each event. Each task was classified as being either PATIENT STATUS or ACTIVITY.

The node PATIENT STATUS represents the tasks in the event where the physician checks the eligibility of the patient for the given clinical guideline, for example, ”Low Risk LR group patients, where the status is only individuals with LDL $>=$ 5 mmol/L (193 mg/dL)”.

The node ACTIVITY represents any other action or task in the rule that is not checking a patient’s status, such as performing a surgery or prescribing a medication, for example the activity ”patients receive treatment for angiotensin converting enzyme inhibitors (ACEI)” taken from study [19].

Any description of attributes for a task was not considered as part of the high-level process, we assumed that the execution of the task would take in consideration its attributes described in the event. We also documented whether a rule checked for the presence or absence of a task.

PATIENT STATUS and ACTIVITY may be followed by a specific

constraint, which will be represented by its characteristic graphical notation, or by another ACTIVITY, this will be represented with an arrow (->). The last step identified the flow of tasks for each rule which were described by using conditional gateways, declarative constraints, and time constraints, as listed in Table 1. Table 1 includes conditional gateways, declarative constraints, and time constraints mainly based on the constraints-based conformance checking language DECLARE used with process mining in the past[12,9,6]. The conditional gateways and the declarative constraints (existence, relation and negation) are directly supported by DECLARE. Temporal constraints Loop and Clock are also supported by DECLARE but use either a ratified condition or a variable for time to determine the existence or not of an activity.

Table 1
Translation protocol constraints and notations.

Type	Constraint	Graphical Notation	Example Rule
Existence	Existence (N)	[Activity1] E(N)	Alcohol/drug use: Documentation of presence of current use (study [20])
	Absence(N)	[Activity1] A(N)	Patient with no apparent contraindication/ intolerance to aspirin is prescribed aspirin (study [21])
	Exactly(N)	[Activity1] X(N)	3 reperfusions via either thrombolysis or angioplasty IF (study [22])
Relation	Precedence	[Activity1]→• [Activity2]	Blood cultures prior to antibiotics (study [23])
	Succession	[Activity1] •→•[Activity2]	No "Succession" examples were identified in the studies
	Response	[Activity1] •→[Activity2]	Specialist assessment: a documented assessment done by a senior physician within 24 h after the suicide attempt (study [24])
Negation	Not Precedence	[Activity1] ↛•[Activity2]	No "NOT Precedence" examples were identified in the studies
	Not Succession	[Activity1] •↛•[Activity2]	No "NOT Succession" examples were identified in the studies
	Not Response	[Activity1] •↛[Activity2]	No "NOT Response" examples were identified in the studies
Temporal	Clock	[Activity1] CLOCK (</ >/ =) X time [Activity2]	Glucose measured within 7 h of birth (study [25])
	Loop	[Activity1] LOOP until [Activity2]	Patients on osteoporosis pharmacotherapy receive dual-energy X-ray absorptiometry (DXA) scans on a yearly basis until BMD is stabilized (study [26])
Conditional	AND	→AND ([Activity1], [Activity2], …)	Avoidance of overtreatment: Proportion of patients with BC not greater than 3 cm who underwent BCS as primary treatment (study [27])
	OR	→OR ([Activity1], [Activity2], …)	Diabetes: percentages of patients with at least 1 A1c test, at least 1 low-density lipoprotein (LDL) test, an eye exam (study [28])
	XOR	→XOR ([Activity1], [Activity2], …)	Diagnosis is obtained through documentation of the review of >= 5 DSM-IV criteria or through specific PHQ results (study [20])

2.5. Complexity classification framework

The complexity framework was based on the nodes, conditional gateways, declarative constraints, and temporal constraints identified in the previous stage. Each type of variable was analyzed and given a complexity weight to describe the level of difficulty of measuring that variable in the high-level process. Additionally, we defined specific categories to account for the different aspects of each variable like variety, which identified the different types of value that a variable can assume; frequency, which identified the number of times that a variable exists in the process; repetition, which described the number of times a task was executed; and scenario, which described the number of different paths a process can have. The categories for frequency, repetition, and scenarios, have established ranges, which were based on the distribution of the observations. Table 2 describes the framework used to calculate the complexity score of each study and defines the level of complexity according to the final scores.

Table 3 describes an example of how we calculated the scores for the complexity classification framework. This example is based on the two rules described in Table 4 and Table 5. It is important to highlight that the complexity classification framework does not take in consideration the complexity of the actual task in the high-level process, like performing a surgery versus prescribing a drug. We aimed at evaluating the complexity of defining a high-level model process that can be used for conformance checking, so our focus is in estimating the level of difficulty to execute the process notations used to measure the adherence of the guideline.

3. Results

3.1. Overview of selected studies

We conducted our search in August 2020 and retrieved a list of 100 studies with a total of 31 studies included in our research project [29–31,22,32,33,25,34,27,35,26,36–41,20,42–48,23,21,49,19,28,24]. The complete list can be found in the supplementary material (additional file1). Fig. 1 shows the study selection process and the final outcomes.

