Journal of Biomedical Informatics 52 (2014) 354–363

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

Creating hospital-specific customized clinical pathways by applying semantic reasoning to clinical data

Hua-qiong Wang [a], Tian-shu Zhou [a], Li-li Tian [b], Yang-ming Qian [b], Jing-song Li [a,*]

[a] EMR and Intelligent Expert System Engineering Research Center, Key Laboratory of Biomedical Engineering, Ministry of Education, College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, China
[b] Navy General Hospital, Beijing, China

ARTICLE INFO

Article history:
Received 29 December 2013
Accepted 23 July 2014
Available online 8 August 2014

Keywords:
Clinical pathway
Semantic transformation
Semantic reasoning
Clinical decision support

ABSTRACT

Objective: Clinical pathways (CPs) are widely studied methods to standardize clinical intervention and improve medical quality. However, standard care plans defined in current CPs are too general to execute in a practical healthcare environment. The purpose of this study was to create hospital-specific personalized CPs by explicitly expressing and replenishing the general knowledge of CPs by applying semantic analysis and reasoning to historical clinical data.

Methods: A semantic data model was constructed to semantically store clinical data. After querying semantic clinical data, treatment procedures were extracted. Four properties were self-defined for local ontology construction and semantic transformation, and three Jena rules were proposed to achieve error correction and pathway order recognition. Semantic reasoning was utilized to establish the relationship between data orders and pathway orders.

Results: A clinical pathway for deviated nasal septum was used as an example to illustrate how to combine standard care plans and practical treatment procedures. A group of 224 patients with 11,473 orders was transformed to a semantic data model, which was stored in RDF format. Long term order processing and error correction made the treatment procedures more consistent with clinical practice. The percentage of each pathway order with different probabilities was calculated to declare the commonality between the standard care plans and practical treatment procedures. Detailed treatment procedures with pathway orders, deduced pathway orders, and orders with probability greater than 80% were provided to efficiently customize the CPs.

Conclusions: This study contributes to the practical application of pathway specifications recommended by the Ministry of Health of China and provides a generic framework for the hospital-specific customization of standard care plans defined by CPs or clinical guidelines.

1. Introduction

Clinical pathways (CPs), defined as structured multidisciplinary care plans [1], have been widely implemented as methods to standardize clinical intervention and potentially improve medical quality [2–5]. Extensive studies have evaluated the effectiveness of CPs for various diseases [6–10], though little information exists describing the use of CPs. Recent clinical practice in China has proven that the utility rate of CPs is unsatisfactory [11,12]. The European Pathway Association (EPA) performed an international survey on the practical implementation of CPs in 23 countries between 2004 and 2005 [11]. According to the statistical results reported by the EPA, in only three of the countries evaluated were approximately 21–40% of patients under pathway-based treatment, while in China and other participating countries, the number was less than 15%. Tao et al. [12] summarized the use of CPs based on 1051 literatures about CPs published in Chinese journals between 2003 and 2009. As reported, there were 162 hospitals in China that implemented electronic or paper CPs, accounting for only 0.82% of the total hospitals. And in 162 hospitals, 82.7% of the hospitals implemented CPs for less than 10 diseases.

There are two main reasons limiting the practicability of CPs. First, the standard care plans pre-determined in CPs are not universally adaptable for different patients in different hospitals. In addition to patient characteristics, which have been a key consideration in creating personalized care plans via pathway customization, hospital characteristics are also important for generating personalized CPs. Merging the treatment experience of current hospitals

* Corresponding author. Address: College of Biomedical Engineering and Instrument Science, Zhejiang University, 38 Zheda Road, Hangzhou 310027, China.
E-mail address: ljs@zju.edu.cn (J.-s. Li).

http://dx.doi.org/10.1016/j.jbi.2014.07.017
1532-0464/

into patient-specific CPs is beneficial in improving the practicability of CPs. Second, care plans defined in current CPs are too general to execute. Standard care plans in the clinical pathway (CP) specifications recommended by the Ministry of Health of China cannot be directly implemented in hospitals due to the general description of medical interventions. For example, anesthesia, which is usually essential before a surgical operation in a practical clinical environment, is frequently ignored in standardized CPs. Additionally, antiseptic, anticoagulant, and anti-infective agents are commonly defined in CPs without detailed names and dosages. Addressing complicated clinical details has become a challenge for implementing CPs.

Taking into account the above two factors, this study proposes a data-driven, decision-making methodology to improve CP customization by applying semantic analysis and reasoning to historical clinical data. Through the analysis of historical clinical data in the hospital, we can generate disease-specific treatment procedures that are frequently used by physicians. These treatment procedures detail medical interventions of standard CPs and are helpful in creating hospital-specific customized implementation strategies.

However, the quality of medical data may be indefinite due to data inconsistency, incompleteness, and ambiguities. Pretreatment is indispensable before completing data statistics and analysis. Semantic web technologies provide a novel approach to address the problem of data complexity [13–17]. This study analyzes and processes the historical clinical data by semantic transformation and reasoning. Common treatment procedures are extracted from clinical data via probability and statistics. After calculating the probabilities of standard CP procedures that appear in historical data, we discuss the process of CP replenishing and detailing, which are realized with the guidance of the historical treatment experience from historical data.

In this paper, complete treatment procedures with pathway orders (standard interventions defined by CPs), deduced pathway orders (detailed interventions generated via semantic reasoning), and orders with high probability (supplementary interventions obtained from clinical data) are provided to efficiently customize CPs. Hospital-specific customized CPs are created by applying semantic reasoning to clinical data, which is beneficial for improving the practicability of standard CPs in hospitals.

2. Related work

2.1. Pathway customization

Numerous published studies have proposed methods of customizing CPs, most of which generate patient-specific care plans by individually analyzing patient characteristics based on patient information in the electronic healthcare records (EHR) [18–20]. The EHR, as integration of subset records of patient encounters in various care delivery settings, contains complete patient health information ranging from patient demographics and past medical history to laboratory and radiology data generated from medical devices and enables different organizations to easily share patients' medical information [21]. Wang et al. [20] proposed interaction between knowledge-based CPs and semantic EHR to improve the practicality of CPs. Serbanati et al. [22] proposed a virtually centralized, longitudinal patient record called the virtual healthcare record, which is a patient-centric model with a complete and authoritative representation of patient data for regional sharing. The virtual healthcare record was designed to have a native function to monitor clinical information and support CP customization. Although customized methods of analyzing patient characteristics are rational and valid for improving CP practicability, EHR systems are not widely implemented in most countries. Consequently, information about patient characteristics is difficult to capture, which constrains the generation of patient-specific CPs.

Compared to the EHR, the electronic medical record (EMR) is adopted by healthcare practitioners to document, monitor, and manage patients' medical process within a care delivery organization [21]. With the rapid increase in the adoption and implementation of EMR systems, hospital characteristics are comparatively convenient to capture from clinical data recorded by EMR systems.

