FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

Comparison of Treatment Plan Models

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieurin

in

Biomedical Engineering

by

Michaela Lehner BSc
Registration Number 0825832

to the Faculty of Informatics
at the Vienna University of Technology

Advisor: Ao.Univ.Prof. Mag. Dr. Horst Eidenberger
Assistance: Mag. Dr. Katharina Kaiser

Vienna, 15th April, 2016

Michaela Lehner   Horst Eidenberger

FAKULTÄT
FÜR !NFORMATIK

Faculty of Informatics

Vergleich von Therapieplan Modellen

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieurin

im Rahmen des Studiums

Biomedical Engineering

eingereicht von

Michaela Lehner BSc
Matrikelnummer 0825832

an der Fakultät für Informatik
der Technischen Universität Wien

Betreuung: Ao.Univ.Prof. Mag. Dr. Horst Eidenberger
Mitwirkung: Mag. Dr. Katharina Kaiser

Wien, 15. April 2016

Michaela Lehner Horst Eidenberger

Erklärung zur Verfassung der Arbeit

Michaela Lehner BSc
Am langen Felde 22/1/9, 1220 Wien

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 15. April 2016

Michaela Lehner

Acknowledgements

I would like to deeply thank my parents Petra and Dietmar, who accompanied and supported me all the years of my studies. Another thanks also goes to my grandparents Christa and Kuno, who made my student life easier with some financial support.

I especially would like to thank my boyfriend Marco. With both his knowledge and his patience and emotional support he was a great help for me.

Another thanks goes to my friends Jacqueline and Konstanze. I thank them for countless hours of studying and preparing together. With them, my studies turned out a lot funnier.

Last but not least, I thank my supervisors Katharina Kaiser and Horst Eidenberger.

This research was carried out as part of the ‘Support for development and transformation of CGPs’ project funded by the Austrian Science Fund (FWF), grant number TRP71-N23.

Danksagung

Ich möchte mich von ganzem Herzen bei meinen Eltern Petra und Dietmar bedanken, die mich all die Jahre meines Studiums begleitet und unterstützt haben. Ein weiterer Dank gilt auch meinen Großeltern Christa und Kuno, die mir mit manch finanzieller Unterstützung das Studentenleben leichter gemacht haben.
Ganz besonders möchte ich meinem Freund Marco danken. Er war sowohl mit seinem Wissen als auch mit seiner Geduld und seinem seelischen Beistand eine große Hilfe für mich.
Ein weiterer Dank gilt meinen Freundinnen Jacqueline und Konstanze. Ich danke ihnen für die unzähligen gemeinsamen Stunden des Lernens und Vorbereitens. Mit ihnen hat sich mein Studium um einiges lustiger gestaltet.
Zu guter Letzt bedanke ich mich bei meinen Betreuern Katharina Kaiser und Horst Eidenberger.

Diese Arbeit wurde im Rahmen des ‘Unterstützung für Entwicklung und Transformation von CGPs’ Projekts durchgeführt und finanziert von Fonds zur Förderung der wissenschaftlichen Forschung (FWF), Fördernummer TRP71-N23.

Abstract

To assist physicians with the treatment planning process so-called clinical practice guidelines are created. They contain general information about a specific clinical condition as well as rules and procedures to treat patients with this condition. As these guidelines are represented in free text form they are difficult to handle for physicians in their daily working process. Therefore the information in the guideline document is formalized to create computerized guidelines. We work with the formalization language Asbru. The creation of these computer-interpretable guidelines is either done manually which is a great effort or automatically using methods of Information Extraction. The aim of this thesis is to verify whether the automatically generated model of a guideline corresponds to the manually generated model of the same guideline. Doing this manually would again be a great effort and therefore I want to investigate methods to automatically evaluate parts of these models. I focus on the procedural knowledge of the models.

To be able to compare two different models I investigated them. At first I developed methods to compare two models. To make a statement about the similarity of two models we started with comparing their activities using similarity metrics in order to identify corresponding plans of two models. In a second step we furthermore looked into their process structure using workflow patterns and tried to find similarities as well.

Then I implemented these methods prototypically and tested them using a 'real-world' example, a guideline for gestational diabetes mellitus. The example was based on the output of the GESHER tool for manually creating computerized treatment plan models and the tool LASSIE for automatically creating computerized treatment plan models. Using this example we evaluated how much of the original guideline text is present in the automatically generated model. Finally, it was evaluated if the automatically generated model finds the same information as present in the manually generated model.

Kurzfassung

Um Ärzte beim Prozess der Behandlungsplanung zu unterstützen werden so genannte klinische Leitlinien erstellt. Diese enthalten allgemeine Informationen zu einem bestimmten klinischen Zustand sowie Regeln und Verfahren um Patienten mit diesem Zustand zu behandeln. Da diese Leitlinien in freiem Text dargestellt sind, ist es für Ärzte schwierig sie bei der täglichen Arbeit zu verwenden. Deshalb wird die Information in dem Leitlinien-Dokument formalisiert um computergestützte Leitlinien zu erstellen. Wir arbeiten mit der Formalisierungssprache Asbru. Die Erstellung dieser computer-interpretierbaren Leilinien geschieht entweder manuell, was ein großer Aufwand ist, oder automatisch mit Methoden der Informationsextraktion. Das Ziel dieser Arbeit ist zu überprüfen, ob das automatisch erstellte Model einer Leitlinie dem manuell erstellten Model der selben Leitlinie entspricht. Dies manuell zu tun wäre wieder ein großer Aufwand und daher möchte ich Methoden untersuchen, um automatisch Teile dieser Modelle zu evaluieren. Ich konzentriere mich auf das prozedurale Wissen der Modelle.

Um zwei verschiedene Modelle vergleichen zu können habe ich sie untersucht. Zuerst habe ich Methoden entwickelt, um zwei Modelle zu vergleichen. Um eine Aussage über die Ähnlichkeit von zwei Modellen machen zu können, haben wir mit einem Vergleich ihrer Aktivitäten mit Hilfe von Ähnlichkeitsmetriken begonnen, um übereinstimmende Pläne der zwei Modelle zu identifizieren. In einem zweiten Schritt haben wir uns darüber hinaus ihre Prozessstruktur unter Verwendung von Workflow-Mustern angeschaut und versucht wieder Gemeinsamkeiten zu finden.

Dann habe ich diese Methoden prototypisch implementiert und mittels eines 'real-world' Beispiels, einer Leitlinie für Gestationsdiabetes Mellitus, getestet. Das Beispiel basierte auf der Ausgabe des GESHER Tools für die manuelle Erstellung von computergestützten Behandlungsplan-Modellen und des Tools LASSIE zum automatischen Erstellen von computergestützten Behandlungsplan-Modellen.

Unter Verwendung dieses Beispiels haben wir evaluiert, wie viel von dem originalen Leitlinien Text im automatisch erzeugten Modell vorkommt. Schließlich wurde evaluiert, ob das automatisch erstellte Modell dieselbe Information findet, die im manuell erstellten Modell vorkommt.

Contents

Acknowledgements **vii**

Danksagung **ix**

Abstract **xi**

Kurzfassung **xiii**

Contents **xv**

List of Figures **xvi**

List of Tables **xvii**

Listings **xviii**

1 Introduction **1**
1.1 Presentation of the problem 1
1.2 Approach to solving the problem 3

2 Related Work **5**
2.1 Clinical guideline modeling 5
2.1.1 Model-centric approaches 6
2.1.2 Document-centric approaches 8
2.1.3 NLP approaches 9
2.2 Evaluation of process models 11

3 Method **17**
3.1 Challenges 17
3.2 Comparison of single activities of the models 19
3.2.1 Preprocessing 20
3.2.2 The comparison 22
3.3 Comparison of the process structure of the models 23
3.3.1 The Exclusive Choice pattern 25
3.3.2 The comparison 29

4 Implementation **31**
4.1 GESHER and LASSIE 31
4.1.1 GESHER 31
4.1.2 LASSIE 33
4.2 Implementation of the method 34
4.2.1 Comparison of single activities of the models 36
4.2.2 Comparison of the process structure of the models 37
4.3 Evaluation 40
4.3.1 Evaluation of ‘activity’ and ‘condition’ information found by LASSIE 41
4.3.2 Evaluation of the plans 42
4.3.3 Evaluation of the process structure 43

5 Conclusion **49**

Bibliography **51**

A GESHER **57**

B LASSIE **61**

C GDM guideline **65**

Glossary **77**

Acronyms **79**

List of Figures

1.1 Part of a formal model. 2

2.1 PRO*forma* task model [19]. 7
2.2 Example of a guideline for hepatitis in pregnancy using the Tallis tool. 7
2.3 Example of a guideline for hepatitis in pregnancy using the DELT/A tool [49]. 8
2.4 BPMN example of a pizza collaboration. 11

3.1 Overview of the comparison method. 17
3.2 Overview of the preprocessing. 20
3.3 Three dimensional view of a workflow [45]. 24

3.4 Sequence pattern (left) and Parallel Split pattern (right) (the patterns are illustrated using the Coloured Petri-Net formalism [36]). 25
3.5 BPMN's Exclusive Choice and Simple Merge. 26
3.6 Structure of procedural approach of the Exclusive Choice pattern. 26
3.7 Structure of declarative approach of the Exclusive Choice pattern. 27
3.8 Structure of declarative approach of the Exclusive Choice pattern in AsbruView. 28

4.1 GESHER user interface - procedural knowledge. 32
4.2 GESHER user interface - declarative knowledge. 32
4.3 Flow chart of LASSIE's knowledge generation [3]. 33
4.4 Tree view of LASSIE's plans of the GDM guideline. 34
4.5 GESHER's plan hierarchy according to the GDM guideline. 35
4.6 Example of a WordNet database entry. 36
4.7 GESHER plan #19969 (with key on the right). 46

List of Tables

3.1 Example of preprocessing activity labels. 22
3.2 Adaptation of the procedural approach. 29

4.1 Scoring key for the evaluation [28, p. 5]. 41
4.2 Evaluation of LASSIE's 'activity' markups. 41
4.3 Evaluation of LASSIE's 'condition' markups. 41
4.4 LASSIE's plans. 42
4.5 GESHER's plans. 42
4.6 Evaluation of the plans. 42
4.7 GESHER's plans without orphaned plans. 43
4.8 Evaluation of the plans without orphaned plans. 43
4.9 LASSIE `filter-precondition` vs. GESHER `if-then-else`. 44
4.10 Structure of two matching plans. 45
4.11 Structure of LASSIE plan #90689. 45
4.12 Structure of GESHER plan #19969 and subplans #19972 and #19973. . . . 46
4.13 Adapted GESHER plan matches LASSIE plan. 47

Listings

3.1 Example of the plan structure. 19
3.2 Example of `if-then-else`. 26
3.3 Example of `filter-precondition`. 28
3.4 Example of adapted procedural approach. 29
4.1 List of stop words. 36
4.2 Part of XSLT template to transform procedural to declarative Exclusive Choice pattern. 37
4.3 Example of an input of the XSLT templates. 38
4.4 Example of an output of the XSLT templates. 39
A.1 GESHER plan #19922. 57
A.2 GESHER plan #19969. 58
A.3 GESHER plan #19973. 58
B.1 LASSIE plan #90689. 61
B.2 LASSIE plan #90697. 62

CHAPTER 1

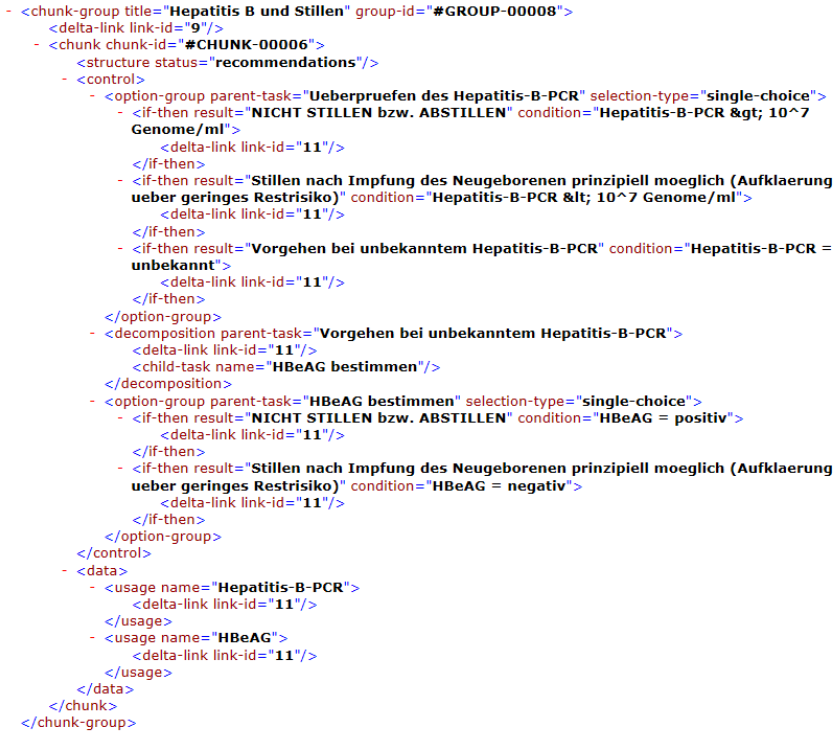
Introduction

1.1 Presentation of the problem

Physicians are daily challenged with the complex task of choosing an appropriate therapy for a patient. In complex and lengthy treatment processes, such as with chronic diseases or cancer, the treatment planning may also be a challenge - for both the physician and the patient.

In order to help the physicians with treatment planning clinical guidelines and protocols are developed. A clinical practice guideline is a set of general rules and policies for the management of patients who have a particular clinical condition. The information in a guideline is in plain text form, in tables or in flow charts. Clinical practice guidelines (CPGs) are statements that assist the physician and the patient in decision making. They represent the best clinical practice. Computer-interpretable guidelines (CIGs) are also very important for quality assurance; uniform standards are created thereby and they are also used for education of physicians [24].

At first these guidelines only exist in free text form which is not very practicable for a physician in the daily working process. That is why computerized treatment plans are created out of clinical guidelines. To execute CPGs in a computer-supported way the information in the guideline has to be formalized [24] (see Figure 1.1 for an example of a formal representation of part of a guideline). In order to do this the computerized guidelines have to be transformed into special CIG formalisms, e.g., Asbru, PRO*forma*, GLIF [11]. Usually treatment plans are created completely manually using special editors. These editors can be divided into two main categories: model-centric and document-centric (for more details see Chapter 2.1). At the moment, using these tools and methods a guideline developer needs not only knowledge about the formal methods but also about the corresponding medical domain [24].



```
- <chunk-group title="Hepatitis B und Stillen" group-id="#GROUP-00008">
    <delta-link link-id="9"/>
  - <chunk chunk-id="#CHUNK-00006">
      <structure status="recommendations"/>
    - <control>
      - <option-group parent-task="Ueberpruefen des Hepatitis-B-PCR" selection-type="single-choice">
        - <if-then result="NICHT STILLEN bzw. ABSTILLEN" condition="Hepatitis-B-PCR &gt; 10^7
          Genome/ml">
            <delta-link link-id="11"/>
          </if-then>
        - <if-then result="Stillen nach Impfung des Neugeborenen prinzipiell moeglich (Aufklaerung
          ueber geringes Restrisiko)" condition="Hepatitis-B-PCR &lt; 10^7 Genome/ml">
            <delta-link link-id="11"/>
          </if-then>
        - <if-then result="Vorgehen bei unbekanntem Hepatitis-B-PCR" condition="Hepatitis-B-PCR =
          unbekannt">
            <delta-link link-id="11"/>
          </if-then>
        </option-group>
      - <decomposition parent-task="Vorgehen bei unbekanntem Hepatitis-B-PCR">
          <delta-link link-id="11"/>
          <child-task name="HBeAG bestimmen"/>
        </decomposition>
      - <option-group parent-task="HBeAG bestimmen" selection-type="single-choice">
        - <if-then result="NICHT STILLEN bzw. ABSTILLEN" condition="HBeAG = positiv">
            <delta-link link-id="11"/>
          </if-then>
        - <if-then result="Stillen nach Impfung des Neugeborenen prinzipiell moeglich (Aufklaerung
          ueber geringes Restrisiko)" condition="HBeAG = negativ">
            <delta-link link-id="11"/>
          </if-then>
        </option-group>
      </control>
    - <data>
      - <usage name="Hepatitis-B-PCR">
          <delta-link link-id="11"/>
        </usage>
      - <usage name="HBeAG">
          <delta-link link-id="11"/>
        </usage>
      </data>
    </chunk>
  </chunk-group>
```

Figure 1.1: Part of a formal model.

Besides creating them manually, which is very time-consuming and challenging, computerized treatment plans or parts thereof can also be generated automatically with complex tools that use Information Extraction (IE) methods [24, 39]. The automatically generated models usually consist of text parts taken from the guideline document whereas the manually created models are developed according to the modeler's knowledge and interpretation of the guideline text. In the area of Business Process Modeling there exist certain rules the modeler has to stick to (e.g., rules about wording) [31]. Unfortunately, in the medical domain such rules have not been adopted yet and therefore the comparison of different models is a big challenge.