We observed that 55% of the studies used data collected from chart reviews, case notes, and encounter notes, which store patient data in a non-structured form. A total of 35% of the studies used registries or medical databases. Using registries is beneficial since patient information is entered in a structured database which limits inconsistencies and reduces errors [19]. The last 10% of the studies used a combination of structured and unstructured data to measure adherence to the CG. The publications reviewed covered a variety of diseases organized into 6 categories: cancer, cardiovascular, infectious, metabolic, mental, and other. Cancer had the highest number of studies, 6 in total. Cardiovascular, infectious, and metabolic diseases each had a total of 5 studies; mental illness had a total of 4. All studies assessed only one medical condition, except for [28] which measured adherence to treatment guidelines of four different medical conditions: asthma, chronic obstructive pulmonary disease (COPD), depression, and diabetes. The conditions and specific diseases assessed in each study are detailed in Table 6. The CGs referenced in the studies varied according to the disease measured. A total of 90% of the studies referenced one guideline whereas the other 10% referenced 2 or 3 guidelines. Additionally, 69% of the studies focused measurement of adherence in the treatment and management phase of the CG, 22% were related to the screening and prevention phase, and 8% were related to the diagnostic phase. Our analysis also compared the medical condition versus the phase of the guideline from which criteria for measuring adherence was obtained (additional file 2).

Table 2
Complexity level classification framework.

Variable/Weight		Variety/Weight		Frequency/Weight		Repetition/Weight		Scenarios/Weight	
Activity and Patient Status	1	one type	1	01–20	1	NA	NA	NA	NA
				21–40	2				
		two types	2	41–60	3				
				> 60	4				
Clock and Loop	2	one type	1	1–3	1	NA	NA	NA	NA
		two types	2	4–7	2				
				> 7	3				
Existence and Exactly	2	one type	1	01–20	1	1–3	1	NA	NA
				21–40	2	4–7	2		
		two types	2	41–60	3	>7	3		
				> 60	4				
Precedence and Response	3	one type	1	1–3	1	NA	NA	NA	NA
		two types	2	4–7	2				
				> 7	3				
AND, XOR, OR, AND(OR)	3	one type	1	1–3	1	NA	NA	01–10	1
		two types	2	4–7	2			11–20	2
		three types	3	>7	3			21–30	3
		four types	4					>30	4
Absence	4	NA	NA	1–3	1	1–3	1	NA	NA
				4–7	2	4–7	2		
				> 7	3	>7	3		
Not Response	4	NA	NA	1–3	1	NA	NA	NA	NA
				4–7	2				
				> 7	3				
Complexity Level		Low				Medium			high
Range		1–20				21–40			>41

Table 3
Example application of the complexity framework on two exemplar rules.

Rule 1	[patient status]→[activity1]→•[activity2]
Patient Status and Activity Frequency: 3 (weight 1) Total score: (2 + 1) x 2 = 6	Variety: 2 (weight 2)
Precedence and Response Frequency:1 (weight 1) Total score: (1 + 1) x 3 = 6	Variety: 1 (weight 1)
Total score rule 1	6 + 6 = 12
Rule 2	[patient status]→XOR ([activity1], [activity2])
Patient Status and Activity Frequency: 3 (weight 1) Total score: (2 + 1) x 2 = 6	Variety: 2 (weight 2)
AND, OR, XOR, AND(OR) Frequency:1 (weight 1) Scenarios: 1 (weight 1) Total score: (1 + 1 + 1) x 3 = 9	Variety: 1 (weight 1)
Total score rule 2	6 + 9 = 15
Final Score Rules 12 + 15 = 27	Total Score Rule 1 + Total Score Rule 2

Table 4
Example 1 of the process used to translate a narrative description of an adherence rule.

Original Text	"Surgery must have preceded chemotherapy for stages I-IIIB to be consideres adherent to National Comprehensive Cancer Network guidelines, whereas for stages IIIC-IV either initial surgery or chemotherpy was characterized as appropriate care [29]"
Rule per event	Event 1: for stages I-IIIB, surgery must have preceded chemotherapy.
Rule per High-level process	*Example Rule 1: [patient status]→[activity1]→•[activity2]*

Table 5
Example 2 of the process used to translate a narrative description of an adherence rule.

Original Text	"Surgery must have preceded chemotherapy for stages I-IIIB to be consideres adherent to National Comprehensive Cancer Network guidelines, whereas for stages IIIC-IV either initial surgery or chemotherpy was characterized as appropriate care [29]"
Rule per event	Event 2: for stages IIIC-IV, either initial surgery or chemotherapy.
Rule per High-level process	*Example Rule 2: [patient status]→XOR ([activity1], [activity2])*

3.2. Data extraction synthesis

After applying the translation protocol to define the high-level processes, we identified a total of 255 rules that measured adherence to the CGs. The average number of rules per study was 8.2 with a 7.4 standard deviation. Breast cancer study [25] had the highest number of rules with a total of 30, as shown in Table 6. Studies in the infectious disease category had the highest average number of rules,12. We identified a total of 483 nodes in the high-level processes, of which 140 were PATIENT STATUS and 343 were ACTIVITY. The average number of nodes per study was 15.6 with a standard deviation of 15.3. For the declarative constraints, our translation protocol identified 17 occurrences for ABSENCE, 12 for EXISTENCE and 2 for EXACTLY. Regarding the relation template, 12 checked for RESPONSE, 11 for PRECEDENCE and none for SUCESSION. The last declarative constraint identified was NOT RESPONSE, which had 2 rules that described a negative relation between activities. Overall, adherence to breast cancer treatment guidelines [25] was the study that had the highest number of declarative constraints, with a total of 1 instance for EXACTLY, 2 for EXISTENCE, 2 for PRECEDENCE, 6 for RESPONSE, 2 for NOT RESPONSE and 11 for ABSENCE. Of the 31 studies reviewed, 2 had loop restrictions and 11 had time restrictions. A total of 32 rules measured time restrictions between activities and a total of 4 rules had conditional loops defined. The average number of temporal constraints per study was 1.2 with a 2.0