2.2. Semantic transformation and mapping

A platform for accessing relational databases as virtual RDF graphs (D2RQ) [23] is commonly used to conveniently access relational databases as semantic RDF graphs [24–26]. However, it offers RDF-based access to the content of relational databases without replicating it into an RDF store, which confines the content of relational databases to semantic reasoning. CEM-OWL is a tool proposed for semantic transformation that provides authoring, reasoning, and querying tools [27]. It transforms the data stored in XML format in EHR systems to semantic data in OWL format. However, in Chinese clinical practice, the medical data recorded by EMR systems and stored in a database like Oracle, are mainly in the form of two-dimensional tables rather than in XML format, which limits the applicability of CEM-OWL. Therefore, instead of using the D2RQ platform and CEM-OWL, this study constructs and customizes a semantic data model according to data structure and practical requirements.

Some studies have adopted semantic similarity matching to address semantic mapping [28–30]. Similarity calculation could achieve more intelligent semantic mapping; however, the mapping process is complex and time-consuming, which will affect the performance of real-time decision support. To achieve data mapping, this study adopts the method of building local ontologies. A separate property is created for each hospital object recording hospital-specific clinical terms. On one hand, because the CP specifications recommended by the Ministry of Health of China describe the treatment procedures generally with "antibacterials" or "chest X-ray" rather than detailed drug names or instrument model numbers, the number of pathway orders for each CP is limited. On the other hand, except for periodic updates for new drugs, materials and instruments, the hospital-specific terms of pathway orders are relatively static, which make the hospital-specific terms easy to establish by the default local ontology. Additionally, even if a new instrument is not updated in local ontologies, the relevant information will be supplemented to treatment procedures from the clinical data if the new instrument has been used frequently. Consequently, the approach of default local ontology could efficiently achieve data mapping and greatly simplify the complexity of semantic reasoning.

3. Material and methods

3.1. Data collection

We performed our study using data from Navy General Hospital in Beijing, China. We compiled statistics on clinical data acquired between August 1, 2010 and July 31, 2011. Over this time period, the hospital had 1,114,693 outpatients, 28,775 inpatients, 163,707 diagnoses, and 2,646,572 executed orders that included laboratory tests, radiation, injections, operations and other order types.

Several types of diagnoses were considered, including: outpatient diagnoses, inpatient diagnoses, key diagnoses, operation complications and other diagnoses. These types of diagnoses were

mutually exclusive and were recorded by physicians in different periods during patients' treatment processes. In practical clinical environment of Navy General Hospital, physicians make a definite, core diagnosis with the type of key diagnoses when a patient's disease is established. The records with diagnostic type "key diagnoses" are chief concerns of CPs. A clinical pathway for a diagnosis means the standardized care plan for patients with this key diagnosis. In addition, the treatment result of diagnoses is a significant factor to evaluate whether a care plan is rational or not. Therefore, the diagnostic type and the treatment result were chief factors we were concerned about. We arranged the query conditions with the diagnostic type "key diagnosis" and the treatment result "cured" and then counted the number of patients with different diagnoses. By querying the diagnostic records with predetermined conditions, 10,359 key diagnostic records were selected with the treatment result "cured". The 10 diagnoses with the largest number of patients are listed in Table 1 with their average length of stay (LOS) and the corresponding standard deviation (SD).

The statistical average of each diagnostic hospitalization in days and the corresponding SD were calculated to evaluate the complexity of patient treatment procedures. In general, the shorter the average LOS is, the simpler the treatment procedures are for this diagnosis. The smaller the SD of the LOS is, the higher the commonality of the patient treatment procedures is. The diagnosis with greater number of patients and smaller SD of the LOS is preferentially chosen. Therefore, deviated nasal septum diagnosis (ICD-10: J34.2), with the second greatest number of patients and the smallest SD of the LOS, was chosen as a study case to illustrate how to improve CP practicability.

3.2. Model construction

Previous work provides technical guidance for constructing CP model and semantic rules [31]. Four phases of ontology methodology have been proposed to develop CP ontology [20]. Protégé [32] was utilized as the ontology editor tool, and the web ontology language (OWL) [33] was adopted for ontology description. Jena semantic web framework provides a platform for semantic transforming and reasoning and Jena rules are defined to describe clinical procedure rules during CP execution.

We established CP terminologies and relationships according to the CP specifications recommended by the Ministry of Health of China. To realize the semantic mapping from terms to practical clinical data, hospital-specific local ontology is indispensable. The hospital-specific terms, which consist of local ontologies, are captured from dictionaries in the hospital data base.

To complement the connection between global terms and hospital-specific terms, the properties **hasHZTerm** ("HZ" is the abbreviation of Chinese expression of Navy General Hospital) and **hasDrugHZTerm** have been introduced to the CP model. The following OWL ontology fragment declares that **hasHZTerm** with a domain **TermElement** and a range **xsd;string** is a functional datatype property. Here, "functional" means that this property is single valued. **hasDrugHZTerm** with a domain **Injection** and a range **xsd;string** is a common datatype property that allows one or more values. Domains **TermElement** and **Injection** are subclasses of the global term class **OrderElement**. Values of these two properties direct to local term class. In this way, global terms and local terms are interconnected.

Property 1:

```
<owl:DatatypeProperty rdf:ID=''hasHZTerm''>
    <rdf:type rdf:resource=''&owl;FunctionalProperty''/>
    <rdfs:domain rdf:resource=''#TermElement''/>
    <rdfs:range rdf:resource=''&xsd;string''/>
</owl:DatatypeProperty>
```

Property 2:

```
<owl:DatatypeProperty rdf:ID=''hasDrugHZTerm''>
    <rdfs:domain rdf:resource=''#Injection''/>
    <rdfs:range rdf:resource=''&xsd;string''/>
</owl:DatatypeProperty>
```

3.3. Semantic transformation

The hospital-specific clinical data are stored in Oracle, a relational database. To support semantic understanding and reasoning, the relational data model needs to be transformed to a semantic data model that is stored in RDF format. Data transformation is the foundation of semantic reasoning. In this study, the main object of semantic reasoning is order information, the transformational relation of which is shown in Table 2.

Class **OrderFact**, a super class, is introduced to represent the order data. As shown in Table 2, each order record acquired by structured query language (SQL) from the relational data model is transformed to an individual of class **OrderFact**. Fields of order records correspond to the properties of individuals. The whole transforming process is implemented based on the Jena semantic web framework.

In the process of semantic transformation, long term order processing is essential. A long term order, relative to a temporary order, needs to be executed repeatedly according to the predetermined frequency within the pre-determined specific period of time. In the Chinese medical environment, long term orders are very commonly used. Data entry of orders with types of care, diet and injection are regularly in the form of long term orders.

Table 1
10 Diagnoses with the largest number of patients.

Diagnosis name	Patient number	Average LOS	SD
Upper respiratory tract infection	233	5.970	3.918
Deviated nasal septum	224	6.915	1.845
Caesarean	223	5.955	3.097
Age-related cataract	213	11.122	8.394
Pneumonia	185	9.459	5.606
Chronic sinusitis	164	6.988	1.981
Lung infection	158	17.911	14.499
Premature rupture of membranes	142	4.852	2.674
Adenoidal hypertrophy	131	6.061	1.900
Uterine smooth muscle tumor	125	8.584	3.349

Table 2
Transformational relation of order information from the relational model to the semantic model.