Comparing two models and checking which concepts in one model correspond to the concepts in the other model is usually done completely manually, which is a great effort. Therefore, it would be a big ease if some parts of this comparison as well as the evaluation of the IE methods were done automatically. The goal of this project is to develop methods to compare parts of process models automatically. The complete automation of the comparison is nearly impossible due to the complexity of the models and the variety of information dimensions.

Comparing an automatically generated model to a manually generated model is problematic due to the fact that the automatically generated model contains text from the guideline document whereas the manually generated model is prone to errors made by the modelers during the development.

1.2 Approach to solving the problem

An entire executable CPG model contains many different information aspects: The most prominent ones are procedural knowledge (e.g., workflows) as well as declarative knowledge (e.g., definition of clinically meaningful terms and temporal patterns) [20]. Procedural knowledge consists of actions, conditions controlling them and information how to order actions (sequentially, parallel, etc.). Normally guideline documents contain not only procedural and declarative knowledge, but also other kinds of information such as effects or intentions of activities, evidence information, information about risks and benefits of activities, etc. [3]. As the procedural model is an important part of a model, I want to focus on it in my thesis.

The goal of this project is the development of methods and their prototypical implementation that allow for the evaluation of the automatically generated procedural model. These methods should compare the automatically generated model with the manually generated model of the same origin and identify similarities and differences between these two models. The methods will be developed for the language Asbru [38]. Asbru is a task-specific and intention-based plan representation language to embody clinical guidelines and protocols as time-oriented skeletal plans [34].

The main objective is to find methods to compare automatically generated procedural models of clinical practice guidelines with a manually generated gold standard in order to evaluate the automatically generated models.

It is important to mention that usually a manually created model contains much more information than that is actually available in the guideline text. Studies showed that only about 50% of the knowledge in the final model comes from the guideline text. Generating a model physicians create much of the knowledge when creating the consensus and therefore it cannot be pinned down to a specific text in the guideline. Much of the knowledge is implicit in the guideline and not made sufficiently explicit [20].

As a result, it is not possible to tell if the automatically generated model is as good as the manually generated gold standard. It is only possible to evaluate if the concepts of the automatically generated model have correspondences in the manually generated model but not vice versa.

Comparing two models, a binary answer (two models are equivalent or not) is not very helpful. But instead it is important to differentiate between slightly different models and completely different models [46].

It also has to be considered when and how such an evaluation can be performed and when it makes sense. For this evaluation a comparison model (i.e., a gold standard) is definitely necessary. We probably have the need for such a comparison model during the development of methods using IE. There it would be helpful to have a tool for testing during the development. But such methods can also be applied when working with several manually generated models. They could for example be compared or similarities could be identified. The methods could also be helpful when studying (manual) modeling. Another possible application might be with models represented in different formalisms:

e.g., having a model for a specific treatment in PRO*forma* to be translated into Asbru or integrating it into an existing Asbru model.

The research for this thesis was carried out in the context of the FWF project ‘Support for development and transformation of CGPs’, grant number TRP71-N23.
The remainder of this thesis is structured as follows. Chapter 2 presents various approaches with examples of clinical guideline modeling (model-centric approaches, document-centric approaches and Natural Language Processing (NLP) approaches are considered) and the evaluation of process models. In Chapter 3 the challenges and the theoretical steps of the comparison of the activities and the process structure of the manually and automatically generated models are introduced and explained. Chapter 4 introduces the two used tools for creating computerized treatment plan models: GESHER for the manually created model and LASSIE for the automatically created model. The methods of the previous chapter are implemented prototypically and finally the evaluation of the obtained results is performed. Chapter 5 concludes the thesis.

CHAPTER 2

Related Work

In this chapter I want to introduce several current approaches of clinical guideline modeling as well as the evaluation of process models.
I will present an overview of the different ways to create CIGs and give examples. This is important to understand that in the medical domain various methods exist to deal with the problem of modeling clinical guidelines and that there does not exist one over all solution that is applicable for every presentation of a problem.
Furthermore, I will take a closer look at the different aspects of process model evaluation and explain why and how process models can be evaluated. Thereby I am trying to find and compare approaches that are useful and applicable for my problem.

2.1 Clinical guideline modeling

In the last years different approaches to manage clinical guidelines and protocols were introduced. To keep track of the diversity the concept of plan management was introduced, where clinical guidelines are seen as (time-oriented) plans [5].
Plan management includes everything from designing a particular plan to the real-world execution and evaluation of such plans. It consists of various tasks that can be differentiated into tasks, which need to be performed mainly at design time (e.g., Plan Generation, Plan Verification, Plan Validation, Plan Visualization) and those which are done mainly during execution time of plans (e.g., Plan Selection, Plan Adaptation, Plan Execution, Plan Modification / Alternatives). An outstanding task is Plan Modification / Alternatives: This covers, on the one hand, the maintaining of clinical guidelines when new medical knowledge is discovered and needs to be included in the clinical guidelines (in the sense of ‘living guidelines’). On the other hand, changes in health condition of the patients or in the medical environment can force a modification of the therapeutic activities [5].

Different frameworks have been developed to implement clinical guidelines in a computer-interpretable format (cf. [34, 22]), such as Asbru, EON, GLIF, Guide, Prodigy and PRO*forma*. These frameworks are created for specific classes of guidelines, specific users and specific organizations. Each framework supports specific guideline representation languages. Various tools and techniques have been developed to ease the guideline modeling and visualization process. They can be classified into model-centric and document-centric approaches [5]. Furthermore some NLP approaches are also worth mentioning.

2.1.1 Model-centric approaches

The model-centric approaches to author clinical guidelines focus on the creation of a conceptual model of the original guideline. But they do not keep the direct connection between these two representations. This category covers many approaches and is more visual-oriented [5]:

- The most widely used visual representation of clinical guidelines are so-called flowchart algorithms. But they were intended to be used on paper and have never been implemented by programmers [5].

- Protégé is an open source ontology development and knowledge acquisition environment. It assists users in the construction of large electronic knowledge bases. Protégé implements various CIG-languages, such as EON, GLIF, PRO*forma* [5, 24].

- AsbruView is a graphical user interface to support the development of guidelines and protocols in Asbru. It is a tool to make Asbru accessible to physicians and to provide a visual overview of the guideline hierarchy and other Asbru-specific components [5, 24].

- Arezzo and the Tallis Toolset are based on the PRO*forma* language for modeling clinical processes (cf. Figures 2.1 and 2.2). They include software and training materials to create, publish and illustrate clinical knowledge applications over the web [5, 24].

- GLARE is a domain-independent system for acquiring, representing and executing clinical guidelines. The flow of the guideline is represented similar to a flowchart [5].

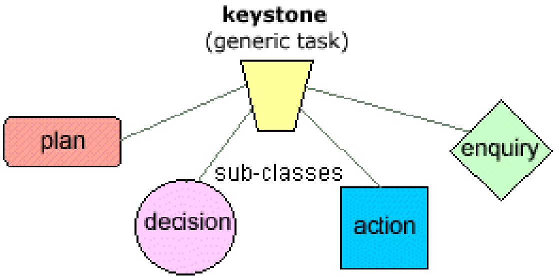


Figure 2.1: PRO*forma* task model [19].

PRO*forma* models a guideline as a set of tasks and data items. Figure 2.1 shows the PRO*forma* task model. The keystone is divided into four task types: Plans are the basic building blocks and can contain any number of any task type. Decisions are taken when options are presented. Actions are clinical procedures which have to be carried out. Enquiries are requests for further information or data [19].

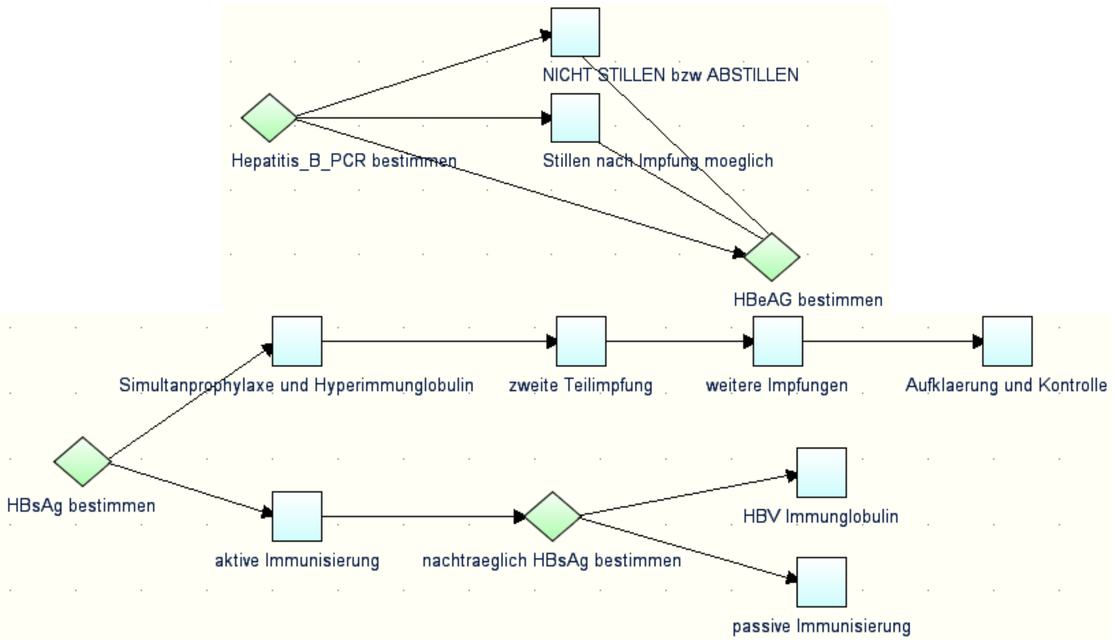

NICHT STILLEN bzw ABSTILLEN
Hepatitis_B_PCR bestimmen
Stillen nach Impfung moeglich
HBeAG bestimmen
Simultanprophylaxe und Hyperimmunglobulin
zweite Teilimpfung
weitere Impfungen
Aufklaerung und Kontrolle
HBsAg bestimmen
HBV Immunglobulin
aktive Immunisierung
nachtraeglich HBsAg bestimmen
passive Immunisierung

Figure 2.2: Example of a guideline for hepatitis in pregnancy using the Tallis tool.

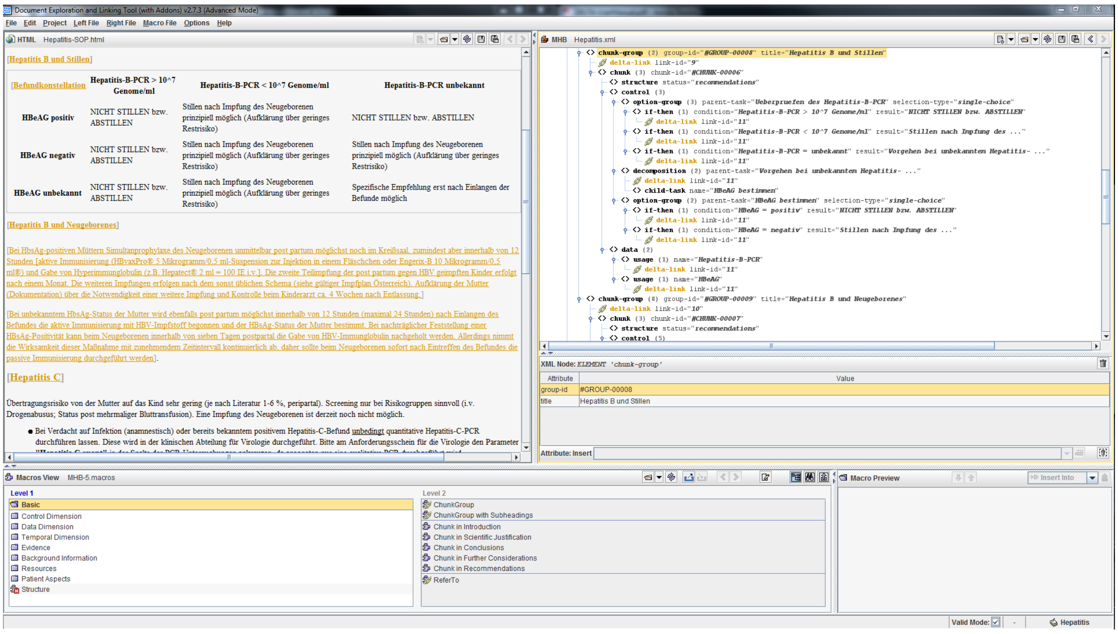
Figure 2.2 shows an example of a guideline modeled with the Tallis tool. Tallis consists of a composer to create, edit and visualize guidelines; a tester to test and debug the application; and an engine to enact guidelines [19].

2.1.2 Document-centric approaches

The document-centric approach to author clinical guidelines preserves the connection of the original guideline written in text and its semi-formal model. Markup-based tools are used to systematically mark up the original guideline in order to generate a semi-formal model of the marked text part. This approach is more text-based and only a few examples exist [5]:

- The GEM Cutter transforms clinical guidelines into the Guideline Elements Model (GEM), which is an XML-based guideline document model [5, 24].

- The Document Exploration and Linking Tool / Addons (DELT/A) (cf. Figure 2.3) supports the translation of HTML documents into an XML language, such as Asbru or MHB. DELT/A allows the definition of links between the original guideline and the target representation, which gives the user the possibility to find out where a certain value in the XML-language notation comes from [5, 24].



[Hepatitis B und Stillen]

[Befundkonstellation	Hepatitis-B-PCR > 10^7 Genome/ml	Hepatitis-B-PCR < 10^7 Genome/ml	Hepatitis-B-PCR unbekannt
HBeAG positiv	NICHT STILLEN bzw. ABSTILLEN	Stillen nach Impfung des Neugeborenen prinzipiell möglich (Aufklärung über geringes Restrisiko)	NICHT STILLEN bzw. ABSTILLEN
HBeAG negativ	NICHT STILLEN bzw. ABSTILLEN	Stillen nach Impfung des Neugeborenen prinzipiell möglich (Aufklärung über geringes Restrisiko)	Stillen nach Impfung des Neugeborenen prinzipiell möglich (Aufklärung über geringes Restrisiko)
HBeAG unbekannt	NICHT STILLEN bzw. ABSTILLEN	Stillen nach Impfung des Neugeborenen prinzipiell möglich (Aufklärung über geringes Restrisiko)	Spezifische Empfehlung erst nach Einlangen der Befunde möglich

[Hepatitis B und Neugeborenes]

[Bei HbsAg-positiven Müttern Simultanprophylaxe des Neugeborenen unmittelbar post partum möglichst noch im Kreißsaal, zumindest aber innerhalb von 12 Stunden [aktive Immunisierung (HBvaxPro® 5 Mikrogramm/0,5 ml-Suspension zur Injektion in einem Fläschchen oder Engerix-B 10 Mikrogramm/0,5 ml®) und Gabe von Hyperimmunglobulin (z.B. Hepatect® 2 ml = 100 IE i.v.]. Die zweite Teilimpfung der post partum gegen HBV geimpften Kinder erfolgt nach einem Monat. Die weiteren Impfungen erfolgen nach dem sonst üblichen Schema (siehe gültiger Impfplan Österreich). Aufklärung der Mutter (Dokumentation) über die Notwendigkeit einer weitere Impfung und Kontrolle beim Kinderarzt ca. 4 Wochen nach Entlassung.]

[Bei unbekanntem HbsAg-Status der Mutter wird ebenfalls post partum möglichst innerhalb von 12 Stunden (maximal 24 Stunden) nach Einlangen des Befundes die aktive Immunisierung mit HBV-Impfstoff begonnen und der HBsAg-Status der Mutter bestimmt. Bei nachträglicher Feststellung einer HBsAg-Positivität kann beim Neugeborenen innerhalb von sieben Tagen postpartal die Gabe von HBV-Immunglobulin nachgeholt werden. Allerdings nimmt die Wirksamkeit dieser Maßnahme mit zunehmendem Zeitintervall kontinuierlich ab, daher sollte beim Neugeborenen sofort nach Eintreffen des Befundes die passive Immunisierung durchgeführt werden].

[Hepatitis C]

Übertragungsrisiko von der Mutter auf das Kind sehr gering (je nach Literatur 1-6 %, peripartal). Screening nur bei Risikogruppen sinnvoll (i.v. Drogenabusus; Status post mehrmaliger Bluttransfusion). Eine Impfung des Neugeborenen ist derzeit noch nicht möglich.