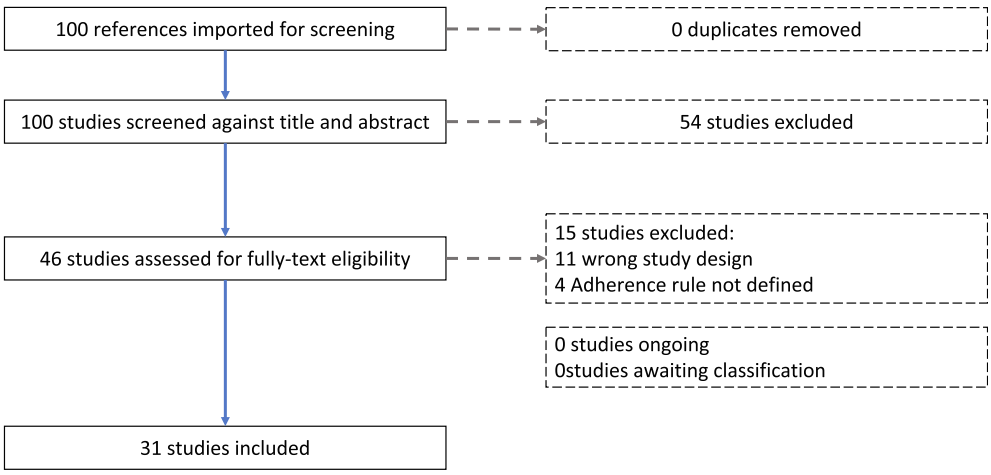

Fig. 1. PRISMA Diagram.

standard deviation. A total of 14 studies included some type of conditional constraint. Eight studies had XOR gateways in their rules, followed by 6 with AND gateways, 2 with OR gateways, and 1 with AND (OR) gateway. The average number of conditional gateways per study was 1.3 with a 2.2 standard deviation. Table 6 shows a summarized result of the data extracted after applying the translation protocol.

A detailed result of the translation protocol for each type of node, constraint and gateway per study is available in the supplementary material (additional file 3).

3.3. Complexity classification

Our proposed framework for calculating the complexity of the rules that measured adherence to clinical guidelines, classified 71% of the studies as having low complexity, 26% as medium complexity and 3% as high complexity. Table 6 shows a summary of the final scores obtained for each study. Only one publication [30] obtained a high complexity level with a final score of 89, which measured adherence to treatment and management guidelines for breast cancer. The second highest score was 38 and the study measured adherence to treatment and management guidelines for ovarian cancer [29]. The outcome of the complexity classification framework, calculated per variable and category for each of the studies, is accessible in the supplementary material (additional file 4).

4. Discussion

Overall, this scoping review shows that the assessment of adherence to clinical guidelines is still a bespoke and heterogeneous process where most studies assess adherence to only a limited number of recommendations—when clinical guidelines frequently contain tens of recommendations— and use heterogeneous and frequently unstructured data sources.

4.1. Lack of standard method to measure adherence

Our results show that the methods used by researchers to measure adherence to CGs is very heterogeneous. No studies used computer interpretable CGs, instead, all used natural language to describe what needed to be checked to confirm adherence. Why is there no standard method to measure adherence to CGs? One main barrier is the way CGs are written. Language is ambiguous and CGs don't intend to formally dictate medical practice; they support physician's decision making and are open to individual interpretation [50]. According to Dykes [51, p. 67] "many guidelines lack explicit recommendations that can be implemented into objective measurement criteria". An ideal method to check the conformance of a CG would be to evaluate its process model. Many CGs have workflow diagrams that describe the tasks required to implement a medical recommendation. We observed only one study [49], regarding treatment guidelines for colon cancer, which measured adherence to CGs based on a workflow diagram.

4.2. Assessment based on sections of the guideline

An interesting pattern identified was that all publications reviewed in our research measured adherence based on sections of the guideline and not the complete guideline. One example is the study that measured adherence to the ESC heart failure treatment guideline [19]. This guideline has multiple recommendations across 13 different topics ranging from definition and diagnosis, prognosis, pharmacological and non-surgical treatments [52]. Our analysis identified only two items related to specific sections of the pharmacological treatment recommendations. In our research no study assessed adherence to a complete CG. Our results showed that 45% of the studies had between 1 and 5 rules, and 29% had between 6 and 10 rules. Both ranges are considered low when compared with the number of recommendations contained in CGs. Although beyond the scope of this study, several reasons might explain this. It is likely that studies are assessing only the most critical sections of the guideline, sections where there is a stronger association with patient outcomes. Another option could be that they are assessing sections where adherence might be less than expected. A third option can also be the availability of clinical data to appropriately assess adherence. Collecting health data that represents the complete lifecycle of a medical condition of a patient is not simple. A variety of information systems coexists within health organizations, they are usually not well-integrated [6], and data quality might be less than ideal.