Relational data model	Semantic data model		
Field	Property	Domain	Range
PATIENT_ID; VISIT_ID	CP:hasPatientData	CP:OrderFact	CP:Patient
ORDER_TYPE	CP:hasOrderType	CP:OrderFact	xsd;string
ORDER_CODE	CP:hasOrderCode	CP:OrderFact	xsd;string
ORDER_TEXT	CP:hasOrderName	CP:OrderFact	xsd;string
REPEAT_INDICATOR	CP:hasRepeatIndicator	CP:OrderFact	xsd;string
START_DATE_TIME	CP:hasStartDate	CP:OrderFact	xsd;string
STOP_DATE_TIME	CP:hasStopDate	CP:OrderFact	xsd;string
	CP:hasExecuteDay	CP:OrderFact	xsd;string
	CP:hasCPFlag	CP:OrderFact	xsd;string

For example, a patient needs to receive "second-class care" between the first and fifth day of hospitalization. The physician enters a five-day long term order instead of five independent orders from the first day to the fifth day. Thus, simply calculating the record number in the order table will neglect the difference between long term orders and temporary orders, which greatly decreases the validity of the statistics. Therefore, semantic processing of long term orders is required before gathering statistics of orders.

A property **hasExecuteDay** is defined to represent the relative execution date of orders. As depicted in the following OWL ontology fragment, **hasExecuteDay** is not functional. If a long term order has its execution date from the second day to the fourth day, the values of property **hasExecuteDay** would be "2", "3" and "4". Every order, whether temporary or long term, has its relative execution date, which means that property **hasExecuteDay** has at least one value.

Property 3:

```
<owl:DatatypeProperty rdf:ID="hasExecuteDay">
    <rdfs:domain rdf:resource="#OrderFact"/>
    <rdfs:range rdf:resource="&xsd;string"/>
</owl:DatatypeProperty>
```

After semantic transformation, every patient's treatment procedures can be extracted from clinical order data, and every day treatment procedures of all patients are merged and listed by simple protocol and RDF query language (SPARQL) [34]. We count the occurrence of each non-repetitive clinical procedure and obtain the corresponding occurrence probability in all patients. Via calculating and analyzing statistical probability of the semantically transformed semantic order data, treatment procedures with probabilistic characteristics are formed based on the historical clinical practice.

3.4. Semantic reasoning

The occurrence of incorrect order records is inevitable. These incorrect order records can be categorized into two types. Random errors resulting from recording mistakes could be eliminated by filtering out the clinical procedures with small probability. Incorrect data recorded during actual medical procedures could be eliminated by semantic reasoning.

Consider the simultaneous temporary and long term orders as an example. In clinical practice, the time of drug dispensing for long term drug orders is fixed. If the drug dispensing time is exceeded when recording a long term drug order, a temporary drug order is required to complete the clinical action of drug dispensing. In this situation, equivalent temporary and long term drug orders are both recorded on the same day for the same patient, while drug dispensing is executed only once. To address this inconsistency between clinical data and the actual medical procedures, the semantic rule **Rule1** has been proposed.

As described in the following rule fragment, two variables (?order1 and ?order2) are defined as individuals of class **OrderFact**. These two orders have the same patient (?patient), the same relative execution date (?day), and the same order name (?name). However, one (?order1) is a long term order and the other (?order2) is a temporary order. After reasoning with this semantic rule, the execution date (?day) of the long term order (?order1) is deleted.

Rule1:

```
@prefix CP: <http://www.owl-ontologies.com/Ontology1332316381.owl#>.
[ErrorData:
(?order1 CP:hasPatientData ?patient)(?order2 CP:hasPatientData ?patient)
(?order1 CP:hasExecuteDay ?day)(?order2 CP:hasExecuteDay ?day)
(?order1 CP:hasOrderName ?name)(?order2 CP:hasOrderName ?name)
(?order1 CP:hasRepeatIndicator "1")(?order2 CP:hasRepeatIndicator "0")
-> remove(2)]
```

After filtering out the probability and reasoning with semantic rules, the treatment procedures extracted from the clinical data can more objectively reflect practical medical procedures. To compare the treatment procedures from clinical data with the standardized care plans from CP model, a property **hasCPFlag** is defined and two Jena rules (**Rule2** and **Rule3**) are proposed.

As depicted in the following OWL ontology fragment, the property **hasCPFlag** is a functional property with a domain **OrderFact** and a range **xsd;int**. The value of **hasCPFlag** represents the relationship between data orders and pathway orders. If a data order is related to a pathway order, "1" signifies that this data order is directly related to a pathway order, while "2" signifies that this data order is a detailed description of a pathway order.

Property 4:

```
<owl:DatatypeProperty rdf:ID="hasCPFlag ">
    <rdf:type rdf:resource="&owl;FunctionalProperty"/>
    <rdfs:domain rdf:resource="#OrderFact"/>
    <rdfs:range rdf:resource="&xsd;int"/>
</owl:DatatypeProperty>
```

The following Jena rule, **Rule2**, decides what type of data orders are pathway orders. Compare the order name (?name) of a data order (?order) with the value (?name) of **hasHZTerm**, which is predefined in local CP ontology. If the order event (?order_event) indirectly has the property value (?name) that is equal to the name of a pathway order of the CP, **SeptumdeviationCP**, then the data order (?order) is a pathway order.

Rule2:

```
[BasicCPOrder:
(?order CP:hasOrderName ?name)
(CP:SeptumdeviationCP CP:hasOrderEvent ?order_event)
(?order_event CP:hasRelatedTerm ?order_term)
(?order_term CP:hasHZTerm ?name)
-> (?order CP:hasCPFlag 1)]
```

In standardized care plans of CP specifications, general clinical actions such as treatment with antiseptic, anticoagulant, and anti-infective agents are most likely proposed without the explicit drug name, dosage, or frequency. According to the standard care plan, patients who enter the CP **SeptumdeviationCP** have to receive antiseptic therapy. However, the standard care plan does not explicitly propose what antimicrobial drugs to prescribe or

how to implement it. Compared to the standard care plan, clinical data include more details. Complete and detailed drug information is included in the data records. To address the semantic inconsistency caused by different description granularity, the following Jena rule, **Rule3**, is proposed. These order records, which detail the CP procedures, are considered as deduced pathway orders with **hasCPFlag** value "2".

Rule3:

```
[DrugCPOrder:(?order CP:hasOrderName ?name)
(CP:SeptumdeviationCP CP:hasOrderEvent ?order_event)
(?order_event CP:hasRelatedTerm ?order_term)
(?order_term CP:hasDrugHZTerm ?name)
-> (?order CP:hasCPFlag 2)]
```

4. Results

4.1. Pathway model

In previous work, we designed four super classes, **ClinicalPathway**, **CPElementBase**, **CPEventModel**, and **Patient** in CP ontology and summarized 84 CP-related classes and 98 individuals [20]. We analyzed the CP specification of deviated nasal septum, established 28 key orders from 6 hospital days, and then replenished 1 individual of class **ClinicalPathways** and 28 corresponding individuals of class **OrderEvent**.