- Bei Verdacht auf Infektion (anamnestisch) oder bereits bekanntem positivem Hepatitis-C-Befund unbedingt quantitative Hepatitis-C-PCR durchführen lassen. Diese wird in der klinischen Abteilung für Virologie durchgeführt. Bitte am Anforderungsschein für die Virologie den Parameter

```
<> chunk-group (2) group-id="#GROUP-00008" title="Hepatitis B und Stillen"
   delta-link link-id="9"
   <> chunk (3) chunk-id="#CHUNK-00006"
      <> structure status="recommendations"
      <> control (3)
         <> option-group (3) parent-task="Ueberpruefen des Hepatitis-B-PCR" selection-type="single-choice"
            <> if-then (1) condition="Hepatitis-B-PCR > 10^7 Genome/ml" result="NICHT STILLEN bzw. ABSTILLEN"
               delta-link link-id="11"
            <> if-then (1) condition="Hepatitis-B-PCR < 10^7 Genome/ml" result="Stillen nach Impfung des ..."
               delta-link link-id="11"
            <> if-then (1) condition="Hepatitis-B-PCR = unbekannt" result="Vorgehen bei unbekannten Hepatitis- ..."
               delta-link link-id="11"
         <> decomposition (2) parent-task="Vorgehen bei unbekanntem Hepatitis- ..."
            delta-link link-id="11"
            <> child-task name="HBeAG bestimmen"
         <> option-group (2) parent-task="HBeAG bestimmen" selection-type="single-choice"
            <> if-then (1) condition="HBeAG = positiv" result="NICHT STILLEN bzw. ABSTILLEN"
               delta-link link-id="11"
            <> if-then (1) condition="HBeAG = negativ" result="Stillen nach Impfung des ..."
               delta-link link-id="11"
      <> data (2)
         <> usage (1) name="Hepatitis-B-PCR"
            delta-link link-id="11"
         <> usage (1) name="HBeAG"
            delta-link link-id="11"
<> chunk-group (8) group-id="#GROUP-00009" title="Hepatitis B und Neugeborenes"
   delta-link link-id="10"
   <> chunk (3) chunk-id="#CHUNK-00007"
      <> structure status="recommendations"
      <> control (5)
```

XML Node: *ELEMENT 'chunk-group'*

Attribute	Value
group-id	#GROUP-00008
title	Hepatitis B und Stillen

Attribute: Insert

Level 1	Level 2
Basic	ChunkGroup
Control Dimension	ChunkGroup with Subheadings
Data Dimension	Chunk in Introduction
Temporal Dimension	Chunk in Scientific Justification
Evidence	Chunk in Conclusions
Background Information	Chunk in Further Considerations
Resources	Chunk in Recommendations
Patient Aspects	ReferTo
Structure	

Figure 2.3: Example of a guideline for hepatitis in pregnancy using the DELT/A tool [49].

Figure 2.3 shows an example of a guideline using the DELT/A tool [49]. On the left side we see the clinical guideline in free text form. The right side shows the corresponding XML model which was created manually to support the generation of the model according to the underlying schema of the guideline representation language.

2.1.3 NLP approaches

Medical documents are mostly unstructured free-text. Their extension and their huge number makes them difficult and time-consuming to process manually. Medical NLP tries to solve these problems by transforming natural language text into standardized semantic structures to make data in written text accessible [14].
One approach is the use of IE. IE is an emerging NLP technology whose function is to process unstructured, natural language text, to locate specific pieces of information in the text and to use them to fill a database. The resulting output can then be processed to yield refined representations, leading to the representation in a specific guideline representation language [24].
A big challenge when authoring clinical guidelines is the detection of individual processes and their relations and dependencies. In this section some approaches that use NLP techniques to detect these processes are presented:

- One important IE approach is the LASSIE approach with the main goal being the acquiring of treatment processes from clinical guidelines. It automates parts of the modeling process, it provides a medical ontology, it structures the information of the guideline and provides a basis for the following transformation of the process information into any guideline representation language (e.g., MHB, Asbru) [24].
  LASSIE's rules to extract the process information are extraction patterns which are based on syntactic and semantic restrictions as well as delimiters. Patterns are defined on three levels: 1) Phrase level patterns are used to identify basic entities (e.g., time, conditions) which build the attributes of actions. 2) Sentence level patterns use phrase level patterns, medical terms and trigger words for medical terms to identify actions and their attributes. The trigger words are mainly verbs and indicate the application or the avoidance of a therapy. 3) Discourse level patterns are based on sentence level patterns, but are extended to consider the structure of the documents. They categorize sentences, merge them to actions and find relationships between actions to structure them. Medical terms (i.e., drugs, surgical procedures) are based on libraries provided by medical institutions [24].
  The extraction of processes from clinical guidelines is done in two steps: 1) The relevant sentences containing treatment instructions are extracted by marking-up the original guideline document. 2) Several sentences are combined to one action, the actions are structured and relations among them are detected. Thereby a representation that is independent of the subsequent guideline representation language is obtained [24].
  LASSIE was extended in the context of the MobiGuide project: various machine learning methods were used to identify clinical actions and the medical terminology systems were extended (cf. Section 4.1.2) [3].

- The VeriCliG project [39] studied how to automatically extract CIGs from clinical guidelines using NLP techniques. This project wants to develop biomedical NLP techniques, especially automated CIG extraction methodologies. The goal is to extract the main control flow structures emerging from the guideline text in order

to express them using known representation languages (e.g., the Business Process Model and Notation (BPMN) standard). To realize this goal the VeriCliG project builds on the work on clinical semantic and syntactic annotation as well as on BPMN model extraction [42].
To describe processes the VeriCliG project uses the terminology coming from the BPMN standard. In BPMN a process consists of five basic components. The first three are static components: activities (e.g., controlling blood glucose levels), actors (e.g., doctors, nurses, patients) and resources (e.g., drugs). The control flow component (e.g., 'if-then-else') is dynamic and the last component is the message flow. To identify these components MetaMap Unified Medical Language System (UMLS) is used to find process evoking words. MetaMap maps biomedical text to the UMLS Metathesaurus. The VeriCliG project wants to adapt BPMN techniques to the clinical setting by combining them with biomedical annotations [42, 41].
The VeriCliG project's CIG extraction methodology consists of: 1) combining annotation resources to extract CIG resources, actors and activities; 2) analyzing syntactic structures to extract CIG control structures; 3) resolving ambiguities and detecting temporal relations to build a CIG [42, 40].
The project intends to help medical staff to save time and resources by developing techniques and technologies to fully or partially assist and automate the task of generating and repairing careflows [39].

- In [15] an approach that transforms sentences of a medical document into semantic representations is presented. Therefore a multiaxial medical nomenclature, a concept-based morpheme lexicon, a word segmentation algorithm and semantic transformation rules for mapping syntactic information to semantic roles are used. The strategy can also be used as a basis of IE [15].
At first a preparser finds sentence boundaries in the text and then each sentence gets decomposed into segments by a syntactic analysis. The next step is to check each sentence for special expressions (e.g., dates, quantities, dosages, etc.) which are then tagged before the indexing process as they often require special interpretation. Next, each word is tagged by decomposing it into its morphemes which are the smallest units of a word. Furthermore the morphological features of each word are determined: word category, language, gender, case, number. The indexing maps every sentence to a set of indices of the Wingert Nomenclature. This nomenclature 'comprises a complete multiaxial terminology of medical terms and allows encoding different aspects of a diagnosis or procedure' [15]. An index consists of a letter representing the axis (topography (T), morphology (M), function (F), procedure (P), diagnosis (D), job (J), information (G), treatment (V), agent (W), aetiology (E)) and a six-digit number for identification. The tagged words are compared with the tagged terms of the Wingert Nomenclature-Lexicon. If they match, the corresponding indices are assigned to the morphemes of the input words. Next, the algorithm checks if it is possible to combine various simple indices to one single index. This step is repeated until a final set of indices is determined. The semantic analysis starts by identifying the central information unit (leading index) and the

set of modifying information (non-leading indices) for each sentence and generates the corresponding concepts. Then the non-leading concepts are linked to their leading concept by relation. The type of the relation is determined from the concept type of each non-leading concept [15, 14].
The method SeReMeD in [13] is a similar approach for the automatic generation of semantic representations. The big difference to the above described approach is the use of MetaMap to map natural language text to UMLS concepts. Contextual relations are automatically identified. For additional semantic relations the UMLS Semantic Network and relationships between concepts defined in the UMLS Metathesaurus are used [13].

2.2 Evaluation of process models

A process model describes the activities of a process together with their execution dependencies [8]. See Figure 2.4 for an example of a process model[1].

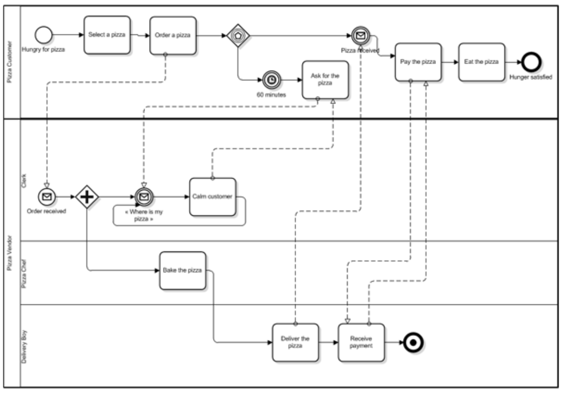


Figure 2.4: BPMN example of a pizza collaboration.

Why evaluate process models?

One reason why process models need to be evaluated is process model matching. The goal of process model matching is the creation of an alignment between process models, i.e., finding similarities between their activities [8].
An important application is the management of large process model repositories which requires effective search techniques. For example, before adding a new process model to a repository, one needs to check that a similar model does not already exist in order to prevent duplication. It is important to identify common or similar processes in order to analyze their overlap and to identify areas for merging [17].
Process model matching is also useful when comparing a gold standard model to a newly

[1] from `http://www.bpmn.org/`, accessed: November 17, 2015

created model as it is the case in the presentation of the problem of this thesis.

Another important part of the evaluation of process models is to verify the model itself and to check if it is consistent.

What can be evaluated?

A single process model can be evaluated in two ways: The model can be checked if it represents the guideline (validation) or it can be checked if the model of the guideline is correct (verification) [21].

When comparing two process models, common labels as well as the structure of the models can be evaluated.

Challenges of evaluation

A big challenge when comparing process models is the wording of the labels. As mentioned in the previous chapter, in the medical domain exist no rules when creating a model from a guideline document. Therefore labels can be ambiguous or different labels can describe the same process by using different phrases and/or grammatical styles.

Existing approaches

This section presents current approaches for the evaluation of process models divided by their application area.

1. Validation of process models
   Validation means checking if the model represents the guideline [21].
   In [30] natural language text is used as a basis for model validation. It is investigated how far a textual representation of the contents of a process model can help someone to understand the content of a model. The objective is to find out how textual information can be automatically provided if the process model is available. An automatic approach is presented for generating natural language texts from Petri nets [30].

2. Verification of process models
   Verification means checking if the model of the guideline is correct [21].
   In [32] a few studies that report results from applying verification techniques to process models are compared. The comparison reveals that most of the samples have error rates of 10% to 20% [32]. Soundness is the classical correctness criterion for process models. ‘Soundness identifies all deadlocks and lack of synchronization for process models with one start and one end node’ [32]. Further criteria are relaxed soundness, interactive verification, EPC soundness, decomposition and reduction (cf. [32]).

A limitation of soundness verification is that it covers only the control flow perspective of the process model. Therefore, soundness is a necessary but insufficient condition for correctness [50].
In [50] four verification tasks are presented: precondition conflicts, effect conflicts, reachability and executability. Based on this formalism, one can detect execution problems in processes with sound control flow, hence enabling verification beyond soundness [50].

Hommersom et al. use formal methods for the verification of clinical practice guidelines. They focus on verification prior to execution. There are roughly two verification approaches: model checking which explores a (finite) model and theorem proving which explores logical derivations of a theory [21].
Interactive theorem proving systems do not construct proofs themselves, but rather support the construction of a proof by a user. The main advantage is that it can handle problems of random complexity, therefore it is especially suitable if the model of the guideline is detailed and contains many complex constructs. By abstracting parts of the guideline, more automated techniques such as automated theorem proving or model checking become possible. Using model checking, temporal properties can be automatically verified for a given state transition system. In principle, model checking is automatic, but the application is limited to finite state transition systems [21].
In [21] several other techniques for checking that a guideline model is internally consistent are also presented.

3. Process model matching
Dijkman et al. present three similarity metrics that can be used to answer requests on process repositories: 1) node matching similarity metrics that measure similarity based on properties of process model elements (such as their labels and their other attributes), 2) structural similarity metrics that measure similarity based on the properties of process model elements as well as the relations between these elements and 3) behavioral similarity metrics that measure similarity based on the intended behavior of process models. These metrics all outperform text-based search engines when it comes to searching through a repository for similar process models [17].

The *Triple-S* matching approach combines similarity scores of syntactic, semantic and structural levels to match pairs of different process models. The syntactic level consists of tokenization and stop word elimination. The Leveshtein distance between each combination of tokens is calculated. The final syntactic score is the minimum average distance between each token. For the semantic level the approach of [54] is used to calculate the semantic similarity. The structural level compares the ratio of in- and outgoing arcs and the relative position in the net. The three scores are combined to a final score that represents how good two activities of different process models match [8].

This approach is an adjustment of [18] where syntactic, linguistic and structural measures are used to compute similarity degrees between process models. Additionally, the linguistic similarity degree also determines synonyms by using a dictionary and considers common synonyms. The structural similarity measures detect homonyms and evaluate the context of the terms [18].
These two methods seem as a good basis to use for my problem. Especially the use of synonyms is an interesting way to approach the problem.

Business process graph matching considers a process as a labeled graph where nodes correspond to tasks or events and edges represent the control flow between nodes. The mapping connects one node in a graph to at most one node in another graph. Therefore the mapping generates a distance between two graphs by calculating and adding 1) the number of inserted nodes (nodes that appear in one graph but not in the other, i.e., nodes that are not part of the mapping), 2) the sum of the distances between nodes that are part of the mapping, 3) the number of inserted edges (edges that appear in one graph but not in the other). Each operation has a cost defined by a cost function. The graph matching algorithm tries to find the mapping with the smallest possible distance (called graph-edit distance which is defined as the minimal cost to transform one graph into another graph) [8, 16]. In [16] some algorithms to compute the graph edit distance of two process models are presented, e.g., the greedy algorithm and the A-star heuristic algorithm.
The *extended semantic greedy matching* approach extends the greedy matching approach. In this approach process models are matched pair-wise based on the similarities of their transitions with the result being a set of complex transition matches between two process models [8].
These strategies could be useful for me to compare the structure of the process models.

The *bag-of-words similarity with label pruning* approach only considers activity labels ignoring other information in the process model such as events or process structure. This approach computes label similarity in various steps: 1) each label is treated like a bag of words (a multi-set of words), 2) word stemming is applied for better comparability (tokenization, stop words removal), 3) similarity scores are computed for each pair of words, 4) the multi-sets for both activity labels are pruned to be equal in the number of words, 5) for each activity pair an overall matching score is computed, 6) all activity pairs with a score above a given threshold are selected. It is interesting to mention that the bag-of-words similarity ignores the grammatical structure of the label [8, 26].
This approach seems very promising to be applied for my problem. However, I do not want to compute similarity scores for every single word but rather for the whole activity label using certain similarity metrics.

Process matching using positional language model is a matching technique made for process models that consist of textual descriptions of activities, i.e., the process models are represented as documents. It adopts positional language models which 'define a document as a sequence of terms with a probability for a term at a document position' [53] and passage-based models which capture 'the probability of a term in such a passage' [53]. This approach uses the textual descriptions to identify similarities between activities which are seen as ordered passages of the documents. A similarity matrix between the activities is created and similarities are derived from this matrix [8, 53].
This strategy is not useful for my problem because the models are not represented as documents.

Another approach is *n-ary semantic cluster matching.* Clustering process model nodes consists of the following four steps executed sequentially: 1) Semantic error detection (errors of modeling are identified and handled) with the main function being the identification of wrong modeled transition nodes. 2) All models are used as input for an n-ary cluster matcher. 3) This cluster matcher uses a semantic similarity measure which consists of three steps for pairwise node comparison. First, node labels are split into single words, stop words and waste characters (e.g., additional spaces) are removed. Secondly, the stem of each word is generated and then the stem sets of both labels are compared. The number of stem matchings divided by the sum of all words yields the similarity value. Thirdly, if this resulting value is above a defined threshold, the labels are checked for antonyms and the occurrence of negation words. 4) Finally, the clusters containing nodes of all considered models are extracted to binary complex matchings. For each pair of models all clusters are searched for the occurrence of nodes from both models. The result is a binary complex matching for each model pair [8].
The *ICoP framework (Identification of Correspondences between Process Models)* focuses on complex correspondences between sets of activities (one activity can be matched to a random number of other activities). The ICoP architecture uses four different components for process model matching: Searchers identify possible matches between two process models by applying different similarity metrics and heuristics and return a set of possible correspondences with assigned scores resulting from a scoring function. Next, boosters are used to aggregate matches, remove matches and adapt their scores. Selectors select the best candidates from the set of potential matches under the condition that matches are not overlapping. Either the individual match scores are used to select the best matches or an evaluator is used which returns a single mapping score for the quality of a mapping. When the selection process is finished the selector presents the final mapping between elements of the process models [8, 51].
These two approaches appear to be rather expensive and complex and therefore not quite applicable for my problem.