4.3. Unstructured data and diversity of standards and sources

Medical records store a patient́s data with a specific structure that allows health professionals to read and interpret the information, usually allowing for fields with free text, medical jargon, and different abbreviations [53]. Consequently, they may contain ambiguous language and are subject to human error. Capturing the patient's clinical attributes in an unstructured form makes it more complex to extract relevant information and significantly affects data quality [54]. Our results showed that a proportion of studies involved manual data extraction; this was necessary to correctly interpret the information and mitigate data quality issues due to unstructured data. On the other hand, the studies that used registries or medical databases, were able to confirm adherence by querying data stored in structured formats.

Table 6
Details of the Medical Conditions and its respective details.

General Medical Condition	Total Studies	Study Id	Diagnoses, risk factors, therapies and chronic conditions	Number of Rules	Number of Nodes	Declarative Constraints	Time Constraints	Conditional Gateways	Final Score	Complexity Level
Cancer	5	29	Breast Cancer [36]	30	71	24	3	10	89	HIGH
		50	Colon Cancer [49]	10	25	0	0	5	19	LOW
		70	Ovarian Cancer [29]	6	18	2	0	5	38	MEDIUM
		97	Breast Cancer [27]	12	22	2	0	1	27	MEDIUM
		30	Early Stage Breast Cancer [47]	3	4	0	0	0	5	LOW
Cardiovascular Disease	7	10	Stroke [45]	2	5	0	1	1	15	LOW
		16	Acute Coronary Syndrome [43]	1	5	0	0	1	12	LOW
		21	Coronary Heart Disease [21]	23	53	0	0	5	23	MEDIUM
		35	Acute Myocardial Infarction [22]	4	4	0	0	0	5	LOW
		67	Heart Failure [19]	2	5	0	0	1	11	LOW
		75	Cardiovascular Disease Risk [34]	11	22	1	0	0	18	LOW
		88	Hypertension [38]	9	9	0	0	0	6	LOW
Infectious Diseases	5	15	Febrile Children [40]	21	21	0	0	0	12	LOW
		38	Community Acquired Pneumonia [32]	20	38	6	3	0	20	LOW
		77	Pneumonia and Sepsis [23]	6	13	1	3	0	16	LOW
		85	Ventilator-associated Pneumonia [36]	7	12	0	0	0	6	LOW
		20	Sepsis [37]	6	13	1	1	0	16	LOW
Metabolic Diseases	5	86	Diabetes [28]	3	15	3	3	2	26	MEDIUM
		100	Diabetes Mellitus [46]	7	14	0	0	0	6	LOW
		83	Neonatal Hypoglycaemia [25]	8	14	4	5	0	23	MEDIUM
		53	Obesity [33]	2	3	2	0	1	19	LOW
		78	Obesity [39]	2	2	0	0	0	5	LOW
Mental Illness	4	48	Depression and Suicide [24]	14	17	4	3	0	26	MEDIUM
		99	Depression [20]	20	31	0	7	3	26	MEDIUM
		86	Depression [28]	2	15	3	3	2	26	MEDIUM
		92	Attention-Deficit/ Hyperactivity Disorder [31]	2	5	2	2	0	18	LOW
Other	8	86	Asthma [28]	2	15	3	3	2	26	MEDIUM
		86	Chronic obstructive pulmonary disease [28]	2	15	3	3	2	26	MEDIUM
		32	Osteoporosis [26]	4	8	3	6	02	29	MEDIUM
		65	Glucocorticoid-induced Osteoporosis [41]	2	5	0	0	1	12	LOW
		49	Breast Reduction [44]	1	1	0	0	0	5	LOW
		60	Tobacco Cessation [42]	5	10	0	0	0	6	LOW
		76	Chronic Non-Cancer Pain [35]	5	9	0	0	0	6	LOW
		69	Stroke Rehabilitation [48]	3	9	1	0	1	20	LOW

4.4. Complexity of the rules that measure adherence

An important highlight of our research was assessing the complexity of the rules defined to measure CG adherence. The first characteristic observed was how publications structured the rules. Forty-five percent of the studies defined rules based on a list of criteria, like quality indicators or simply a list of items to be checked for adherence. Nine percent of the publications defined a unique rule and 8% defined a sub-group of rules to measure adherence. We noticed that the studies based on a list of criteria, had a higher number of rules that were not complex, mostly containing one node and no declarative or temporal constraints. On the other hand, both the publications that defined a unique rule or a sub-group of rules had higher complexity rates with increased number of nodes and constraints, but fewer rules per study. It was not clear if criteria were equally important when calculating adherence to the CGs; none of the studies defined a mechanism to weigh different criteria. It is important to note that medical recommendations based on different levels of research evidence must translate into different enforcement levels when measuring adherence to CGs. Overall, these results demonstrate that even though CGs are complex by nature, most research that measures adherence to the CGs does not implement complex rules to check compliance.

4.5. Trends for declarative and temporal constraints

CGs are dynamic, ambiguous, and complex[5]; as a result, techniques for measuring adherence need to allow for flexibility and variability[6,9]. The constraint-based workflow protocol used to translate

the text-based rules into a high-level process revealed specific patterns regarding the criteria used to measure adherence. The first pattern is related to declarative constraints. Our observations showed that 45% of the studies had text-based rules that used declarative language to describe the criteria for measuring adherence to the guideline. The most common constraint was ABSENCE with 17 occurrences, of which 15 checked for the non-execution of a task in the process. The criteria that measure adherence based on the non-execution of an activity may allow researchers to assess compliance in a more flexible manner, by not restricting possible scenarios that can be compliant to the CG. Temporal constraints were the second pattern analyzed. Of the total number of rules translated, 13% had time restrictions. A total of 33 incidences measured linear time restrictions based on hours, days, weeks, months, and years; and a total of 4 incidences measured circular time restrictions where a task was periodically executed until a specific condition was met. As a result, these trends show that the criteria used to measure adherence does not translate into a complex process model and the use of declarative and temporal constraints allow for a high-level of variability common to CGs and appropriately represent the time-critical behavior of the healthcare processes, respectively.