As depicted in Fig. 1, the individual **SeptumDeviationCP** is used for deviated nasal septum with an average LOS of 6 days. Three order events of the CP for deviated nasal septum are listed with their related order terms and execution dates. Every order term is assigned a value of the property **hasHZTerm**. The order term **AntisepticDrug**, which is a subclass of **Injection**, has several values of the property **hasDrugHZTerm**. The values of **hasHZTerm** and **hasDrugHZTerm** are responsible for recognizing whether a data order is a pathway order or not. The pathway orders in the standardized care plan from the CP model are listed according to the execution date as shown in Fig. 1.

4.2. Semantic data model

By semantic transformation, 224 individuals of class **Patient** and 11,473 individuals of class **OrderFact** are imported. Each individual of class **Patient** has values of properties **hasPaitentId** and **hasVisitId**. As shown in Fig. 2, each individual of class **OrderFact** includes the following 9 properties: **hasPatientData**, **hasOrderType**, **hasOrderCode**, **hasOrderName**, **hasRepeatIndication**, **hasStartDate**, **hasStopDate**, **hasExecuteDay**, and **hasCPFlag**. With the exception of the self-defined properties **hasExecuteDay** and **hasCPFlag**, values of properties are automatically imported from clinical data records by semantic transformation. The value of the property **hasExecuteDay** is obtained in the phase of long term order processing and further determined by error correction. The value of the property **hasCPFlag** is determined by semantic reasoning in the phase of pathway order recognition.

4.3. Long term order processing

In the process of semantic transformation, we mainly processed the relative execution date of long term orders. The property **hasExecuteDay** with at least one value was specially added. Values of the property **hasExecuteDay** recorded the relative execution date of long term orders. We queried the semantic data according to execution date by SPARQL and calculated the probability of each order. To make comparing the results concise and easy to understand, orders with probabilities of more than 30% were chosen. Fig. 3 depicts the results of long-term order processing.

According to the results shown in Fig. 3, the differences between treatment procedures before and after long term order processing are mainly reflected in the orders since the third day. In original treatment procedures, there were only 5 orders on the third day and 1 order on the fourth day. In treatment procedures after long term order processing, 36 orders existed after the third day. Take the nursing orders on the third day as an example. "First-class care", "second-class care" and "third-class care" are nursing orders with different nurse levels. As shown in Fig. 3, only 58% of patients received "second-class care" in original procedures. In reality, 71% of patients received "first-class care", 68% of patients received "second-class care", and 36% of patients received

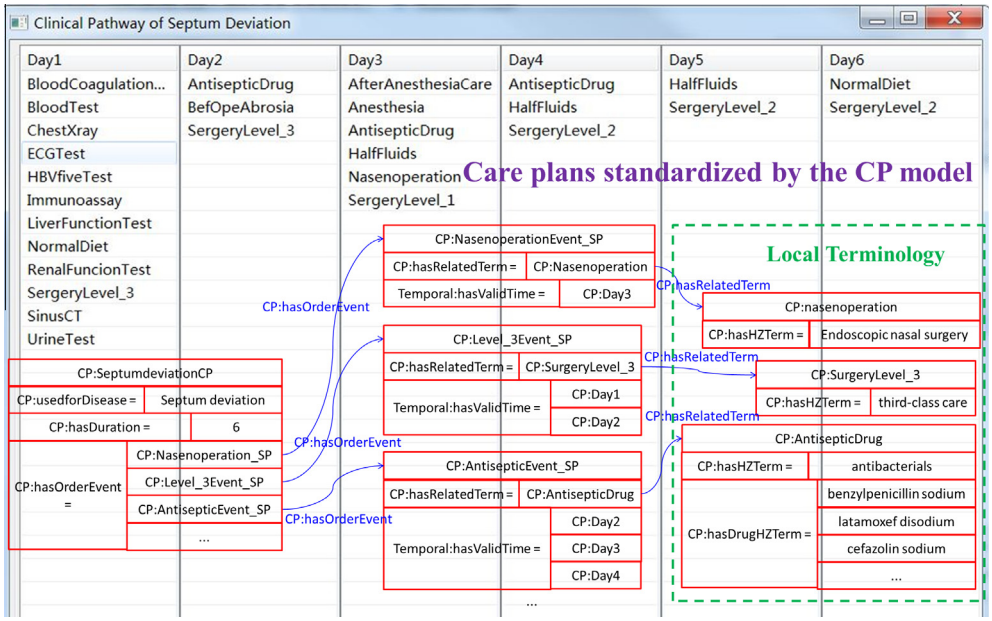

Fig. 1. Care plans standardized by the CP model.

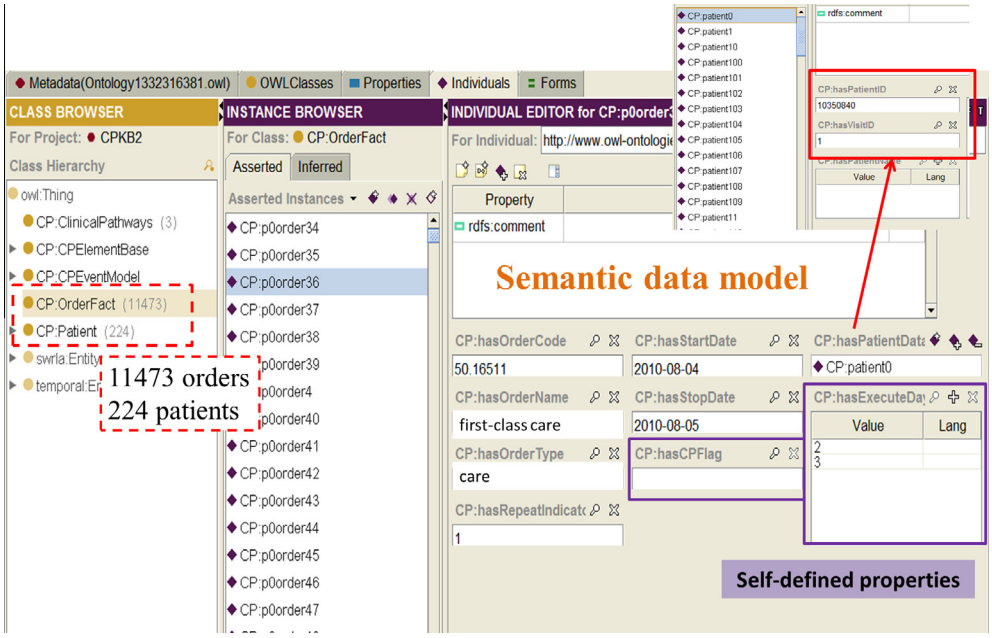

Fig. 2. Semantic data model after semantic transformation.