CHAPTER 3

Method

In order to compare the procedural models we will use a two-step approach:
First, we start with the comparison of the plan names of the two models to identify corresponding plans. When comparing an automatically created model with a manually created one, we can make a statement if the automatically created model can be found in the manually created model. Second, among these corresponding plans we then investigate their process structure using workflow patterns to evaluate if the plans with the same meaning also have a corresponding structure. See Figure 3.1 for a short overview.

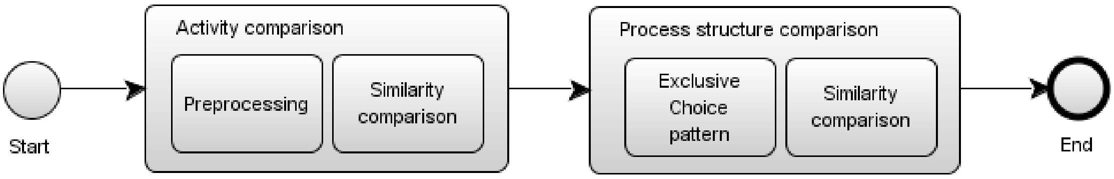


Figure 3.1: Overview of the comparison method.

For the following approach I started with an analysis of existing work (cf. Chapter 2). I adapted some of those strategies for the requirements of this presentation of the problem. However, first of all I want to address the challenges and problems that can occur when dealing with CPG models.

3.1 Challenges

The ideal case in CPG modeling would be having standards and rules for modeling like it is the case with Business Process Management (BPM) [31]. As already mentioned in Section 1.1, there are no exact rules for modeling clinical guidelines as there are for BPM guidelines. Unfortunately, those BPM standards still have not found their way into the medical modeling domain. That is one of the reasons why the manually as well as the automatically generated plans are quite difficult to process and to compare with each other. Therefore I want to present various challenges and problems that are encountered when dealing with manually and automatically generated models.

The names/labels of the plans of manually created models are prone to typing and grammatical errors. Errors found in the plans include switched letters (e.g., ‘*bec**ua**se*’, ‘*current**yl***’), wrong letters (e.g., ‘*gest**s**tional*’, ‘*meas**y**re*’, ‘*k**a**t**e**nuria*’), missing letters (e.g., ‘*measu**rm**ent*’, ‘*abn**om**al*’) and too many letters (e.g., ‘*posi**tt**ive*’, ‘*po**ss**itives*’, ‘*carbohyd**o**rates*’, ‘*enough**t***’).

Another difficulty are abbreviations (e.g., ‘*pls*’ for ‘please’, ‘*u*’ for ‘you’, ‘*w*’ for ‘with’, ‘*r*’ for ‘are’) and strange word combinations (e.g., ‘*GoodBP2Weeks*’, ‘*EVERY2DAYS*’). Such word combinations with missing whitespaces are nearly impossible to handle for the computer programs.

Sometimes information is left out and some very weird sentences are formed (e.g., ‘*('when 2 abnomal in a week= true' from twice)*’, ‘*GOOD after BAD from once a week*’). These sentences are difficult to understand for the human reader, therefore they are also pretty hard if not impossible to process for the computer programs.

Another problem with manually created models are plans that are not activated/called. Those plans exist, but there is no plan-pointer pointing to them, hence those plans are never used (e.g., from a total number of 223 plans, 128 plans are not activated - about 57,4%). This unactivated plans might be discarded or forgotten plans resulting from the fact that various people worked on them without an agreement about the exact modeling strategy. The plans were created manually, nothing was taken from the original guideline text and therefore some plans got lost during the working process.

In contrast, all of the automatically generated model's plans are activated and actually used.

The automatically generated model consists of text chunks that are taken from the original guideline text. Therefore these plans are sometimes very long and confusing, for example:

> ‘*Obstetric Delivery AND should undergo an oral glucose tolerance test 6 to 12 weeks after delivery, using a two-hour 75 gram oral glucose tolerance test AND Those with prediabetes should be counseled about their subsequent risk AND They should have yearly assessment of glycemic status AND should receive appropriate education and treatment AND She should also be given advice regarding contraception and the planning of future pregnancies AND should be counseled regarding the importance of good metabolic control prior to any future pregnancies AND should be counseled regarding their risk AND Drug therapy*’

However, these long text chunks are split up in different plan-pointers, which makes them easier to handle. As these plans are taken from the original guideline text, all typing and grammatical errors (there are hardly any) must originate from the original guideline text.

It is also possible that the automatically created model extracts too much information from the guideline document (e.g., the beginning of the next sentence is erroneously extracted with the previous sentence) or that it extracts too little information from the text (e.g., a sentence is cut off in the middle). These erroneously split up sentences and incomplete text chunks could also be seen as grammatical errors. Another option is that the model simply extracts irrelevant information.

Looking at the process structure of different models also leads to problems. Processes might have the same semantic meaning but they can be expressed differently. There are various ways to model a single process. Without exact rules each modeler can choose among a set of variants to model a certain process.

3.2 Comparison of single activities of the models

In this step we want to check if there are corresponding activities within two models. The manually created model consists of much more activities than the automatically created model. A lot of the manually created model's activities are not originating from the guideline text, but are created by the modeler due to physicians' input. In contrast, the automatically created model's activities are text chunks taken directly from the guideline text.

To fully understand a process model the ambiguity of activity labels is a significant part that has to be considered. There are different grammatical styles used in models: the verb-object style ('examine patient') and the action-noun style ('patient examining'). Studies showed that verb-object labels are considered significantly less ambiguous and more useful than action-noun labels or labels that follow neither of these styles [31].

Generally, the plan names are used for the comparison. If the label `structured-text` contains more information than the label `name` then this text is used for the comparison instead of the plan name. See Listing 3.1 for a brief example of a plan and a subplan in Asbru.

```
<hybrid-asbru-plan id="17513" name="Ketonuria">
    <plan-body>                    <!-- parent plan with subplans -->
        <semi-formal-plan-body>
            <subplans type="sequentially">
                <plan-activation>
                    <plan-schema>
                        <plan-pointer id="90686" name="Ketones"/>
                    </plan-schema>
                </plan-activation>
                <plan-activation>
                    <plan-schema>
                        <plan-pointer id="90851" name="Ketonuria"/>
                    </plan-schema>
                </plan-activation>
                <wait-for>
                    <all></all>
                </wait-for>
            </subplans>
        </semi-formal-plan-body>
    </plan-body>
</hybrid-asbru-plan>

<hybrid-asbru-plan id="90851" name="Ketonuria"> <!-- subplan -->
    <plan-body>
        <structured-text>The patient measures ketonuria</structured-text>
        <semi-formal-plan-body>
            <single-action isMandatory="false">
                <plan-activation>
                    <patientDataEntry answerType="numeric" patient-data-
                        entry-type="physical-examination" name="" glID=""
                        concept-key="90962" concept-name="Ketonuria"/>
                </plan-activation>
            </single-action>
        </semi-formal-plan-body>
    </plan-body>
</hybrid-asbru-plan>
```

Listing 3.1: Example of the plan structure.

The comparison is done by a syntactic and semantic comparison of the two texts. The syntax might be quite different because of the various grammatical styles described above. The semantics of a word might also vary due to the context. Therefore we look for example for the existence of semantically equivalent or similar verbs (i.e., synonyms). It would also be a possibility to look for hyponyms and hypernyms of nouns but this is too expensive and not very effective. A hyponym is the minor term of a word and a hypernym is the generic term of a word, e.g., 'finger' is a hyponym of 'hand'; 'hand' is a hypernym of 'finger'.

3.2.1 Preprocessing

This step is necessary because of the challenges and errors introduced in Section 3.1. Figure 3.2 shows the individual steps.

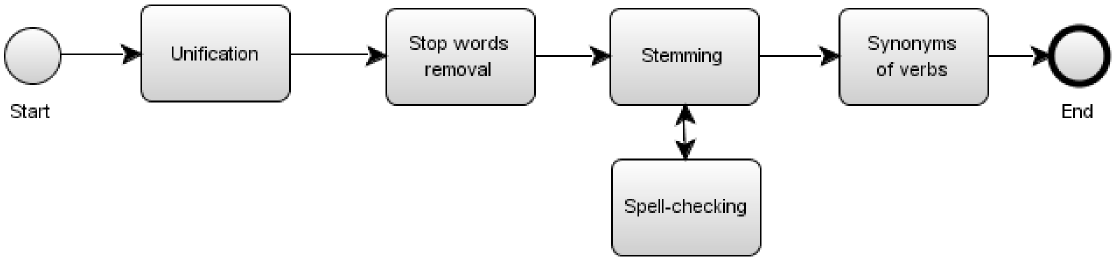


Figure 3.2: Overview of the preprocessing.

1. Unification

For a better handling and easier comparison of the labels they need to be unified:

1. All words are written in lower case.

2. All special characters and multiple whitespaces are deleted.

3. Written-out numbers are replaced by actual figures (e.g., ‘one’ becomes ‘1’).

4. All abbreviations are spelled in full.

5. The word ‘woman’ is replaced by ‘patient’ to standardize the patient’s gender and to be independent of the guideline text.

2. Stop words removal

In the next step so-called stop words are deleted. These are very frequent words which are not so important for the comparison of the texts because they have little information value, e.g., ‘and’, ‘for’, ‘in’ [6].

3. Stemming

Then each word gets stemmed which means that it is reduced to its word stem, e.g., ‘examining’ becomes ‘examine’, ‘ketonuria’ becomes ‘ketonuria’. Stemming generally cuts off the end of a word and removes derivational affixes. Another quite similar technique would be lemmatization which uses a vocabulary and morphological analysis of a word. It returns the base or dictionary form of a word (called the lemma), e.g., ‘examining’ becomes ‘examine’, ‘ketonuria’ becomes ‘ketonurium’ [29].

4. Spell-checking

If a word cannot be stemmed it is assumed that this word is misspelled. Therefore such a word is spell-checked and then the corrected word gets stemmed.

5. Synonyms of verbs

Synonyms of verbs are also identified to be used for the subsequent comparison.

With these few steps the activity labels are standardized and better suitable for comparison. See Table 3.1 for an example of the above described steps.

original label	**preprocessed label**
If the patient was NOT COMPLIANT with the prescribed diet the nurse insists the patient on the importance of eating enough carbohydrates	*patient compliant prescribe diet nurse insist patient importance eat enough carbohydrate*

Table 3.1: Example of preprocessing activity labels.

3.2.2 The comparison

We only compare plan names that have the same level (e.g., `hybrid-asbru-plan`). For the actual comparison of the names similarity metrics are used (Levenshtein distance, Smith-Waterman edit distance and Jaccard similarity using n-grams). Because of the challenges explained in Section 3.1 a simple string comparison is not suitable.

The Levenshtein distance ($dis_{levenshtein}$) is defined as the minimum number of character insertions, deletions and substitutions needed to transform *word 1* into *word 2*. It works best for strings with (nearly) similar length [9].
The Smith-Waterman edit distance ($dis_{smith-waterman}$) is a string comparison technique originally developed to find the optimal alignment between biological sequences (e.g., DNA, protein sequences). It is similar to the Levenshtein edit distance with the addition that it allows gaps and also character specific match scores and costs. The five basic operations (exact match between two characters, approximate match between two similar characters, mismatch between two different characters, gap start penalty, gap continuation penalty) each have a different match score. The final best score is the highest value in the dynamic programming score matrix and with this a similarity value is calculated. The Smith-Waterman edit distance works very well for compound names that contain initials or abbreviated names because it allows gaps [10].
Using n-grams, a string is split into short substrings of length n (usually $n = 2$ (Bigrams) or $n = 3$ (Trigrams)). The Jaccard similarity using n-grams ($sim_{jaccard}$) is defined as the number of common n-grams (intersection) divided by the number of all n-grams (union) [6]:

$$sim_{jaccard}(word1, word2) = \frac{|ngrams(word1) \cap ngrams(word2)|}{|ngrams(word1) \cup ngrams(word2)|}$$

These three similarity metrics are combined. Two plan names are defined as similar if the three similarity metrics reach certain thresholds. I figured out the thresholds by experimenting and testing different values and looking at the resulting values of corresponding plans of artificially generated and existing data (in the domain of Gestational Diabetes Management [3]).
Further, I differentiate according to the plan name length. If the plan name is a single word it has high thresholds and is only compared to short names (one to three words). As mentioned above, I also use synonyms of verbs for the comparison of plan names. When comparing a single word to one to three words I look for common synonyms. Even

if the two names are below the thresholds, I consider them as a match if they have a synonym in common. For short names (one to three words) the thresholds are higher than for middle (four to six words) and long (seven and more words) names.
The following list shows how the three similarity metrics are combined and what their approximate thresholds are:

- single word: $((dis_{levenshtein} \geq\sim 0,4)$ `AND` $(sim_{jaccard} \geq\sim 0,4))$ `OR` $(dis_{smith-waterman} = 1)$
- one to three words: $((dis_{levenshtein} \geq\sim 0,5)$ `AND` $(sim_{jaccard} \geq\sim 0,5))$ `OR` $(dis_{smith-waterman} = 1)$
- four to six words: $((dis_{levenshtein} \geq\sim 0,2)$ `AND` $(sim_{jaccard} \geq\sim 0,2)$ `AND` $(dis_{smith-waterman} \geq\sim 0,4))$ `OR` $(dis_{smith-waterman} \geq\sim 0,6)$
- seven and more words: $((dis_{levenshtein} \geq\sim 0,4)$ `AND` $(sim_{jaccard} \geq\sim 0,4))$ `OR` $((dis_{levenshtein} \geq\sim 0,1)$ `AND` $(dis_{smith-waterman} \geq\sim 0,5))$

3.3 Comparison of the process structure of the models

After comparing the plan names, in the next step the process structure of the plans is compared. It is examined whether plans that correspond due to their plan names also correspond in their structure.
We will use workflow patterns [47] to describe the process structure of the models.
A pattern is the abstraction from a concrete form which keeps recurring in specific nonarbitrary contexts [48].
Workflows are case-based, i.e., every piece of work is executed for a specific case. Examples for cases are an examination, a specific medical test, an order or a request for information. The goal of workflow management is to handle cases as efficiently and effectively as possible. A workflow process is designed to handle similar cases. Cases are handled by executing tasks in a specific order. The workflow process definition specifies which tasks need to be executed and in what order. Since tasks are executed in a specific order, it is useful to identify conditions which correspond to causal dependencies between tasks. Each task has pre- and postconditions: the preconditions should hold before the task is executed and the postconditions should hold after execution of the task. A task which needs to be executed for a specific case is called a work item. Most work items are executed by a resource. A resource is either a machine (e.g., a printer, an ECG, a CT or X-ray) or a person (e.g., physician, nurse). A resource class is a group of resources with similar characteristics. If a resource class is based on the capabilities (i.e., functional requirements) of its members, it is called a role. If the classification is based on the structure of the organization, such a resource class is called an organizational unit (e.g., team or department). A work item which is being executed by a specific resource is called an activity. Work items link cases and tasks. Activities link cases, tasks and resources [45].

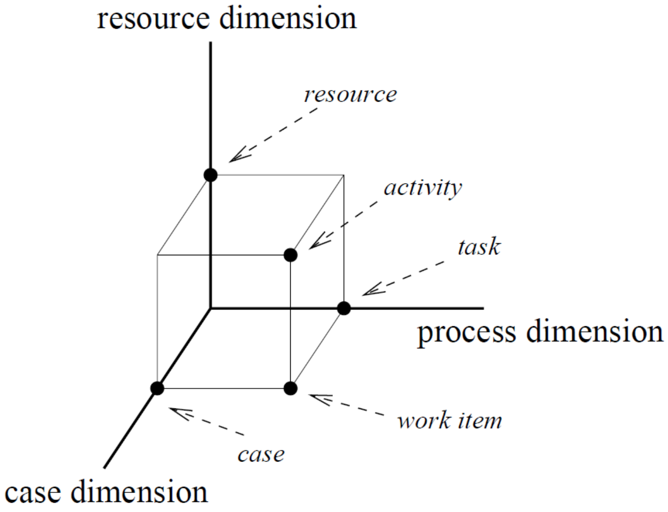


Figure 3.3: Three dimensional view of a workflow [45].

Figure 3.3 shows that a workflow has three dimensions: The case dimension signifies the fact that all cases are handled individually. Clearly the cases influence each other indirectly via the sharing of resources and data. In the process dimension, the workflow process, i.e., the tasks and the routing along these tasks, is specified. The process dimension defines which tasks need to be executed and in what order. In the resource dimension, the resources are grouped into roles and organizational units. A workflow can be visualized as a number of dots in the three dimensional view shown in Figure 3.3. Each dot represents either a work item (case + task) or an activity (case + task + resource) [45].
The ultimate goal of workflow management is to make sure that the proper activities are executed by the right person at the right time [45].
In general, we can distinguish three fundamental relations between activities of a process model. The execution of two activities might happen either in strict order, exclusively or in interleaving order. These relations state potential dependencies [52].

Workflow patterns correspond to routing constructs encountered when modeling and analyzing workflows. They are typically realized in a specific language using one or more constructs available for this language. However, several patterns are difficult, if not impossible, to realize because sometimes workflow constructs available for a given language are not sufficient to realize a given pattern [48].