4.6. Limitations

Our translation protocol classified the nodes of the high-level process as PATIENT STATUS and ACTIVITY, it did dis-aggregate on the specific type of task being executed like performing surgery, administering chemotherapy, or receiving a specific drug treatment. Identifying all the different clinical activities in the rules that measure adherence, would require a more complex workflow protocol to translate the rules into a detailed process model. Finally, the method to define the levels of complexity, has not yet been externally validated, but have included detailed examples so the reader can appropriately reproduce it.

5. Conclusions

This research systematically reviewed how publications measure adherence to CGs. Our interpretive review of the data and the translation of text-based rules into a high-level process, allowed us to identify relevant patterns and gaps when checking compliance. Even though CGs are complex and ambiguous by nature, our results demonstrate that in the past 20 years the assessment of guideline adherence has been very simplistic. Most studies only assess sections of the guideline, and there is no standard method to do so. The patterns identified when translating the rules for measuring adherence into a high-level process revealed that most of them have a simple structure for execution, and few used more complex declarative and temporal constraints. Although using PM techniques to check compliance with CGs is feasible, our research highlighted the following gaps that limit a successful implementation of this technology and how they may be addressed:

- Lack of clear and concise rules for measuring adherence: Metrics should be developed to establish how to classify and measure adherence to CGs. Using concise rules will determine the structure of a CG and each of its components, allowing the generation of automated tools to measuring adherence and use process mining tools.
- Process models not defined: Process mining should be the tool to generate the executed process models using healthcare data from information systems, to provide a way to compare and measure the adherence to the CG. As a CG is written, and concise rules are followed, a process model should be generated automatically to calculate the adherence.
- No use of computer-interpretable guidelines: Generating CGs in only natural language limits the usability of the CG in real healthcare environments. Generating computer-interpretable guidelines, following multiple available methodologies, will become an input for the development of real time decision support systems, which can really make an impact in healthcare processes.
- Use of unstructured data which require human interpretation: Rules and codes to store data should be defined and established to avoid generation of unstructured data, which eventually will avoid ambiguous misinterpretations that may lead to human errors. Process Mining tools will also be favorably benefited with the use of better structured data, allowing an easier generation and analysis of process models to measure adherence.
- Data not representing every aspect of the CGs, such as different weights of research evidence and clinical recommendations: CGs must be build based on clinical and research evidence. The more details from clinical data from multiple sources a CG has, the better guide it can be in the daily practice and as part of a potential clinical decision support system.

We conclude health organizations must overcome these barriers to use PM technologies and conformance checking techniques to automatically assess adherence to CGs. Most importantly, CGs should define standards for measuring adherence so that the quality of services can be thoroughly evaluated.

Funding and Conflicts of Interest

This research did not receive external funding and the authors declare no competing interests. Dr. Capurro receives funding for his research through the Australian Medical Research Future Funds on topics unrelated to this study. Data used for this review can be sought by contacting the corresponding author.

CRediT authorship contribution statement

Eimy Oliart: Investigation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing. **Eric Rojas:** Formal analysis, Writing – original draft, Writing – review & editing. **Daniel Capurro:** Conceptualization, Methodology, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jbi.2022.104076.

References

[1] P. Varkey, M.K. Reller, R.K. Resar, Basics of quality improvement in health care, Mayo Clin. Proc. 82 (6) (2007) 735–739, https://doi.org/10.4065/82.6.735.
[2] Institute of Medicine, Guidelines for Clinical Practice: From Development to Use, National Academies Press, 1992. https://doi.org/10.17226/1863.
[3] M. Ghasemi, D. Amyot, Process mining in healthcare: A systematised literature review, Int. J. Electronic Healthcare 9 (1) (2016) 60–88, https://doi.org/10.1504/IJEH.2016.078745.
[4] W.M. Van Der Aalst, Process Mining: Discovery, Conformance and Enhancement of Business Processes, Springer Science & Business Media, 2011.
[5] E. Rojas, J. Munoz-Gama, M. Sepúlveda, D. Capurro, Process mining in healthcare: A literature review, J. Biomed. Informat. (2016) 224–236, https://doi.org/10.1016/j.jbi.2016.04.007.
[6] M. Rovani, F.M. Maggi, M. De Leoni, W.M. Van Der Aalst, Declarative process mining in healthcare, Expert Syst. Appl. (23) (2015) 9236–9251, https://doi.org/10.1016/j.eswa.2015.07.040.
[7] C. Combi, G. Pozzi, P. Veltri, Process Modeling and Management for Healthcare, CRC Press, 2018, https://doi.org/10.1201/b21884.
[8] M.A. Grando, M.H. Schonenberg, W. van der Aalst, Semantic-Based Conformance Checking of Computer Interpretable Medical Guidelines, Communications in