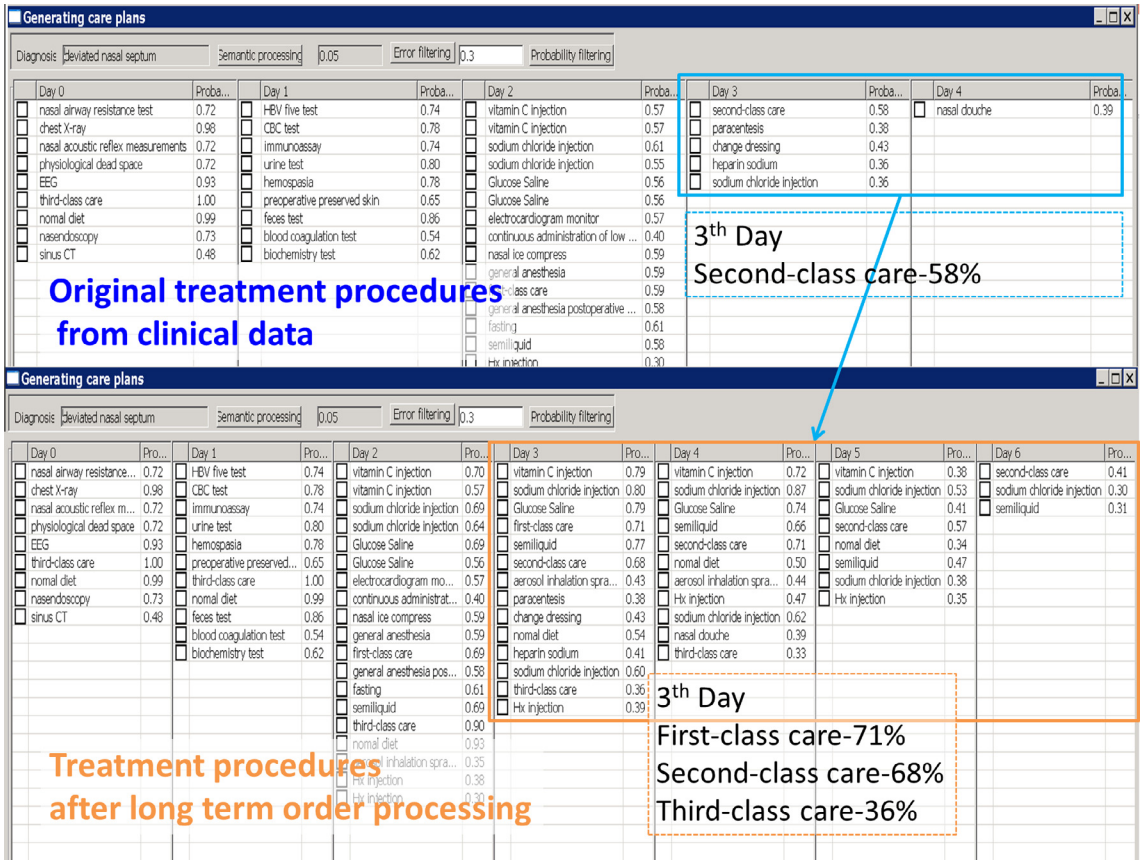

Fig. 3. Results of long term order processing.

"third-class care" according to the results of long term order processing. Orders regarding types of nursing, diet, and injection are commonly entered in the form of long term orders. Thus, the effect of long term order processing is significant for these types of orders.

4.4. Error correction

4.4.1. Random error elimination

The number of clinical data records generated by EMR systems is enormous, which leads to a lot of reliance on erroneous data. The

random error caused by mistaken recording could be reduced by filtering out the orders with probabilities less than the pre-determined minimum, which could be modified via the user interface by physicians. The orders less than the pre-determined minimum are considered as events of small probability. In addition to random errors, orders with small probability included the rare but correct orders that were recorded by physicians for special patients. However, these orders with small probability provided little guidance for ordinary patients. With experts' suggestions, 5% was determined as the default minimum. Based on the statistical results, the number of orders that were considered events of small probability was 2370. After error filtering, the number of orders decreased from 11,473 to 9103.

4.4.2. Procedural error elimination

Rule1 has been proposed to eliminate the repetition of long term and temporary orders. The reasoning results of executing the Jena rule **Rule1** are shown in Fig. 4. We considered the orders in the second day as an example. There exist reduplicate injection orders for injections such as vitamin C (70%, 57%), sodium chloride (69%, 64%), and glucose saline (69%, 54%) in preoperative treatment procedures. After reasoning, we can see that the recurrence of the above 3 orders is removed.

Long term order processing and reasoning with **Rule1** make the treatment procedures from clinical data more consistent with clinical practice. The effectiveness of data is strengthened, which makes the comparison of data orders and standardized pathway orders in the CP model more meaningful.

4.5. Pathway order recognition and perfection

Jena rules **Rule2** and **Rule3** have been proposed to distinguish a pathway order from a data order. As depicted in Fig. 5, different item backgrounds in each child table illustrate the different reasoning results after executing **Rule2** and **Rule3**. Orders with a blue background are pathway orders, while orders with a red background or a symbol "*" are deduced pathway orders, which specify and detail the general knowledge of pathway orders in the CP model. Process mining techniques have been utilized to mine the sequence of clinical orders. In Fig. 5, all orders in one day are displayed in time sequence on the basis of the time property recorded in clinical data. The direct reasoning results are in Chinese to keep them in accordance with the EMR systems. We have translated the orders into English to present the reasoning results in a user-friendly way as shown in Fig. 5.

For the general actions predefined in the CP model, the clinical data provide the detailed content to support the CP execution. According to the statistical results of clinical historical data, different antibacterial drugs have been utilized for different patients. Cefazolin sodium, latamoxef disodium, cefotiam hydrochloride and benzylpenicillin sodium are common antibacterial drugs for patients with a deviated nasal septum. As a detailed description of the pathway order "antibacterials", Fig. 6 presents the probabilities of these four antibacterial drugs being prescribed between the first day and the third day.

As depicted in Fig. 5, we can obviously discover that not all pathway orders are executed with high probability. Pathway orders such as "third-class care" are 100% for all patients on the first and second day, while pathway orders like "HBV five test" are only 1% implemented on the first day. The number of pathway orders with execution probabilities greater than some pre-determined value is easy to count. Probability of pathway orders refers to the probability of pathway orders that appear in historical data, while percentage of pathway orders is defined as the percentage of pathway orders with some probability in all pathway orders. As shown in Fig. 5, two pathway orders named "third-class care" were executed with a probability of 100% for all patients on the first and second day, and one pathway order "normal diet" was executed with a probability of 99%. Then the number of pathway orders with the probability greater than or equal to 99% is 3. As