For example: Basic Control-Flow Patterns are a class of patterns that capture elementary aspects of process control: Sequence, Parallel Split (AND-split), Synchronization (AND-join), Exclusive Choice (XOR-split), Simple Merge (XOR-join) (see Figure 3.4) [36, 48].

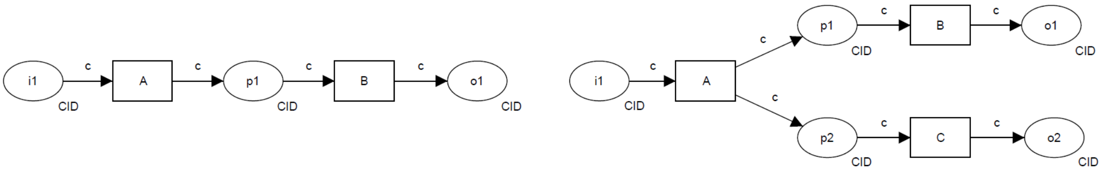


Figure 3.4: Sequence pattern (left) and Parallel Split pattern (right) (the patterns are illustrated using the Coloured Petri-Net formalism [36]).

Another series of patterns characterizes more complex branching and merging concepts: Multi-choice (OR-split), Synchronizing Merge, Multi-merge, Discriminator [36, 48]. However, there exist many more kinds of patterns, cf. [36, 48].
For some of the above mentioned patterns Asbru provides more than one possibility for modeling. This has to be taken into account when comparing models.

There are 8 basic patterns and many more complex patterns (43 patterns in total according to [33]). However, just a few of these pattern types appear in clinical guidelines. In general, only about half of the workflow patterns used in business process modeling languages are supported by CIG languages [33]. Furthermore, only about half of the workflow patterns are applicable for modeling CPGs [25]. But usually there are a lot of patterns (of the same type) present in a guideline. According to [43] the most frequent patterns appearing in a guideline are Sequence and Exclusive Choice. I decided to exemplarily look at the Exclusive Choice pattern because it is more complex as it is linked with a condition.

3.3.1 The Exclusive Choice pattern

First of all I want to give a brief overview on the Exclusive Choice pattern for better understanding.
The Exclusive Choice pattern directs the control flow to a particular task. It depends on a logical condition based on a value of specific data or on a user decision. ‘In the medical domain, this pattern allows enabling a particular action under certain clinical circumstances. It also allows the choice among alternative courses of action that is common in CPGs’ [25]. The Exclusive Choice pattern is usually followed by a Simple Merge joining the branches to the subsequent task [25]. Figure 3.5 presents the structure of the two patterns combined in BPMN.

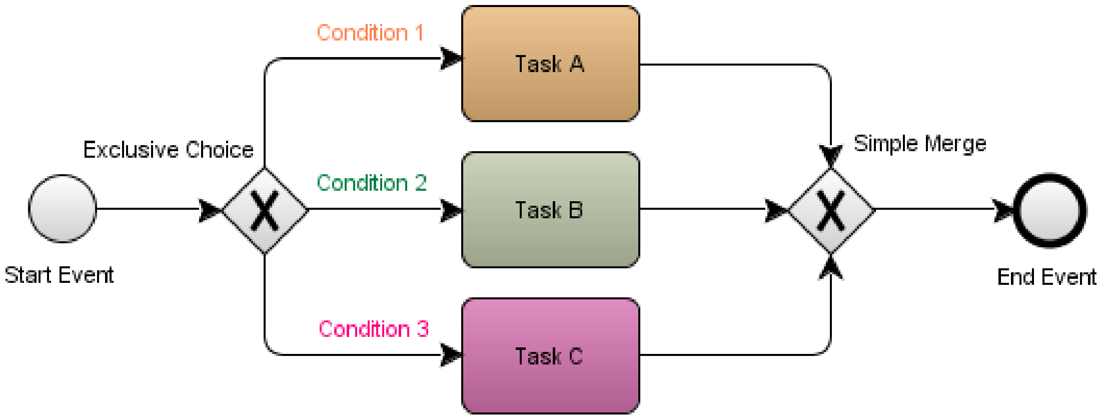


Figure 3.5: BPMN's Exclusive Choice and Simple Merge.

In Asbru there exist two ways to model the Exclusive Choice pattern: the procedural approach and the declarative approach [25].

- The procedural approach: The Exclusive Choice pattern is modeled using the `if-then-else` construct [25]. Figure 3.6 shows the structure of this pattern and Listing 3.2 presents an example of the Exclusive Choice in GESHER, a tool for editing computerized CPGs.

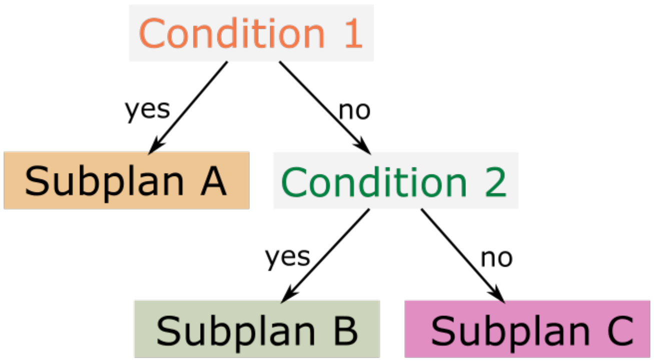


Figure 3.6: Structure of procedural approach of the Exclusive Choice pattern.

```
<hybrid-asbru-plan name="Exclusive Choice">
    <plan-body>
      <semi-formal-plan-body>
         <single-action>
            <if-then-else>                <!-- if -->
               <simple-expression name="Condition 1">
                  <concept name="Condition 1" />
               </simple-expression>
               <then-branch>              <!-- then -->
```

```
                        <plan-activation>
                            <plan-schema>
                                <plan-pointer name="Subplan A" />
                            </plan-schema>
                        </plan-activation>
                    </then-branch>
                    <else-branch>                   <!-- else -->
                        <plan-activation>
                            <plan-schema>
                                <plan-pointer name="Subplan B" />
                            </plan-schema>
                        </plan-activation>
                    </else-branch>
                </if-then-else>
            </single-action>
        </semi-formal-plan-body>
    </plan-body>
</hybrid-asbru-plan>
```

Listing 3.2: Example of `if-then-else`.

- The declarative approach: In this case alternatives are modeled as subplans and all subplans have mutually-exclusive `filter-preconditions` that ensure the execution of only one subplan. When one subplan is executed the parent plan completes and thereby prevents the execution of the remaining subplans. This is achieved by using the expression `wait-for=one` in the parent plan [25]. Figure 3.7 shows the structure of this pattern and Figure 3.8 shows the representation in AsbruView. Listing 3.3 presents an example of the Exclusive Choice in LASSIE, a tool for modeling computerized CPGs.

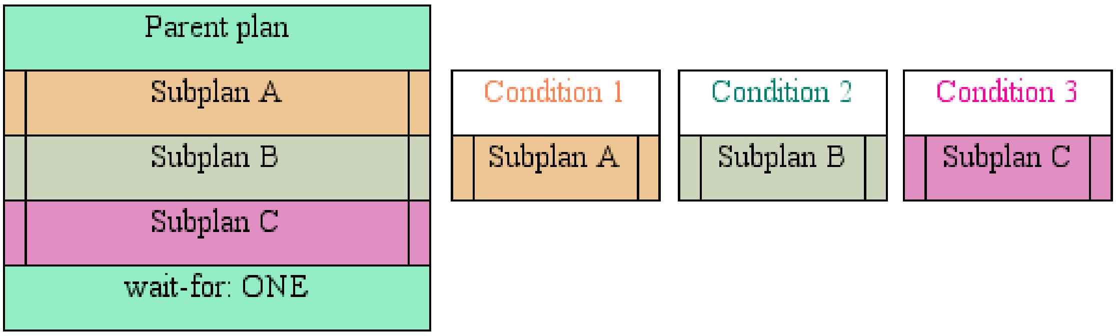


Figure 3.7: Structure of declarative approach of the Exclusive Choice pattern.

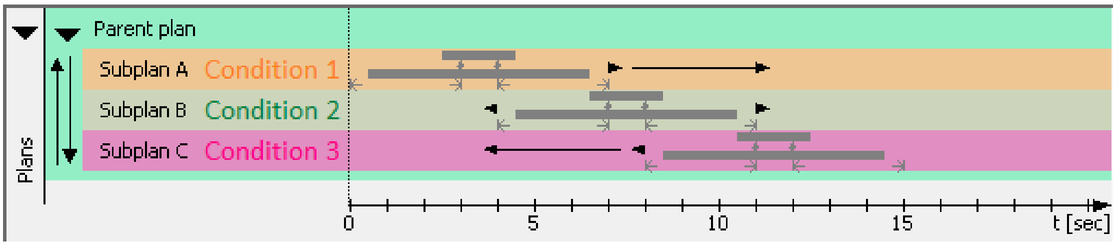


Figure 3.8: Structure of declarative approach of the Exclusive Choice pattern in Asbru-View.

```
<hybrid-asbru-plan name="Exclusive Choice">     <!-- Parent plan -->
    <plan-body>
        <subplans type="any-order">
            <plan-activation>
                <plan-schema>
                    <plan-pointer name="Subplan A" />
                </plan-schema>
            </plan-activation>
            <plan-activation>
                <plan-schema>
                    <plan-pointer name="Subplan B" />
                </plan-schema>
            </plan-activation>
            <wait-for>
                <one/>
            </wait-for>
        </subplans>
    </plan-body>
</hybrid-asbru-plan>

<hybrid-asbru-plan name="Subplan A">   <!-- Subplan A -->
    <conditions>
        <filter-precondition>          <!-- filter-precondition -->
            <concept name="Condition 1" />
        </filter-precondition>
    </conditions>
    <plan-body>
        <single-action>
            <plan-activation>
                <drug-prescription name="DO SOMETHING" />
            </plan-activation>
        </single-action>
    </plan-body>
</hybrid-asbru-plan>
```

Listing 3.3: Example of `filter-precondition`.

`if-then-else` and `filter-precondition` do not match linguistically, but they mean the same.

3.3.2 The comparison

Based on the description of the two different approaches to model the Exclusive Choice pattern in the previous section, one can see that the structure cannot be compared directly. In the procedural approach the condition for a subplan is present in the respective parent plan. In contrast, in the declarative approach the condition for a subplan is located directly in the subplan and the respective parent plan only points to the subplan.
So in order to be able to compare the structure of two models that use these two different approaches, the structure of the models has to be standardized.

The procedural approach can be adapted to be better comparable with the declarative approach. The `if` label is the condition for the subplan with the `then` label. Consequently, the negated `if` label is the condition for the subplan with the `else` label. See Table 3.2 for a description of the adaptation of the procedural approach.

procedural apporach	**adapted procedural approach**
`if` *Condition 1*	
`then` Subplan A	Subplan A *Condition 1*
`else` Subplan B	Subplan B NOT *Condition 1*

Table 3.2: Adaptation of the procedural approach.

So the plans using the procedural approach have to be rewritten in the way that each subplan also contains the condition of the parent plan leading to this subplan. See Listing 3.4 for an example of the adapted procedural approach.

```
<hybrid-asbru-plan name="Exclusive Choice">     <!-- Parent plan -->
    <plan-body>
        <subplans>
            <plan-activation>
                <plan-schema>
                    <plan-pointer name="Subplan A" />
                </plan-schema>
            </plan-activation>
            <plan-activation>
                <plan-schema>
                    <plan-pointer name="Subplan B" />
                </plan-schema>
            </plan-activation>
        </subplans>
    </plan-body>
</hybrid-asbru-plan>

<hybrid-asbru-plan name="Subplan A">   <!-- Subplan A -->
```

```
    <conditions>                       <!-- condition -->
        <concept name="Condition 1" />
    </conditions>
    <plan-body>
        <single-action>
            <plan-activation>
                <drug-prescription name="ACTION" />
            </plan-activation>
        </single-action>
    </plan-body>
</hybrid-asbru-plan>

<hybrid-asbru-plan name="Subplan B">   <!-- Subplan B -->
    <conditions>                       <!-- negated condition -->
        <concept name="NOT Condition 1" />
    </conditions>
    <plan-body>
        <single-action>
            <plan-activation>
                <drug-prescription name="ACTION" />
            </plan-activation>
        </single-action>
    </plan-body>
</hybrid-asbru-plan>
```

Listing 3.4: Example of adapted procedural approach.

Thereby the structure of the plans using the adapted procedural approach is similar to the structure of the plans which use the declarative approach. This adaptation makes it possible to compare the structure of two models.

CHAPTER 4

Implementation

In this chapter I introduce the two different tools that were used for generating computerized treatment plan models: GESHER for manually creating models and LASSIE for automatically creating models.
Next, I show how I implemented the methods described in Chapter 3 prototypically and finally I present the evaluation of the obtained results.

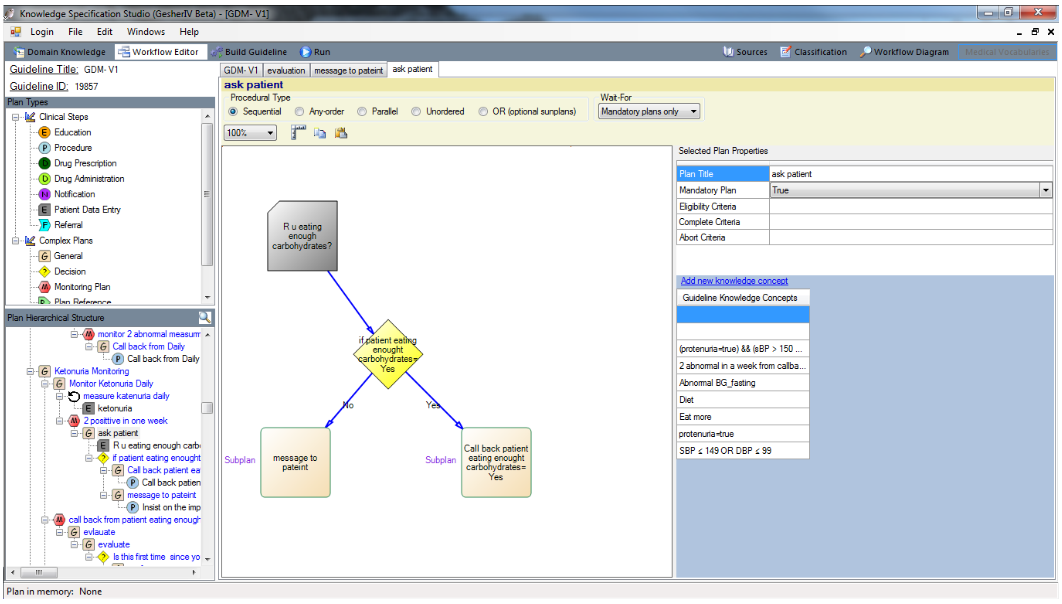
4.1 GESHER and LASSIE

We work with two tools for creating computerized treatment plan models: GESHER and LASSIE.

4.1.1 GESHER

GESHER is a client application developed in C#. Modeling a CIG with the GESHER tool is based on the CPG document as well as on the expertise coming from medical experts and knowledge engineers. The experts orientate by the guideline document and adapt it. A lot of information does not have its origin in the original guideline document but is created by the domain experts as a result of their expertise and experience. GESHER has three important services: specification of procedural clinical knowledge, specification of declarative clinical knowledge and exploration of the knowledge embodied by a given clinical guideline [20]. In GESHER the procedural knowledge is directly modeled using a model-centric approach (see Figure 4.1). A hybrid approach applying both document-centric and model-centric techniques is used for modeling the declarative knowledge (see Figure 4.2) [3].
Figure 4.1 shows the hierarchical plan builder in GESHER that is used for specifying the procedural aspects of the guideline. The user selects plans from different types (*upper left box*) and adds them to the hierarchical flow chart (*center box*). For each plan several properties can be defined (*middle right box*) and for composite plans the procedural

aspects are specified (*upper right box*). The sub-plan hierarchy is shown in a tree-view display (*lower left box*). A list of declarative concepts is defined (*lower right box*) [3, 20]. For examples of GESHER plans see Appendix A.