Computer and Information, Science 273 (2011) 285–300, https://doi.org/10.1007/978-3-642-29752-6.
[9] A. Burattin, F.M. Maggi, A. Sperduti, Conformance checking based on multi-perspective declarative process models, Expert Syst. Appl. 65 (2016) 194–211.
[10] A. Bottrighi, F. Chesani, P. Mello, M. Montali, S. Montani, P. Terenziani, Conformance checking of executed clinical guidelines in presence of basic medical knowledge, Lecture Notes in Business Information Processing 100 LNBIP (PART 2), 2012, pp. 200–211. https://doi.org/10.1007/978-3-642-28115-0_20.
[11] D. Borrego, I. Barba, Conformance checking and diagnosis for declarative business process models in data-aware scenarios, Expert Syst. Appl. 11 (2014) 5340–5352, https://doi.org/10.1016/j.eswa.2014.03.010.
[12] W.M. van Der Aalst, M. Pesic, H. Schonenberg, Declarative workflows: Balancing between flexibility and support, Comput. Sci.-Res. Develop. 23 (2) (2009) 99–113.
[13] I. Barba, A. Lanz, B. Weber, M. Reichert, C. Del Valle, Optimized time management for declarative workflows, Lecture Notes in Business Information Processing 113 LNBIP, 2012, pp. 195–210. https://doi.org/10.1007/978-3-642-31072-0_14.
[14] F.M. Maggi, M. Westergaard, Using timed automata for a priori warnings and planning for timed declarative process models, Int. J. Cooperative Informat. Syst. 23 (1) (2014), https://doi.org/10.1142/S0218843014400036.
[15] L. Anselma, P. Terenziani, S. Montani, A. Bottrighi, Towards a comprehensive treatment of repetitions, periodicity and temporal constraints in clinical guidelines, Artif. Intell. Med. 38 (2) (2006) 171–195, https://doi.org/10.1016/j.artmed.2006.03.007.
[16] A.C. Tricco, E. Lillie, W. Zarin, K.K. O'Brien, H. Colquhoun, D. Levac, D. Moher, M. D. Peters, T. Horsley, L. Weeks, et al., Prisma extension for scoping reviews (prisma-scr): checklist and explanation, Ann. Intern. Med. 169 (7) (2018) 467–473.
[17] Veritas Health Innovation, Covidence systematic review software.
[18] K. Seers, What is a qualitative synthesis? Evidence-Based Nursing 15 (4) (2012) 101, https://doi.org/10.1136/ebnurs-2012-100977.
[19] M.G. Crespo-Leiro, J. Segovia-Cubero, J. González-Costello, A. Bayes-Genis, S. López-Fernández, E. Roig, M. Sanz-Julve, C. Fernández-Vivancos, M. de Mora-Martín, J.M. García-Pinilla, A. Varela-Román, L. Almenar-Bonet, A. Lara-Padrón, L. de la Fuente-Galán, J. Delgado-Jiménez, Adherence to the ESC Heart Failure Treatment Guidelines in Spain: ESC Heart Failure Long-term Registry, Revista Española de Cardiología (English Edition) 68 (9) (2015) 785–793, https://doi.org/10.1016/j.rec.2015.03.008.
[20] S.K. Dobscha, M.S. Gerrity, K. Corson, A. Bahr, N.M. Cuilwik, Measuring adherence to depression treatment guidelines in a VA primary care clinic, Gen. Hosp. Psychiatry 25 (4) (2003) 230–237, https://doi.org/10.1016/S0163-8343(03)00020-3.
[21] M. Kamyar, B.J. Johnson, J.J. McAnaw, R. Lemmens-Gruber, S.A. Hudson, Adherence to clinical guidelines in the prevention of coronary heart disease in type II diabetes mellitus, Pharmacy World Sci. 30 (1) (2008) 120–127, https://doi.org/10.1007/s11096-007-9154-7.
[22] C.R. Levy, T.A. Radcliff, E.T. Williams, E. Hutt, Acute Myocardial Infarction in Nursing Home Residents: Adherence to Treatment Guidelines Reduces Mortality, But Why Is Adherence So Low? J. Am. Medical Directors Assoc. (1) (2009) 56–61, https://doi.org/10.1016/j.jamda.2008.08.009.
[23] S.A. Trent, Z.J. Jarou, E.P. Havranek, A.A. Ginde, J.S. Haukoos, Variation in Emergency Department Adherence to Treatment Guidelines for Inpatient Pneumonia and Sepsis: A Retrospective Cohort Study, Acad. Emerg. Med. 26 (8) (2019) 908–920, https://doi.org/10.1111/acem.13639.
[24] T. Forsner, A.Å. Wistedt, M. Brommels, Y. Forsell, An approach to measure compliance to clinical guidelines in psychiatric care, BMC Psychiatry 8 (2008) 1–9, https://doi.org/10.1186/1471-244X-8-64.
[25] S.L. Sundercombe, C.H. Raynes-Greenow, A.E. Carberry, R.M. Turner, H.E. Jeffery, Audit of a clinical guideline for neonatal hypoglycaemia screening, J. Paediatr. Child Health 49 (10) (2013) 833–838, https://doi.org/10.1111/jpc.12293.
[26] C.R. Mountjoy, S.P. Shrader, K.R. Ragucci, Compliance with osteoporosis treatment guidelines in postmenopausal women, Ann. Pharmacother. 43 (2) (2009) 242–250, https://doi.org/10.1345/aph.1L464.