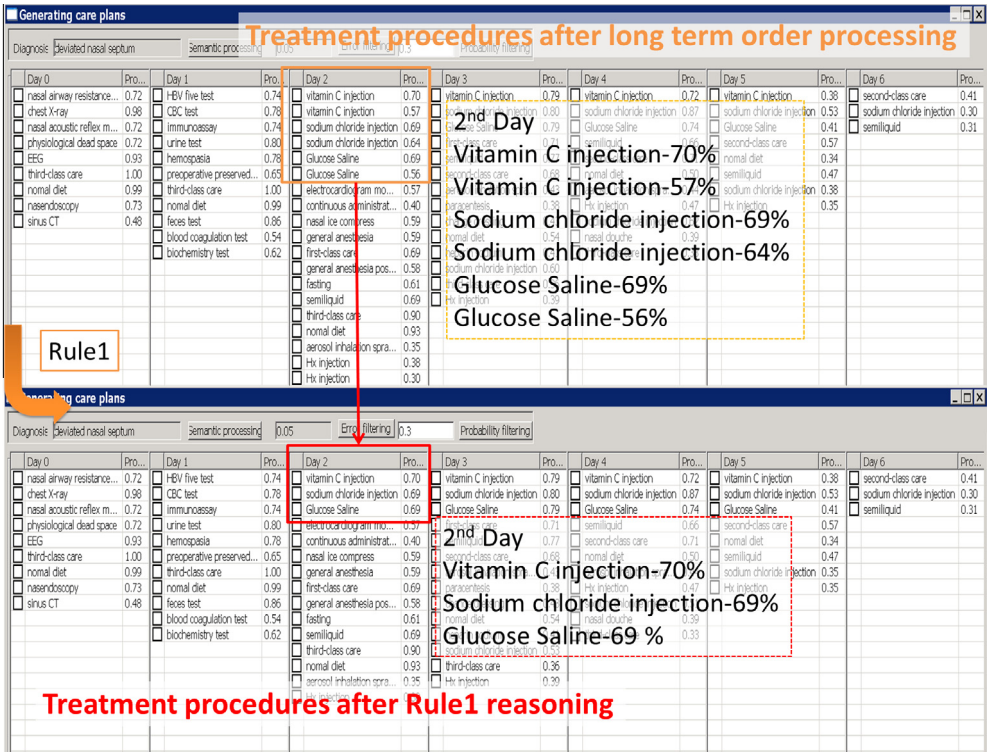

Fig. 4. Reasoning results of executing **Rule1**.

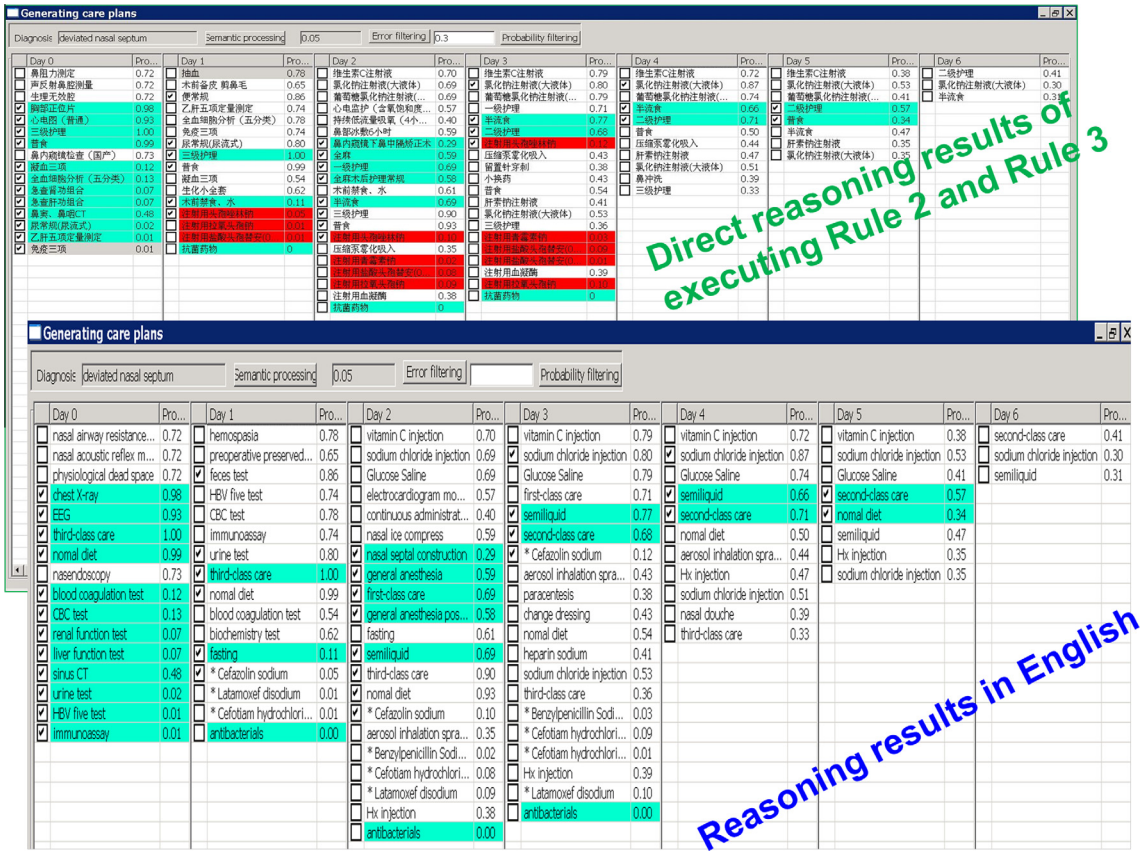

Fig. 5. Reasoning results of executing **Rule2** and **Rule3**.

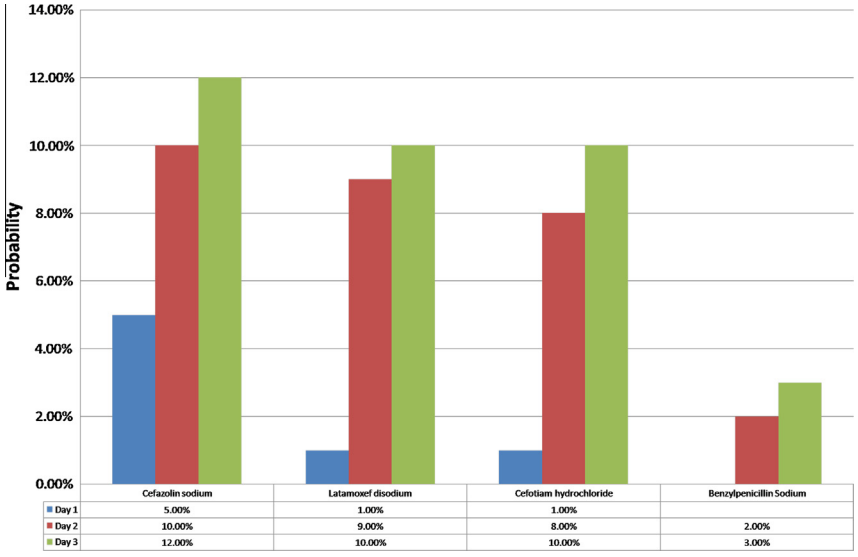

Fig. 6. A detailed description of the pathway order "antibacterials".

the number of pathway orders for deviated nasal septum is 28, the percentage of pathway orders with the probability greater than or equal to 99% is 10.71% (3/28). We calculated the percentage of each pathway order with different probabilities and plotted the practical statistical data. The plot results are shown in Fig. 7. By curve fitting, the relationship between the percentage of pathway orders (y) and the probability (x) is described with the following formula.

$$y = -0.821x + 0.908; \quad k = 0.821,\ y0 = 0.908 \tag{1}$$

As the value of k becomes smaller, the percentage of pathway orders (y) will be larger for the same probability (x), which means that the standardized care plans defined by the CP model better resemble the treatment procedures executed in the current hospital and represent more adaptability to the current hospital.