Guideline Title: GDM- V1

Guideline ID: 19857

Plan Types

- Clinical Steps
  - E Education
  - P Procedure
  - D Drug Prescription
  - D Drug Administration
  - N Notification
  - E Patient Data Entry
  - F Referral
- Complex Plans
  - G General
  - ? Decision
  - M Monitoring Plan
  - P Plan Reference

Plan Hierarchical Structure

- M monitor 2 abnormal measurm
  - G Call back from Daily
    - P Call back from Daily
- G Ketonuria Monitoring
  - G Monitor Ketonuria Daily
    - measure katenuria daily
      - E ketonuria
    - M 2 posittive in one week
      - G ask patient
        - E R u eating enough carb
        - ? if patient eating enought
          - G Call back patient ea
            - P Call back patien
          - G message to pateint
            - P Insist on the imp
  - M call back from patient eating enough
    - G evlauate
      - G evaluate
        - ? Is this first time since yo

ask patient

Procedural Type: ● Sequential ○ Any-order ○ Parallel ○ Unordered ○ OR (optional sunplans)

Wait-For: Mandatory plans only

Selected Plan Properties

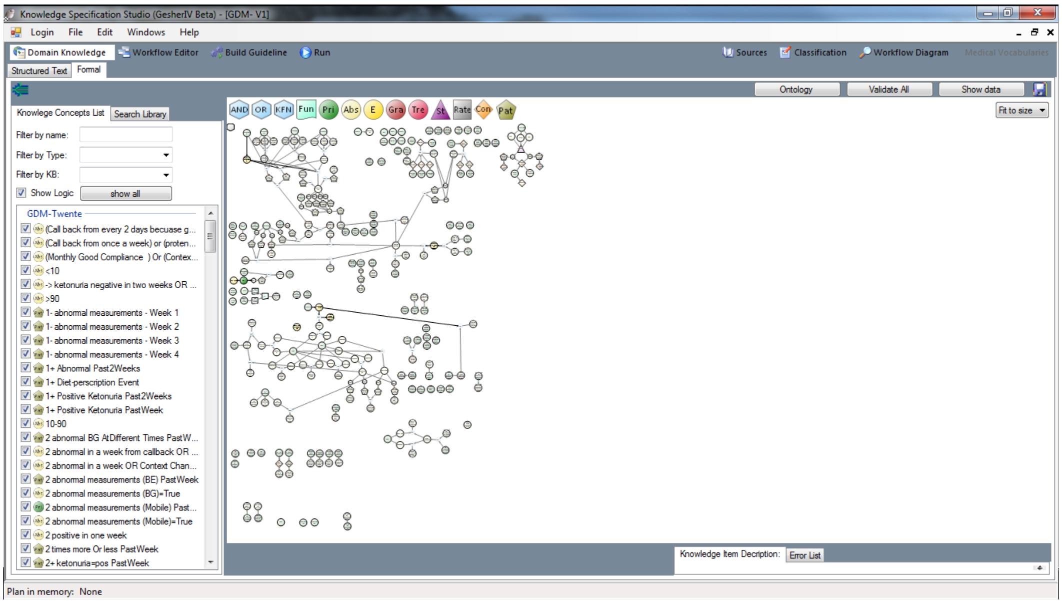
Plan Title	ask patient
Mandatory Plan	True
Eligibility Criteria	
Complete Criteria	
Abort Criteria	

Add new knowledge concept

Guideline Knowledge Concepts
(protenuria=true) && (sBP > 150 ...
2 abnormal in a week from callba...
Abnormal BG_fasting
Diet
Eat more
protenuria=true
SBP ≤ 149 OR DBP ≤ 99

Plan in memory: None

Figure 4.1: GESHER user interface - procedural knowledge.



Filter by name:

Filter by Type:

Filter by KB:

☑ Show Logic

GDM-Twente

- ☑ (Call back from every 2 days becuase g...
- ☑ (Call back from once a week) or (proten...
- ☑ (Monthly Good Compliance) Or (Contex...
- ☑ <10
- ☑ -> ketonuria negative in two weeks OR ...
- ☑ >90
- ☑ 1- abnormal measurements - Week 1
- ☑ 1- abnormal measurements - Week 2
- ☑ 1- abnormal measurements - Week 3
- ☑ 1- abnormal measurements - Week 4
- ☑ 1+ Abnormal Past2Weeks
- ☑ 1+ Diet-perscription Event
- ☑ 1+ Positive Ketonuria Past2Weeks
- ☑ 1+ Positive Ketonuria PastWeek
- ☑ 10-90
- ☑ 2 abnormal BG AtDifferent Times PastW...
- ☑ 2 abnormal in a week from callback OR ...
- ☑ 2 abnormal in a week OR Context Chan...
- ☑ 2 abnormal measurements (BE) PastWeek
- ☑ 2 abnormal measurements (BG)=True
- ☑ 2 abnormal measurements (Mobile) Past...
- ☑ 2 abnormal measurements (Mobile)=True
- ☑ 2 positive in one week
- ☑ 2 times more Or less PastWeek
- ☑ 2+ ketonuria=pos PastWeek

Plan in memory: None

Figure 4.2: GESHER user interface - declarative knowledge.

4.1.2 LASSIE

LASSIE is an acronym for 'modeLing treAtment proceSSes using Information Extraction'. LASSIE is developed in Java and uses the NLP framework GATE. It uses IE techniques to automatically identify procedural information in medical natural language text (i.e., CPGs) and consists of manually generated rules using a medical dictionary to extract information. LASSIE's single source of information is the original guideline document itself. In the first version LASSIE's multi-step process uses intermediate representations to transform clinical guidelines and protocols (CGPs) into a formal representation language. The intermediate representations are XML-based documents that are only used to semi-automatically generate a formal representation in any guideline representation language. Therefore LASSIE automates parts of the modeling process and structures the guideline information. LASSIE identifies text chunks in the guideline document that refer to procedural knowledge (i.e., activities and associated conditions), structures them in a hierarchy and transforms them into Asbru [3, 23].
Figure 4.3 describes the flow chart of LASSIE's steps [3].

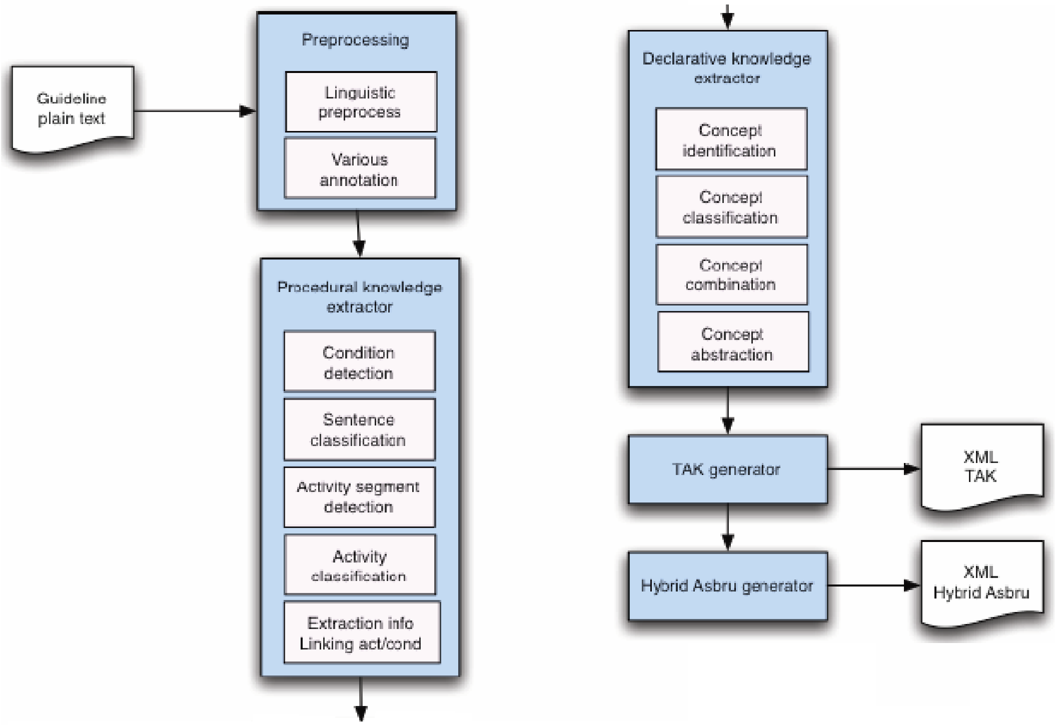


Figure 4.3: Flow chart of LASSIE's knowledge generation [3].

Leveraging LASSIE independently of the clinical specialty machine learning models for extracting parts of the procedural knowledge were trained and the UMLS is used to provide an extensive set of clinical terminology systems [3].
For examples of LASSIE plans see Appendix B.

4.2 Implementation of the method

As a basis for the implementation of the method presented in Chapter 3 we used a guideline for gestational diabetes mellitus as a proof of concept. In this guideline gestational diabetes mellitus (GDM) ‘is defined as carbohydrate intolerance that begins or is first recognized during pregnancy’ [35].
See Appendix C for more information about the GDM guideline and the text itself.
See Figure 4.4 for a tree visualization of LASSIE’s plans of the GDM guideline and Figure 4.5 for a tree visualization of GESHER’s plans of the GDM guideline.

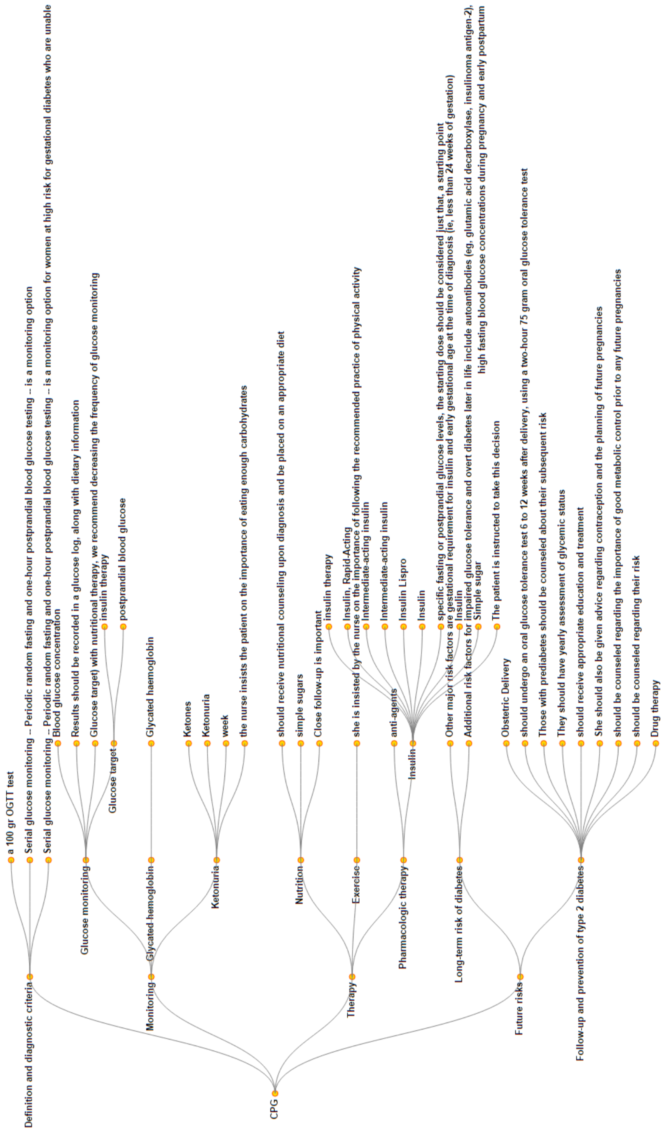
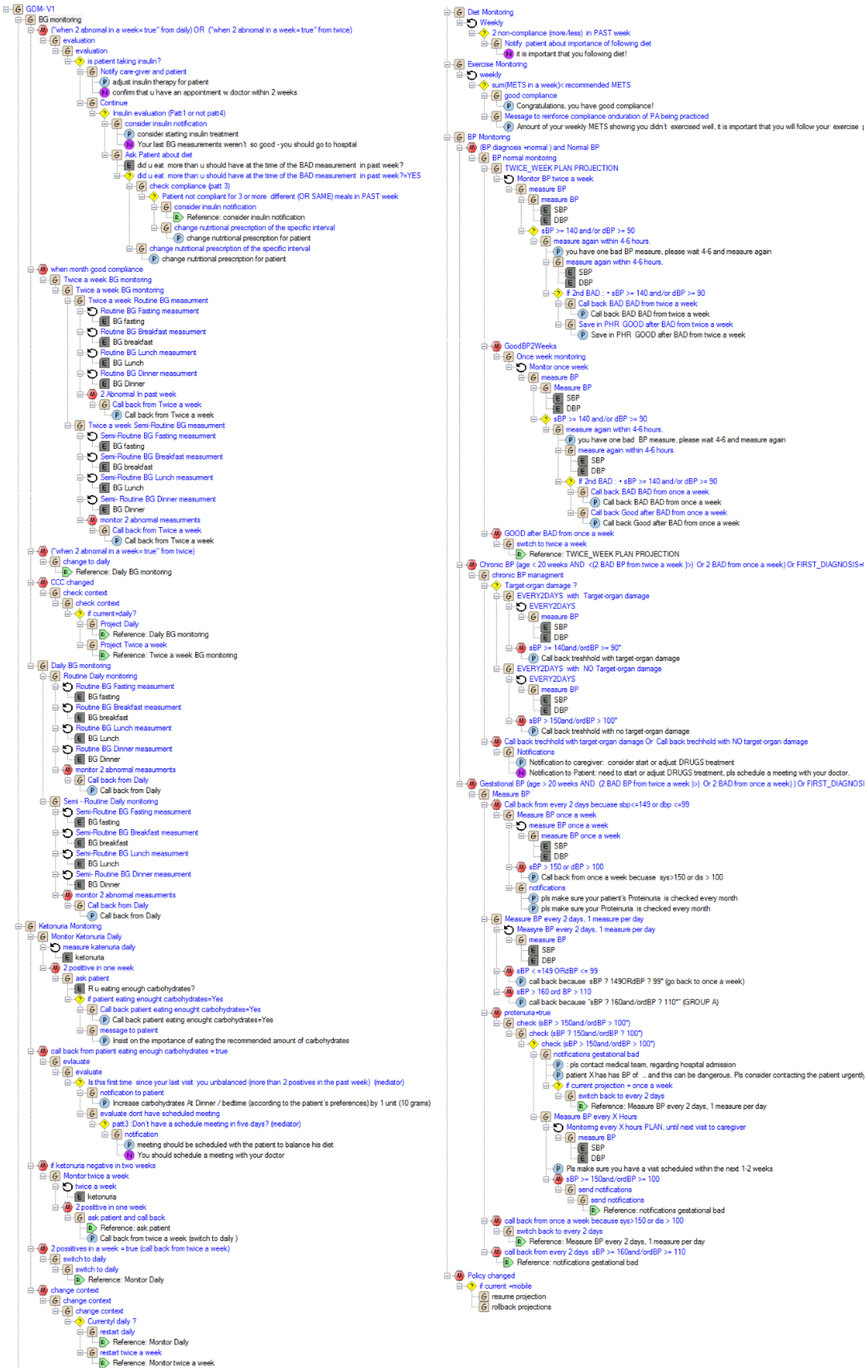

CPG
Definition and diagnostic criteria
a 100 gr OGTT test
Serial glucose monitoring -- Periodic random fasting and one-hour postprandial blood glucose testing -- is a monitoring option
Serial glucose monitoring -- Periodic random fasting and one-hour postprandial blood glucose testing -- is a monitoring option for women at high risk for gestational diabetes who are unable
Monitoring
Glucose monitoring
Blood glucose concentration
Results should be recorded in a glucose log, along with dietary information
Glucose target) with nutritional therapy, we recommend decreasing the frequency of glucose monitoring
Glucose target
insulin therapy
postprandial blood glucose
Glycated hemoglobin
Glycated haemoglobin
Ketonuria
Ketones
Ketonuria
week
the nurse insists the patient on the importance of eating enough carbohydrates
Therapy
Nutrition
should receive nutritional counseling upon diagnosis and be placed on an appropriate diet
simple sugars
Close follow-up is important
Exercise
she is insisted by the nurse on the importance of following the recommended practice of physical activity
Pharmacologic therapy
anti-agents
Insulin
insulin therapy
Insulin, Rapid-Acting
Intermediate-acting insulin
Intermediate-acting insulin
Insulin Lispro
Insulin
specific fasting or postprandial glucose levels, the starting dose should be considered just that, a starting point
Insulin
Simple sugar
The patient is instructed to take this decision
Future risks
Long-term risk of diabetes
Other major risk factors are gestational requirement for insulin and early gestational age at the time of diagnosis (ie, less than 24 weeks of gestation)
Additional risk factors for impaired glucose tolerance and overt diabetes later in life include autoantibodies (eg, glutamic acid decarboxylase, insulinoma antigen-2), high fasting blood glucose concentrations during pregnancy and early postpartum
Follow-up and prevention of type 2 diabetes
Obstetric Delivery
should undergo an oral glucose tolerance test 6 to 12 weeks after delivery, using a two-hour 75 gram oral glucose tolerance test
Those with prediabetes should be counseled about their subsequent risk
They should have yearly assessment of glycemic status
should receive appropriate education and treatment
She should also be given advice regarding contraception and the planning of future pregnancies
should be counseled regarding the importance of good metabolic control prior to any future pregnancies
should be counseled regarding their risk
Drug therapy