[27] A. Cowppli-Bony, B. Tretarre, E. Marrer, G. Defossez, L. Daubisse-Marliac, G. Coureau, P. Minicozzi, A.S. Woronoff, P. Delafosse, F. Molinie, Compliance with clinical guidelines for breast cancer management: A population-based study of quality-of-care indicators in France, PLoS ONE 14 (10) (2019) 1–17, https://doi.org/10.1371/journal.pone.0224275.
[28] C.R. Cantrell, J.L. Priest, C.L. Cook, J. Fincham, S.P. Burch, Adherence to treatment guidelines and therapeutic regimens: A US claims-based benchmark of a commercial population, Population Health Manage. 14 (1) (2011) 33–41, https://doi.org/10.1089/pop.2010.0018.
[29] R.E. Bristow, J. Chang, A. Ziogas, H. Anton-Culver, Adherence to treatment guidelines for ovarian cancer as a measure of quality care, Obstet. Gynecol. 121 (6) (2013) 1226–1234, https://doi.org/10.1097/AOG.0b013e3182922a17.
[30] M. Lebeau, S. Mathoulin-Pélissier, C. Bellera, C. Tunon-De-Lara, A. Daban, F. Lipinski, D. Jaubert, P. Ingrand, V. Migeot, Breast cancer care compared with clinical guidelines: An observational study in France, BMC Public Health 11 (2011), https://doi.org/10.1186/1471-2458-11-45.
[31] P.L. Geltman, L.E. Fried, L.N. Arsenault, A.M. Knowles, D.A. Link, J.N. Goldstein, J. M. Perrin, K.A. Hacker, A planned care approach and patient registry to improve adherence to clinical guidelines for the diagnosis and management of attention-deficit/hyperactivity disorder, Academic Pediatrics (3) (2015) 289–296, https://doi.org/10.1016/j.acap.2014.12.002.
[32] S.S. Al-Abri, S. Al-Maashani, Z.A. Memish, N.J. Beeching, An audit of inpatient management of community-acquired pneumonia in Oman: A comparison with regional clinical guidelines, J. Infect. Public Health 3 (2012) 250–256, https://doi.org/10.1016/j.jiph.2012.03.002.
[33] N. Farran, P. Ellis, M. Lee Barron, Assessment of provider adherence to obesity treatment guidelines, J. Am. Acad. Nurse Practition. 25 (3) (2013) 147–155, https://doi.org/10.1111/j.1745-7599.2012.00769.x.
[34] A.S. Chopra, M. Miyatani, B.C. Craven, Cardiovascular disease risk in individuals with chronic spinal cord injury: Prevalence of untreated risk factors and poor adherence to treatment guidelines, J. Spinal Cord Med. 1 (2018) 2–9, https://doi.org/10.1080/10790268.2016.1140390.
[35] R. Sekhon, N. Aminjavahery, C.N. Davis, M.J. Roswarski, C. Robinette, Compliance with Opioid Treatment Guidelines for Chronic Non-Cancer Pain (CNCP) in Primary Care at a Veterans Affairs Medical Center (VAMC), Pain Medicine (United States) 14 (10) (2013) 1548–1556, https://doi.org/10.1111/pme.12164.
[36] M. Sakaguchi, N. Shime, N. Iguchi, A. Kobayashi, K. Takada, L.E. Morrow, Effects of adherence to ventilator-associated pneumonia treatment guidelines on clinical outcomes, J. Infect. Chemotherapy 4 (2013) 599–606, https://doi.org/10.1007/s10156-012-0522-4.
[37] J.M. De Miguel-Yanes, J.A. Andueza-Lillo, V.J. González-Ramallo, L. Pastor, J. Muñoz, Failure to implement evidence-based clinical guidelines for sepsis at the ED, Am. J. Emerg. Med. 24 (5) (2006) 553–559, https://doi.org/10.1016/j.ajem.2006.01.012.
[38] A.R. Adedeji, J. Tumbo, I. Govender, Adherence of doctors to a clinical guideline for hypertension in Bojanala district, North-West Province, South Africa, Afr. J. Primary Health Care Family Med. 7 (1) (2015) 1–6, https://doi.org/10.4102/phcfm.v7i1.776.
[39] E.R. Barnes, L.A. Theeke, J. Mallow, Impact of the Provider and Healthcare team Adherence to Treatment Guidelines (PHAT-G) intervention on adherence to national obesity clinical practice guidelines in a primary care centre, J. Eval. Clin. Practice 21 (2) (2015) 300–306, https://doi.org/10.1111/jep.12308.
[40] D.L. Schriger, L.J. Baraff, K. Buller, M.A. Shendrikar, S. Naga, E.J. Lin, V. J. Mikulich, S. Cretin, Implementation of Clinical Guidelines Using a Computer Charting System, Jama 278 (19) (1997) 1585, https://doi.org/10.1001/jama.1997.03550190049043.
[41] D. Kirigaya, T. Nakayama, T. Ishizaki, S. Ikeda, T. Satoh, Management and treatment of osteoporosis in patients receiving long-term glucocorticoid treatment: current status of adherence to clinical guidelines and related factors, Internal Med. (Tokyo, Japan) 22 (2011) 2793–2800.