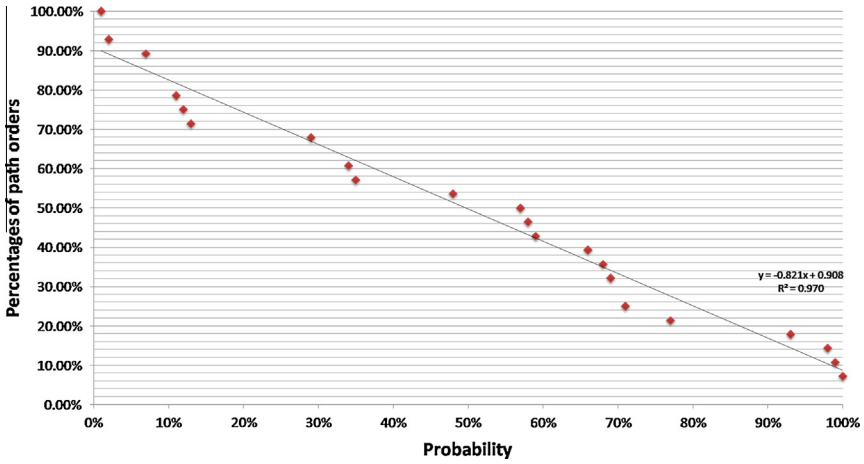

Fig. 7. Percentage of each pathway order with different probabilities.

Based on the above analysis, 3 types of orders are automatically selected to improve the practicability of CPs: (1) pathway orders, which are directly defined in the CP model; (2) deduced pathway orders with the highest probabilities; and (3) orders with probabilities greater than the datum probability (i.e., 80%), set by physicians. To improve the practicability of CPs in the current hospital, the pathway orders directly or indirectly defined in the CP model and the common orders with high execution probabilities are combined to effectively reduce the inconsistencies between the standard care plans and the practical treatment procedures. This will provide much more complete decision support for clinical physicians to develop more efficient and intelligent clinical procedures.

5. Discussion

In the result section, the CP for deviated nasal septum is taken as an example to illustrate how to create hospital-specific customized CPs via semantic reasoning. The CP for deviated nasal septum is selected via adequate consideration of the number of patients, average LOS and corresponding SD. However, as the order structure for all diagnoses defined in EMR systems is identical, similar processes of semantic transformation and semantic reasoning also apply to other diagnosis.

According to the reasoning results shown in Fig. 7, the percentage of pathway orders with the probability greater than 80% is 25%, which indicates that only 25% of pathway orders in standard care plans pre-determined in CPs were executed in practical treatment procedures for more than 80% of patients with deviated nasal septum. As shown in Fig. 5, there are 7 supplemented clinical orders with probability greater than 80%, which indicates that 7 orders were executed in practical treatment procedures for 80% of patients with deviated nasal septum. In general, the treatment procedures executed in the current hospital for deviated nasal septum are not consistent with the standardized care plans defined by the CP specifications issued by the authority.

Treatment procedures based on clinical data and standard care plans defined by the CP model are not totally consistent, which is mainly due to the following four factors. First, orders such as "second-class or third-class care" and "antibacterials" in the standard care plans defined by CPs are too ambiguous and general to be implemented, while the clinical data are recorded with explicit and complete information. Second, practical treatment procedures of one hospital differ from those of other hospitals because of differences in medical resources and the healthcare environment. For example, a hospital without CT equipment cannot execute the pathway order "chest CT". Third, the treatment procedures differ from each other for different patients. Finally, non-standard medical actions may exist, which induce different treatment procedures.

With the exception of the incorrect data recorded by non-standard medical actions, which have been removed by probability filtering, the other three factors are perfectly representative of standard care plans. The detailed information of treatment procedures is beneficial to explicitly express the ambiguous and general knowledge in CP specifications. The historical treatment procedures based on clinical data provide rich data support for every medical term in the CP model, formulating complete, detailed decision support on the basis of the current hospital's long-standing practice, and help the physicians to efficiently form individualized treatment procedures. The individualized treatment procedures based on the standard CP and the historical treatment procedures are recorded in the form of CP data, which will further optimize the CP model and assist with the development of CP specifications.

The clinical data that we analyzed is mainly about diagnoses and orders. In these two data tables, the information associated with the patient serves as unique patient identifiers, including properties PATIENT_ID and VISIT_ID. The unique patient identifiers, which are automatically created by healthcare information systems, cannot be used to deduce sensible patient-specific information. Therefore, concerns about privacy issues associated with the data analysis in this study are alleviated.

The limitation in this study is that the hospital-specific customized treatment procedures generated by standard care plans and clinical data are not adaptable to all hospitals. In this study, orders with high execution probability may be too hospital-specific to apply to other hospitals and not totally correlated to correct practice. However, the procedures are instrumental in the practical implementation of standard CPs and in improving the practicability of CPs in the current hospital. Pathway orders and deduced pathway orders are priority selection to standardize clinical treatment and orders with high execution probability are provided to support physicians' decision-making processes, remarkably improving efficiency. Physicians are the final decision makers who can modify the customized treatment procedures to meet practical clinical environments. Besides, as the CP specifications issued by the Ministry of Health of China list pathway orders without time specifications in one day, the comparison between pathway orders and clinical orders is based on individual orders, without considering the relationships between orders. Lack of quantitative evaluation of the system is another limitation of this study. This system was designed via the cooperation with Navy General Hospital and was developed under the guidance of clinical

experts. Some legal and economic issues restricted practical implementation of the system in hospitals. However, this study provided feasible technical solutions for the hospital-specific customization of standard care plans defined by CPs.

Relative to CPs with essential steps in the care of patients, clinical guidelines provide recommendations for best practice yet do not provide implementation details [14]. Recommendations in guidelines also need detailing and replenishing for the electronic implementation. Therefore, the dynamic customization processes of clinical guidelines are common to CPs. The information from clinical data enables clinical guidelines to provide comprehensive and detailed decision support to physicians.

6. Conclusions

The increasing amount of data digitally collected in healthcare systems provides great potential for extracting useful knowledge to improve medical services. The combination with clinical data recorded by EMR systems makes the standard care plans defined by CPs much more explicit and detailed. By explicitly expressing the general knowledge of CPs, more detailed and complete decision support is provided for clinical physicians to develop more efficient and intelligent clinical procedures. On the basis of CP knowledge and hospital-specific clinical data, the practicability of standard CPs in current hospitals will be effectively improved. This study contributes to the practical application of CP specifications recommended by the Ministry of Health of China and provides a generic framework for the hospital-specific customization of standard care plans defined by CPs or clinical guidelines.