Figure 4.4: Tree view of LASSIE’s plans of the GDM guideline.



```
GDM- V1
  BG monitoring
    ("when 2 abnomal in a week= true" from daily) OR  ("when 2 abnomal in a week= true" from twice)
      evaluation
        evaluation
          is patient taking insulin?
            Notify care-giver and patient
              adjust insulin therapy for patient
              confirm that u have an appointment w doctor within 2 weeks
            Continue
              Insulin evaluation (Patt 1 or not patt4)
                consider insulin notification
                  consider starting insulin treatment
                  Your last BG measurements weren't  so good - you should go to hospital
                Ask Patient about diet
                  did u eat  more than u should have at the time of the BAD measurement  in past week?
                  did u eat  more than u should have at the time of the BAD measurement  in past week?=YES
                    check compliance (patt 3)
                      Patient not compliant for 3 or more  different (OR SAME) meals in PAST week
                        consider insulin notification
                          Reference: consider insulin notification
                        change nutritional prescription of the specific interval
                          change nutritional prescription for patient
                    change nutritional prescription of the specific interval
                      change nutritional prescription for patient
    when month good compliance
      Twice a week BG monitoring
        Twice a week BG monitoring
          Twice a week Routine BG measurment
            Routine BG Fasting measurment
              BG fasting
            Routine BG Breakfast measurment
              BG breakfast
            Routine BG Lunch measurment
              BG Lunch
            Routine BG Dinner measurment
              BG Dinner
            2 Abnormal In past week
              Call back from Twice a week
                Call back from Twice a week
          Twice a week Semi-Routine BG measurment
            Semi-Routine BG Fasting measurment
              BG fasting
            Semi-Routine BG Breakfast measurment
              BG breakfast
            Semi-Routine BG Lunch measurment
              BG Lunch
            Semi- Routine BG Dinner measurment
              BG Dinner
            monitor 2 abnormal measurments
              Call back from Twice a week
                Call back from Twice a week
    ("when 2 abnomal in a week= true" from twice)
      change to daily
        Reference: Daily BG monitoring
    CCC changed
      check context
        check context
          if current=daily?
            Project Daily
              Reference: Daily BG monitoring
            Project Twice a week
              Reference: Twice a week BG monitoring
    Daily BG monitoring
      Routine Daily monitoring
        Routine BG Fasting measurment
          BG fasting
        Routine BG Breakfast measurment
          BG breakfast
        Routine BG Lunch measurment
          BG Lunch
        Routine BG Dinner measurment
          BG Dinner
        monitor 2 abnormal measurments
          Call back from Daily
            Call back from Daily
      Semi - Routine Daily monitoring
        Semi-Routine BG Fasting measurment
          BG fasting
        Semi-Routine BG Breakfast measurment
          BG breakfast
        Semi-Routine BG Lunch measurment
          BG Lunch
        Semi- Routine BG Dinner measurment
          BG Dinner
        monitor 2 abnormal measurments
          Call back from Daily
            Call back from Daily
  Ketonuria Monitoring
    Monitor Ketonuria Daily
      measure katenuria daily
        ketonuria
      2 posittive in one week
        ask patient
          R u eating enough carbohydrates?
          if patient eating enought carbohydrates=Yes
            Call back patient eating enought carbohydrates=Yes
              Call back patient eating enought carbohydrates=Yes
            message to pateint
              Insist on the importance of eating the recommended amount of carbohydrates
    call back from patient eating enough carbohydrates = true
      evlauate
        evaluate
          Is this first time  since your last visit  you unbalanced (more than 2 positives in the past week)  (mediator)
            notification to patient
              Increase carbohydrates At Dinner / bedtime (according to the patient's preferences) by 1 unit (10 grams)
            evaluate dont have scheduled meeting
              patt3 :Don't have a schedule meeting in five days? (mediator)
                notification
                  meeting should be scheduled with the patient to balance his diet
                  You should schedule a meeting with your doctor
    if ketonuria negative in two weeks
      Monitor twice a week
        twice a week
          ketonuria
        2 posittive in one week
          ask patient and call back
            Reference: ask patient
            Call back from twice a week (switch to daily )
    2 possitives in a week = true (call back from twice a week)
      switch to daily
        switch to daily
          Reference: Monitor Daily
    change context
      change context
        change context
          Curentyl daily ?
            restart daily
              Reference: Monitor Daily
            restart twice a week
              Reference: Monitor twice a week
  Diet Monitoring
    Weekly
      2 non-compliance (more/less)  in PAST week
        Notify  patient about importance of following diet
          it is important that you following diet!
  Exercise Monitoring
    weekly
      sum(METS in a week)< recommended METS
        good compliance
          Congratulations, you have good compliance!
        Message to reinforce compliance onduration of PA being practiced
          Amount of your weekly METS showing you didn't  exercised well, it is important that you will follow your  exercise
  BP Monitoring
    (BP diagnosis =normal ) and Normal BP
      BP normal monitoring
        TWICE_WEEK PLAN PROJECTION
          Monitor BP twice a week
            measure BP
              measure BP
                SBP
                DBP
              sBP >= 140 and/or dBP >= 90
                measure again within 4-6 hours.
                  you have one bad BP measure, please wait 4-6 and measure again
                  measure again within 4-6 hours.
                    SBP
                    DBP
                  If 2nd BAD : • sBP >= 140 and/or dBP >= 90
                    Call back BAD BAD from twice a week
                      Call back BAD BAD from twice a week
                    Save in PHR  GOOD after BAD from twice a week
                      Save in PHR  GOOD after BAD from twice a week
        GoodBP2Weeks
          Once week monitoring
            Monitor once week
              measure BP
                Measure BP
                  SBP
                  DBP
                sBP >= 140 and/or dBP >= 90
                  measure again within 4-6 hours.
                    you have one bad  BP measure, please wait 4-6 and measure again
                    measure again within 4-6 hours.
                      SBP
                      DBP
                    If 2nd BAD : • sBP >= 140 and/or dBP >= 90
                      Call back BAD BAD from once a week
                        Call back BAD BAD from once a week
                      Call back Good after BAD from once a week
                        Call back Good after BAD from once a week
        GOOD after BAD from once a week
          switch to twice a week
            Reference: TWICE_WEEK PLAN PROJECTION
    Chronic BP (age < 20 weeks AND  <(2 BAD BP from twice a week )>)  Or 2 BAD from once a week) Or FIRST_DIAGNOSIS=
      chronic BP managment
        Target-organ damage ?
          EVERY2DAYS  with  Target-organ damage
            EVERY2DAYS
              measure BP
                SBP
                DBP
            sBP >= 140and/ordBP >= 90*
              Call back treshhold with target-organ damage
          EVERY2DAYS  with  NO  Target-organ damage
            EVERY2DAYS
              measure BP
                SBP
                DBP
            sBP > 150and/ordBP > 100*
              Call back treshhold with no target-organ damage
        Call back trechhold with target-organ damage Or  Call back trechhold with NO target-organ damage
          Notifications
            Notification to caregiver:  consider start or adjust DRUGS treatment
            Notification to Patient: need to start or adjust DRUGS treatment, pls schedule a meeting with your doctor.
    Geststional BP (age > 20 weeks AND  (2 BAD BP from twice a week )>)  Or 2 BAD from once a week) ) Or FIRST_DIAGNOSI
      Measure BP
        Call back from every 2 days becuase sbp<=149 or dbp <=99
          Measure BP once a week
            measure BP once a week
              measure BP once a week
                SBP
                DBP
            sBP > 150 or dBP > 100
              Call back from once a week becuase  sys>150 or dis > 100
            notifications
              pls make sure your patient's Proteinuria  is checked every month
              pls make sure your Proteinuria  is checked every month
        Measure BP every 2 days, 1 measure per day
          Measyre BP every 2 days, 1 measure per day
            measure BP
              SBP
              DBP
          sBP < =149 ORdBP <= 99
            call back because  sBP ? 149ORdBP ? 99* (go back to once a week)
          sBP > 160 ord BP > 110
            call back because "sBP ? 160and/ordBP ? 110*" (GROUP A)
        protenuria=true
          check (sBP > 150and/ordBP > 100*)
            check (sBP ? 150and/ordBP ? 100*)
              check (sBP > 150and/ordBP > 100*)
                notifications gestational bad
                  : pls contact medical team, regarding hospital admission
                  patient X has has BP of  ... and this can be dangerous. Pls consider contacting the patient urgently
                  if current projection = once a week
                    switch back to every 2 days
                      Reference: Measure BP every 2 days, 1 measure per day
                Measure BP every X Hours
                  Monitoring every X hours PLAN, until next visit to caregiver
                    measure BP
                      SBP
                      DBP
                  Pls make sure you have a visit scheduled within the next 1-2 weeks
                  sBP >= 150and/ordBP >= 100
                    send notifications
                      send notifications
                        Reference: notifications gestational bad
        call back from once a week because sys>150 or dis > 100
          switch back to every 2 days
            Reference: Measure BP every 2 days, 1 measure per day
        call back from every 2 days  sBP >= 160and/ordBP >= 110
          Reference: notifications gestational bad
  Policy changed
    if current =mobile
      resume projection
      rollback projections
```

Figure 4.5: GESHER's plan hierarchy according to the GDM guideline.

]

4.2.1 Comparison of single activities of the models

I used the well-known programming language Java to implement the various steps of the comparison of single activities.

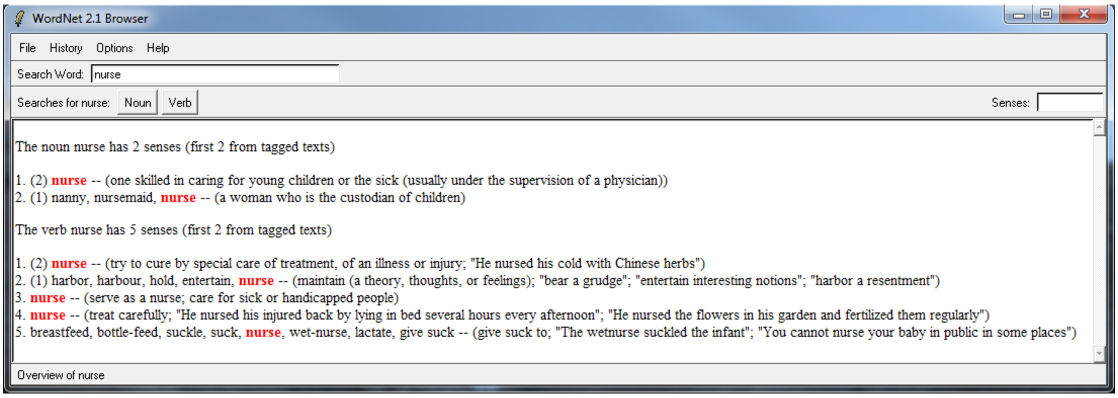
1. Preprocessing

For the stop word removal I used a list of common English words[1], see Listing 4.1.

```
'tis 'twas a able about across after ain't all almost also am among an and any are
aren't as at be because been but by can can't cannot could could've couldn't dear did
didn't do does doesn't don't either else ever every for from get got had has hasn't
have he he'd he'll he's her hers him his how how'd how'll how's however i i'd i'll i'm
i've if in into is isn't it it's its just least let like likely may me might might've
mightn't most must must've mustn't my neither no nor not of off often on only or other
our own rather said say says shan't she she'd she'll she's should should've shouldn't
since so some than that that'll that's the their them then there there's these they
they'd they'll they're they've this tis to too twas us wants was wasn't we we'd we'll
we're were weren't what what'd what's when when when'd when'll when's where where'd
where'll where's which while who who'd who'll who's whom why why'd why'll why's will
with won't would would've wouldn't yet you you'd you'll you're you've your
```

Listing 4.1: List of stop words.

Next, each word gets stemmed. I tried both stemming and lemmatization. However, stemming yielded better results in the case of this presentation of the problem (cf. the 'ketonuria' example in Section 3.2.1) and it is also faster. I stemmed the words using WordNet which is a lexical database of English words [44]. Figure 4.6 shows an example of a WordNet database entry.



Search Word: nurse

Searches for nurse:

Senses:

The noun nurse has 2 senses (first 2 from tagged texts)

1. (2) **nurse** -- (one skilled in caring for young children or the sick (usually under the supervision of a physician))
2. (1) nanny, nursemaid, **nurse** -- (a woman who is the custodian of children)

The verb nurse has 5 senses (first 2 from tagged texts)

1. (2) **nurse** -- (try to cure by special care of treatment, of an illness or injury; "He nursed his cold with Chinese herbs")
2. (1) harbor, harbour, hold, entertain, **nurse** -- (maintain (a theory, thoughts, or feelings); "bear a grudge"; "entertain interesting notions"; "harbor a resentment")
3. **nurse** -- (serve as a nurse; care for sick or handicapped people)
4. **nurse** -- (treat carefully; "He nursed his injured back by lying in bed several hours every afternoon"; "He nursed the flowers in his garden and fertilized them regularly")
5. breastfeed, bottle-feed, suckle, suck, **nurse**, wet-nurse, lactate, give suck -- (give suck to; "The wetnurse suckled the infant"; "You cannot nurse your baby in public in some places")

Figure 4.6: Example of a WordNet database entry.

The words that cannot be stemmed are spell-checked using an English medical dictionary and then the program tries to stem the corrected word again. Spell-checking all words at the beginning would be too expensive and time consuming.
I also look for synonyms of verbs to use them for the subsequent comparison. For that I also used WordNet [44].

[1] from `http://www.textfixer.com/resources/common-english-words.php`, accessed: April 5, 2016

2. The comparison

For the implementation of the similarity metrics introduced in Section 3.2.2 I used the Java library SimMetrics[2]. It provides different kinds of similarity and distance metrics.

4.2.2 Comparison of the process structure of the models

The results from the activity comparison are the input for the structure comparison.
I used XSLT templates to unify the modelings to one general model pattern to enable the comparison. As a proof of concept I implemented the unification of the Exclusive Choice pattern that can be modeled in two different ways in Asbru (cf. Section 3.3.1). I accomplished the unification by developing XSLT templates that take XML files as input and can generate output files of different formats or languages (cf. [25]). In our case the output is still XML, but the concepts of the underlying language are changed by preserving the underlying semantics (see Listing 4.2). See Listing 4.3 for an example of an input of the XSLT templates and Listing 4.4 for the respective output of the XSLT templates. And then I compared the equivalent plans I received in the previous step.

```
<xsl:template match="if-then-else">
   <xsl:element name="hybrid-asbru-plan">
      <xsl:attribute name="id">
         <xsl:value-of select="../../../@id" />
      </xsl:attribute>
      <xsl:attribute name="name">
         <xsl:value-of select="../../../@name" />
      </xsl:attribute>
      <xsl:element name="plan-body">
         <xsl:element name="semi-formal-plan-body">
            <xsl:attribute name="type">any-order</xsl:attribute>
            <xsl:element name="wait-for">
               <xsl:element name="one" />
            </xsl:element>
            <xsl:element name="subplans">
               <xsl:for-each select="//then-branch/plan-activation/plan-schema/plan-pointer">
               <!-- specify the child plans -->
                  <xsl:apply-templates select="." mode="plan-activation" />
               </xsl:for-each>
            </xsl:element>
         </xsl:element>
      </xsl:element>
   </xsl:element>
</xsl:template>
```

Listing 4.2: Part of XSLT template to transform procedural to declarative Exclusive Choice pattern.

[2] from https://sourceforge.net/projects/simmetrics/, accessed: April 6, 2016

```
<hybrid-asbru-plan id="19969" name="if patient eating enought carbohydrates
   =Yes">
    <plan-body>
      <semi-formal-plan-body>
         <single-action>
            <if-then-else>
               <simple-expression name="Eating enough carbohydorates=yes">
                  <concept name="Eating enough carbohydorates=yes" />
               </simple-expression>
               <then-branch>
                  <plan-activation>
                     <plan-schema>
                        <plan-pointer id="19972" name="Call back patient
                            eating enought carbohydrates=Yes" />
                     </plan-schema>
                  </plan-activation>
               </then-branch>
               <else-branch>
                  <plan-activation>
                     <plan-schema>
                        <plan-pointer id="19973" name="Insist on the
                            importance of eating the recommended amount of
                            carbohydrates" />
                     </plan-schema>
                  </plan-activation>
               </else-branch>
            </if-then-else>
         </single-action>
      </semi-formal-plan-body>
   </plan-body>
</hybrid-asbru-plan>

<hybrid-asbru-plan id="19972" name="Call back patient eating enought
    carbohydrates=Yes">
    <plan-body>
      <semi-formal-plan-body>
         <single-action>
            <plan-activation>
               <procedure name="Call back patient eating enought
                   carbohydrates=Yes" />
            </plan-activation>
         </single-action>
      </semi-formal-plan-body>
   </plan-body>
</hybrid-asbru-plan>

<hybrid-asbru-plan id="19973" name="Insist on the importance of eating the
    recommended amount of carbohydrates">
   <plan-body>
      <semi-formal-plan-body>
         <single-action>
            <plan-activation>
```

```
               <procedure name="Insist on the importance of eating the recommended amount of carbohydrates" />
            </plan-activation>
         </single-action>
      </semi-formal-plan-body>
   </plan-body>
</hybrid-asbru-plan>
```

Listing 4.3: Example of an input of the XSLT templates.

```
<hybrid-asbru-plan id="19969" name="if patient eating enought carbohydrates=Yes">
    <plan-body>
      <semi-formal-plan-body>
         <subplans type="any-order">
            <wait-for>
               <one/>
            </wait-for>
            <plan-activation>
               <plan-schema>
                  <plan-pointer id="19972" name="Call back patient eating enought carbohydrates=Yes" />
               </plan-schema>
            </plan-activation>
            <plan-activation>
               <plan-schema>
                  <plan-pointer id="19973" name="Insist on the importance of eating the recommended amount of carbohydrates" />
               </plan-schema>
            </plan-activation>
         </subplans>
      </semi-formal-plan-body>
   </plan-body>
</hybrid-asbru-plan>

<hybrid-asbru-plan id="19972" name="Call back patient eating enought carbohydrates=Yes">
    <conditions>
        <filter-precondition>
            <concept name="Eating enough carbohydorates=yes" />
        </filter-precondition>
    </conditions>
    <plan-body>
      <semi-formal-plan-body>
         <single-action>
            <plan-activation>
               <procedure name="Call back patient eating enought carbohydrates=Yes" />
            </plan-activation>
         </single-action>
      </semi-formal-plan-body>
    </plan-body>
```

```
</hybrid-asbru-plan>

<hybrid-asbru-plan id="19973" name="Insist on the importance of eating the
    recommended amount of carbohydrates">
    <conditions>
        <filter-precondition>
            <concept name="NOT Eating enough carbohydorates=yes" />
        </filter-precondition>
    </conditions>
    <plan-body>
      <semi-formal-plan-body>
         <single-action>
            <plan-activation>
               <procedure name="Insist on the importance of eating the
                   recommended amount of carbohydrates" />
            </plan-activation>
         </single-action>
      </semi-formal-plan-body>
    </plan-body>
</hybrid-asbru-plan>
```

Listing 4.4: Example of an output of the XSLT templates.