[42] M.B. Conroy, N.E. Majchrzak, C.B. Silverman, Y. Chang, S. Regan, L.I. Schneider, N.A. Rigotti, Measuring provider adherence to tobacco treatment guidelines: A comparison of electronic medical record review, patients survey, and provider survey, Nicotine Tobacco Res. 7 (SUPPL. 1) (2005) 35–43, https://doi.org/10.1080/14622200500078089.
[43] Z. Castañeda-Amado, L. Calixto-Aguilar, C.L. Munarriz, F.A.M. Palomino, Medical compliance to evidence-based clinical guidelines on secondary prevention of coronary heart disease in a hospital from Lima, Peru: a retrospective study, Medwave 5 (2017) e6989–e6989.
[44] D. Aydin, L.B. Hansen, P. Ikander, J.B. Drejoe, S. Lambaa, L.R. Holmich, Poor adherence to clinical guidelines for women undergoing breast reduction, Danish Medical J. 1 (2016) A5179.
[45] A. Bowen, P. Knapp, A. Hoffman, D. Lowe, Psychological services for people with stroke: Compliance with the UK National Clinical Guidelines, Clin. Rehabilit. 19 (3) (2005) 323–330, https://doi.org/10.1191/0269215505cr799oa.
[46] C.-L. Lai, Y.-H. Hou, The association of clinical guideline adherence and pay-for-performance among patients with diabetes, J. Chinese Medical Assoc.: JCMA 2 (2013) 102–107, https://doi.org/10.1016/j.jcma.2012.06.024.
[47] L. Laliberte, M.L. Fennell, G. Papandonatos, The relationship of membership in research networks to compliance with treatment guidelines for early-stage breast cancer, Med. Care 43 (5) (2005) 471–479, https://doi.org/10.1097/01.mlr.0000160416.66188.f5.
[48] I.J. Hubbard, D. Harris, M.F. Kilkenny, S.G. Faux, M.R. Pollack, D.A. Cadilhac, Adherence to clinical guidelines improves patient outcomes in Australian audit of stroke rehabilitation practice, Arch. Phys. Medicine Rehabilit. 6 (2012) 965–971, https://doi.org/10.1016/j.apmr.2012.01.011.
[49] R. Chagpar, Y. Xing, Y.J. Chiang, B.W. Feig, G.J. Chang, Y.N. You, J.N. Cormier, Adherence to stage-specific treatment guidelines for patients with colon cancer, J. Clin. Oncol. 30 (9) (2012) 972–979, https://doi.org/10.1200/JCO.2011.39.6937.
[50] R. Gatta, M. Vallati, C. Fernandez-Llatas, A. Martinez-Millana, S. Orini, L. Sacchi, J. Lenkowicz, M. Marcos, J. Munoz-Gama, M. Cuendet, B. de Bari, L. Marco-Ruiz, A. Stefanini, M. Castellano, Clinical Guidelines: A Crossroad of Many Research Areas, in: Challenges and Opportunities in Process Mining for Healthcare, Lecture Notes in Business Information Processing 362 LNBIP, 2019, pp. 545–556, https://doi.org/10.1007/978-3-030-37453-2_44.
[51] P.C. Dykes, Practice guidelines and measurement: state-of-the-science, Nursing Outlook 51 (2) (2003) 65–69.
[52] J.J. Mcmurray, S. Adamopoulos, S.D. Anker, A. Auricchio, M. Böhm, K. Dickstein, V. Falk, G. Filippatos, C. Fonseca, M.A. Gomez-Sanchez, T. Jaarsma, L. Kober, G.Y. Lip, A.P. Maggioni, A. Parkhomenko, B.M. Pieske, B.A. Popescu, P.K. Ronnevik, F. H. Rutten, J. Schwitter, P. Seferovic, J. Stepinska, P.T. Trindade, A.A. Voors, F. Zannad, A. Zeiher, J.J. Bax, H. Baumgartner, C. Ceconi, V. Dean, C. Deaton, R. Fagard, C. Funck-Brentano, D. Hasdai, A. Hoes, P. Kirchhof, J. Knuuti, P. Kolh, T. Mcdonagh, C. Moulin, Ž. Reiner, U. Sechtem, P.A. Sirnes, M. Tendera, A. Torbicki, A. Vahanian, S. Windecker, L.A. Bonet, P. Avraamides, H.A. Ben Lamin, M. Brignole, A. Coca, P. Cowburn, H. Dargie, P. Elliott, F.A. Flachskampf, G.F. Guida, S. Hardman, B. Iung, B. Merkely, C. Mueller, J.N. Nanas, O.W. Nielsen, S. Orn, J.T. Parissis, P. Ponikowski, ESC Guidelines for the diagnosis and treatment of acute

and chronic heart failure 2012, Eur. J. Heart Failure 14 (8) (2012) 803–869. https://doi.org/10.1093/eurjhf/hfs105.

[53] R. Gatta, M. Vallati, C. Fernandez-Llatas, A. Martinez-Millana, S. Orini, L. Sacchi, J. Lenkowicz, M. Marcos, J. Munoz-Gama, M.A. Cuendet, B. de Bari, L. Marco-Ruiz, A. Stefanini, Z. Valero-Ramon, O. Michielin, T. Lapinskas, A. Montvila, N. Martin, E. Tavazzi, M. Castellano, What role can process mining play in recurrent clinical guidelines issues? A position paper, Int. J. Environ. Res. Public Health 17 (18) (2020) 1–19, https://doi.org/10.3390/ijerph17186616.

[54] D. Capurro, M. Yetisgen, E. Eaton, R. Black, P. Tarczy-Hornoch, Availability of Structured and Unstructured Clinical Data for Comparative Effectiveness Research and Quality Improvement: A Multi-Site Assessment, eGEMs (Generating Evidence & Methods to improve patient outcomes) 2 (1) (2014) 11, https://doi.org/10.13063/2327-9214.1079.