Acknowledgments

This work was supported by the National Natural Science Foundation (Grant No. 61173127), National High-tech R&D Program (No. 2013AA041201) and Zhejiang University Top Disciplinary Partnership Program.

References

[1] Campbell H, Hotchkiss R, Bradshaw N, et al. Integrated care pathways. BMJ 1998;316:133–7.
[2] Wakamiya S, Yamauchi K. What are the standard functions of electronic clinical pathways? Int J Med Inform 2009;78(8):543–50.
[3] Rotter T, Kinsman L, James E, Machotta A, Willis J, Snow P, et al. The effects of clinical pathways on professional practice, patient outcomes, length of stay, and hospital costs: Cochrane systematic review and meta-analysis. Eval Health Prof 2012;35(1):3–27.
[4] Vanhaecht K, Ovretveit J, Elliott MJ, Sermeus W, Ellershaw J, Panella M. Have we drawn the wrong conclusions about the value of care pathways? Is a Cochrane review appropriate? Eval Health Prof 2012;35(1):28–42.
[5] Deneckere S, Euwema M, Van Herck P, Lodewijckx C, Panella M, Sermeus W, et al. Care pathways lead to better teamwork: results of a systematic review. Soc Sci Med 2012;75(2):264–8.
[6] Takegami K, Kawaguchi Y, Nakayama H, Kubota Y, Nagawa H. Impact of a clinical pathway and standardization of treatment for acute appendicitis. Surg Today 2003;33(5):336–41.
[7] Sheffield KM, Ramos KE, Djukom CD, Jimenez CJ, Mileski WJ, Kimbrough TD, et al. Implementation of a critical pathway for complicated gallstone disease: translation of population-based data into clinical practice. J Am Coll Surg 2011;212(5):835–43.
[8] Chevalley T, Chevalley T, Hoffmeyer P, Bonjour JP, Rizzoli R. An osteoporosis clinical pathway for the medical management of patients with low-trauma fracture. Osteoporos Int 2002;13(6):450–5.
[9] Kulkarni RP, Ituarte PHG, Gunderson D, Yeh MW. Clinical pathways improve hospital resource use in endocrine surgery. J Am Coll Surg 2011;212(1):35–41.
[10] Prestom S, Markar S, Baker C, Soon Y, Singh S, Low D. Impact of a multidisciplinary standardized clinical pathway on perioperative outcomes in patients with oesopageal cancer. Br J Surg 2013;100(1):105–12.
[11] Vanhaecht K, Bollmann M, Bower K, Gallagher C, Gardini A, Guezo J, et al. Prevalence and use of clinical pathways in 23 countries: an international survey by the European Pathway Association. Int J Care Pathway 2006;10(1):28–34.
[12] Tao H, Liu P, Liang J, Ke X, Guo S, Qu H. Survey and analysis of the hospitals implementing clinical pathway. Chin Hosp Manage 2010;30(2):28–30.
[13] Dang J, Hedayati A, Hampel K, Toklu C. An ontological knowledge framework for adaptive medical workflow. J Biomed Inform 2008;41(5):829–36.
[14] Gooch P, Roudsari A. Computerization of workflows, guidelines, and care pathways: a review of implementation challenges for process-oriented health information systems. J Am Med Inform Assoc 2011;18(6):738–48.
[15] Argüello Casteleiro M, Des J, Prieto MJF, Perez R, Paniagua H. Executing medical guidelines on the web: towards next generation healthcare. Knowl-Based Syst 2009;22(7):545–51.
[16] Patel C, Gomadam K, Khan S, Garg V. TrialX: using semantic technologies to match patients to relevant clinical trials based on their Personal Health Records. Web Semant: Sci Serv Agent World Wide Web 2010;8(4):342–7.
[17] Ye Y, Jiang Z, Diao X, Yang D, Du G. An ontology-based hierarchical semantic modeling approach to clinical pathway workflows. Comput Biol Med 2009;39(8):722–32.
[18] Blaser R, Schnabel M, Biber C, Bäumlein M, Heger O, Beyer M, et al. Improving pathway compliance and clinician performance by using information technology. Int J Med Inform 2007;76(2–3):151–6.
[19] González-Ferrer A, ten Teije A, Fdez-Olivares J, Milian K. Automated generation of patient-tailored electronic care pathways by translating computer-interpretable guidelines into hierarchical task networks. Artif Intell Med 2013;57(2):91–109.
[20] Wang HQ, Li JS, Zhang YF, Suzuki M, Araki K. Creating personalised clinical pathways by semantic interoperability with electronic health records. Artif Intell Med 2013;58(2):81–9.
[21] Garets D, Davis M. Electronic medical records vs electronic health records: yes, there is a difference: a HIMSS analytics white paper; 2006. <http://www.himssanalytics.org/docs/WP_EMR_EHR.pdf>.
[22] Serbanati LD, Ricci FL, Mercurio G, Vasilateanu A. Steps towards a digital health ecosystem. J Biomed Inform 2011;44(4):621–36.
[23] D2RQ Accessing Relational Databases as Virtual RDF Graphs. <http://d2rq.org/>.
[24] Kama AA, Primadhanty A, Choquet R, Teodoro D, Enders F, Duclos C, et al. Data Definition Ontology for clinical data integration and querying. Stud Health Technol Inform 2012;180:38–42.
[25] Zappa A, Splendiani A, Romano P. Towards linked open gene mutations data. BMC Bioinform 2012;13(Suppl 4):S7.
[26] Kostic P, Novicic D, Blesic S, Vasiljevic Z, Milanovic JG, Pavlovic S, et al. Semantic search engine as tool for clinical decision support in register for acute coronary syndrome. Telfor J 2011;3(1):66–71.
[27] Tao C, Jiang G, Oniki TA, Freimuth RR, Zhu Q, Sharma D, et al. A semantic-web oriented representation of the clinical element model for secondary use of electronic health records data. J Am Med Inform Assoc 2013;20(3):554–62.
[28] Batet M, Sánchez D, Valls A. An ontology-based measure to compute semantic similarity in biomedicine. J Biomed Inform 2011;44(1):118–25.
[29] Köhler S, Schulz MH, Krawitz P, Bauer S, Dölken S, Ott CE, et al. Clinical diagnostics in human genetics with semantic similarity searches in ontologies. Am J Hum Genet 2009;85(4):457–64.
[30] Mabotuwana T, Lee MC, Cohen-Solal EV. An ontology-based similarity measure for biomedical data – application to radiology reports. J Biomed Inform 2013;46(5):857–68.
[31] Hu Z, Li JS, Zhou TS, Yu HY, Suzuki M, Araki K. Ontology-based clinical pathways with semantic rules. J Med Syst 2012;36(4):2203–12.
[32] The Protégé Ontology Editor. <http://protege.stanford.edu/>.
[33] Web Ontology Language (OWL). <http://www.w3.org/2004/OWL/>.
[34] SPARQL Query Language for RDF. <http://www.w3.org/TR/rdf-sparql-query/>.