4.3 Evaluation

For the evaluation we compare the GDM guideline modeled with GESHER and modeled with LASSIE.

First, we evaluate the 'activity' and 'condition' information found by LASSIE to see how much of the original guideline text is present in LASSIE. Second, we evaluate how many of the GESHER plans and LASSIE plans correspond to know if and how good the automatically generated model finds the same information as present in the manually generated model. Finally, the process structure of matching plans is evaluated to examine if corresponding plans also have a corresponding structure.

For the evaluation of the information found by LASSIE and the plans recall and precision measures are used. However, these values cannot be used for the evaluation of the process structure based on workflow patterns.

The recall score measures the ratio of correct information extracted from the text against all the available information present in the text. The precision score measures the ratio of correct information that was extracted against all the information that was extracted [24].

We evaluated our results according to Table 4.1 [28]:

actual (ACT)	the number of slot fillers generated by the system (= COR + PAR + INC)
correct (COR)	the number of correct slot fillers generated by the system
partial (PAR)	the number of partially correct slot fillers generated by the system
incorrect (INC)	the number of incorrect slot fillers generated by the system
missing (MIS)	the number of slot fillers erroneously not generated by the system
possible (POS)	the number of slot fillers according to the key target templates
recall (REC)	the ratio of COR plus 0,5*PAR slot fillers to POS slot fillers
precision (PRE)	the ratio of COR plus 0,5*PAR slot fillers to ACT slot fillers

Table 4.1: Scoring key for the evaluation [28, p. 5].

The F-Measure combines recall and precision to get one value for the evaluation of a system [27]:

$$F = \frac{(\beta^2 + 1) * precision * recall}{\beta^2 * precision + recall}$$

Here β is a weighting parameter. In the case of this evaluation, I chose $\beta = 1$ which is the harmonic mean [27].

4.3.1 Evaluation of 'activity' and 'condition' information found by LASSIE

LASSIE can identify 'activity' and 'condition' text chunks in guidelines, but also use pre-annotated documents for improved processing output.
In this section we want to evaluate how many of the original manually created 'activity' and 'condition' markups are found by LASSIE. For the original manually created markups see Appendix C. The results are displayed in Table 4.2 and 4.3.

actual (ACT)	39
correct (COR)	15
partial (PAR)	12
incorrect (INC)	12
missing (MIS)	22
possible (POS)	49
recall (REC)	0,4286
precision (PRE)	0,5385
F-Measure	0,4773

Table 4.2: Evaluation of LASSIE's 'activity' markups.

actual (ACT)	27
correct (COR)	19
partial (PAR)	4
incorrect (INC)	4
missing (MIS)	17
possible (POS)	40
recall (REC)	0,525
precision (PRE)	0,7778
F-Measure	0,6269

Table 4.3: Evaluation of LASSIE's 'condition' markups.

Originally there are 49 'activity' and 40 'condition' markups in the guideline document. LASSIE finds 39 'activity' and 27 'condition' markups. Of these markups 15 'activity' and

19 ‘condition’ markups are correctly identified, 12 ‘activity’ and 4 ‘condition’ markups are partially correctly identified and 12 ‘activity’ and 4 ‘condition’ markups are wrongly identified. 22 ‘activity’ and 17 ‘condition’ markups of the original markups are not identified at all by LASSIE.
All in all precision (0,5385 and 0,7778) and recall (0,4286 and 0,525) as well as the F-Measure (0,4773 and 0,6269) are very promising for both ‘activity’ and ‘condition’ markups.

4.3.2 Evaluation of the plans

The next step is to evaluate if or how much the plans of the automatically generated model correspond to the plans of the manually generated model.
Each LASSIE plan name is compared to each GESHER plan name according to Section 3.2. This comparison yielded the results displayed in Table 4.4 and 4.5.

total number of plans	54
total number of matches	20
matches in %	37,037%
total number of mismatches	34
mismatches in %	62,963%

Table 4.4: LASSIE’s plans.

total number of plans	223
total number of matches	45
matches in %	20,179%
total number of mismatches	178
mismatches in %	79,821%

Table 4.5: GESHER’s plans.

The GDM guideline consists of 54 LASSIE and 223 GESHER plans including the orphaned plans. Eventually, 20 of LASSIE’s plans could be matched to 45 of GESHER’s plans.
These numbers led to the following evaluation results shown in Table 4.6:

recall	0,101
precision	0,37
F-Measure	0,159

Table 4.6: Evaluation of the plans.

Recall is a measure of the percentage of information available that is actually found. It provides information about how good the system finds what is relevant. In this case about 10% of correct information was found by the LASSIE system [12].
Precision is a measure of the percentage correctness of the information produced. It tells how good the system sorts out what is irrelevant. So about 37% of the extracted information by LASSIE was actually correct information [12].

At this point it is important to mention that 57,4% of GESHER’s plans (128 plans) are never used in the model (orphaned plans). If we subtract these plans from to total number of GESHER plans, LASSIE actually finds 28,57% of the GESHER plans (see Table 4.7 and 4.8).

total number of plans	95
total number of matches	45
matches in %	47,368%
total number of mismatches	50
mismatches in %	52,632%

Table 4.7: GESHER's plans without orphaned plans.

recall	0,2857
precision	0,37
F-Measure	0,323

Table 4.8: Evaluation of the plans without orphaned plans.

These values do not seem to be that great. An attempt to explain these results are the challenges presented in Section 3.1. Furthermore, it has to be considered that the GESHER CIG contains additional concepts (see Hatsek et al. [20]: only about 50% of the information comes from the original document) and therefore LASSIE can never find these concepts as they are not present in the original guideline text. It is also often the case that only parts of a guideline are used to generate the CIG model. It is common that entire chapters are left out by the modeler because the information is considered irrelevant. But in contrast LASSIE looks through the whole guideline document and extracts information without leaving out any part of the guideline text.

4.3.3 Evaluation of the process structure

The input for the structure comparison are the matching plans resulting from the activity comparison.
I identified and searched for the Exclusive Choice pattern manually in the models:

- GESHER uses the procedural approach.
  19 GESHER plans of the GDM guideline with this pattern are found.
  See Listing A.1 for an example of a GESHER plan.

- LASSIE uses the declarative approach.
  21 LASSIE plans of the GDM guideline with this pattern are found.
  See Listing B.2 for an example of a LASSIE plan.

The same amount of patterns and plans respectively were found by our automatic identification method using XSLT templates.
Among these plans the matches resulting from the activity comparison were examined. Table 4.9 shows all LASSIE plans with `filter-precondition` (first column) and all GESHER plans with `if-then-else` (third column). The corresponding GESHER matches to the LASSIE plans with `filter-precondition` are shown in the second column. The corresponding LASSIE matches to the GESHER plans with `if-then-else` are shown in the fourth column. There are three matches where both the LASSIE plan

has `filter-precondition` and the GESHER plan has `if-then-else` (highlighted gray). In the other cases there is either no match to the plan or the matched plan does not have the pattern. Some plans also have more than one matching plan.

LASSIE	GESHER	GESHER	LASSIE
`filter-precondition`	Match	`if-then-else`	Match
90681	19908	19908	17518
"	19916	"	90681
90685	no Match	"	90854
90686	19860	19913	17518
90689	19972	19919	no Match
"	19973	19922	90697
"	19974	19958	no Match
90695	no Match	19969	no Match
90696	no Match	19980	no Match
90697	19922	19984	no Match
90699	no Match	20002	no Match
90700	no Match	20011	no Match
90701	no Match	20035	no Match
90705	no Match	20041	no Match
90712	no Match	20050	no Match
90846	no Match	20060	no Match
90849	20091	20134	no Match
"	20093	20137	no Match
"	20094	20191	no Match
"	20115	20193	no Match
"	20117	20210	no Match
"	20118		
90850	no Match		
90852	19975		
90853	no Match		
90854	19908		
90855	no Match		
90859	no Match		
90860	no Match		
21		19	

Table 4.9: LASSIE `filter-precondition` vs. GESHER `if-then-else`.

There are only a few matching plans containing the selected pattern. I compared the activities of the plans manually because the language of the conditions is too different. It is difficult to identify the semantics of the conditions and to compare them as they often use negations. But with the transformation introduced in Section 3.3.2 the plans have the same structure and are easier to compare.

Comparing the process structure of the LASSIE and GESHER plans must be seen critically. LASSIE looks for key words in the original guideline text such as ‘if...’:

$$\text{‘}\underbrace{\textbf{\textit{If}}\textit{ the patient was NOT COMPLIANT with the prescribed diet},}_{\textit{condition}}$$

$$\underbrace{\textit{the nurse insists the patient on the importance of eating enough carbohydrates}}_{\textit{activity}}\text{.’}$$

The first part of this sentence is identified as a condition because of the trigger word ‘if’. As a consequence the second part after the comma must be the corresponding activity. Whereas GESHER’s manually generated plans reflect the interpretation of the guideline text by the modeler.

The LASSIE plan #90697 with the `filter-precondition` structure (Listing B.2 in Appendix B) and the GESHER plan #19922 with the `if-then-else` structure (Listing A.1 in Appendix A) are a match (see Table 4.10).
These two plans match because their activity labels are very similar, especially the `filter-precondition` and the `if`.

LASSIE plan #90697	**GESHER plan #19922**
`filter-precondition` *If the patient was not compliant for three or more different meals in a period of one week*	`if` *Patient not compliance for 3 or more different (OR SAME) meals in PAST week*
`plan-body` *insulin therapy should be started*	`then` *Reference: consider insulin notification*
	`else` *change nutritional prescription for patient*

Table 4.10: Structure of two matching plans.

However, there are also matches where only the LASSIE plan has the `filter-precondition` structure or only the GESHER plan has the `if-then-else` structure. The LASSIE plan #90689 with the `filter-precondition` structure (Listing B.1 in Appendix B) and the GESHER plan #19973 (Listing A.3 in Appendix A) are a match. These plans do not share the same structure. However, the superior GESHER plan #19969 (Listing A.2 in Appendix A) which calls the plan #19973 does in fact have the `if-then-else` structure (see Tables 4.11 and 4.12).

LASSIE plan #90689
`filter-precondition` *If the patient was NOT COMPLIANT with the prescribed diet*
`plan-body` *the nurse insists the patient on the importance of eating enough carbohydrates*

Table 4.11: Structure of LASSIE plan #90689.

GESHER plan #19969	GESHER plan #19972	GESHER plan #19973
`if` *Eating enough carbohydorates=yes*		
`then (plan-pointer ID=19972)`	`plan-body` *Call back patient eating enought carbohydrates=Yes*	
`else (plan-pointer ID=19973)`		`plan-body` *Insist on the importance of eating the recommended amount of carbohydrates*

Table 4.12: Structure of GESHER plan #19969 and subplans #19972 and #19973.

Figure 4.7 visualizes the procedural knowledge of the GESHER plan #19969 and its subplans.

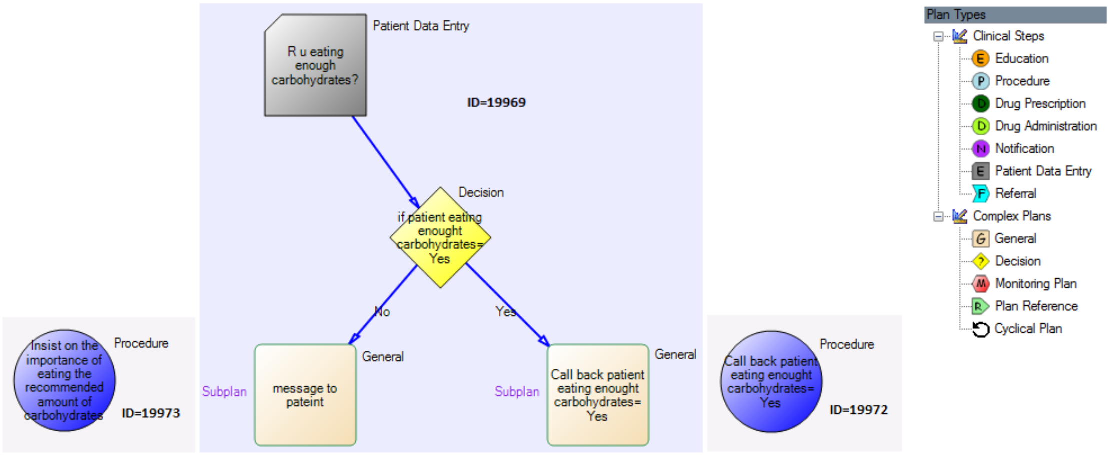


Figure 4.7: GESHER plan #19969 (with key on the right).

The two structurally equivalent plans #90689 (LASSIE) and #19969 (GESHER) are not a match because their activity labels are too different. The labels (especially the `filter-precondition` and the `if`) do not so much differ in their meaning, but they differ a lot in the words that are used to describe the condition. The subplan #19973 of #19969 (GESHER) matches the plan #90689 (LASSIE), but has only a simple structure. If we transform the plan #19973 (GESHER) according to 3.3.2, then we get an activity and a structure match between the LASSIE plan #90689 and the adapted GESHER plan #19973 (see Table 4.13).

LASSIE plan #90689	**adapted GESHER plan #19973**
`filter-precondition` *If the patient was NOT COMPLIANT with the prescribed diet*	`condition` *Eating enough carbohydorates=**NO***
`plan-body` *the nurse insists the patient on the importance of eating enough carbohydrates*	`plan-body` *Insist on the importance of eating the recommended amount of carbohydrates*

Table 4.13: Adapted GESHER plan matches LASSIE plan.

CHAPTER 5

Conclusion

In order to compare two models of a CPG I developed methods for the comparison of the activities and the process structure of the models and implemented them using Java and XSLT templates. I evaluated these methods by testing them with a guideline for GDM.

At first the approach to solve the presented problem of evaluating automatically and manually created treatment plan models seemed pretty straightforward and manageable. However, while implementing the presented methods prototypically it turned out that there might not exist 'the' one overall solution to this problem.
In summary, the automatic model generated with LASSIE finds 43% of the original manually created 'activity' markups and 53% of the original manually created 'condition' markups. 37% of LASSIE's plans match the manually created plans using the GESHER tool. All in all LASSIE finds 10% of the GESHER plans. If we ignore the GESHER plans that are never used, LASSIE actually finds 29% of the GESHER plans.
Looking at the process structure of the two different models shows that there are hardly any matching plans that also have a corresponding structure.

A big challenge when comparing the two models appeared to be the two different ways in which the automatical and manual model were created. This circumstance made the comparison of the two models harder and could also be seen as one of the main reasons why not more matches were found. Another reason for not obtaining better results is the general lack of common rules for creating a CPG model such as [31].

All in all, it would be desirable to have standards and strict rules in the area of CPG modeling. This would not only make the creation of a model and the comparison of two different models easier, but also improve the readability and usability of the models in the daily working process of medical staff.

Bibliography

[1] GATE general architecture for text engineering. ©The University of Sheffield, 1995-2015. `https://gate.ac.uk/`. Accessed: November 2, 2015.

[2] Gestational diabetes. From Wikipedia, the free encyclopedia, `https://en.wikipedia.org/wiki/Gestational_diabetes`. Accessed: August 18, 2015.

[3] Deliverable 3.2: Knowledge Acquisition Tools Prototype Design Documents. MobiGuide Guiding Patients Anytime Everywhere, 2013.

[4] Deliverable 4.1: CIGs KB. MobiGuide Guiding Patients Anytime Everywhere, 2014.

[5] Wolfgang Aigner, Katharina Kaiser, and Silvia Miksch. Visualization Methods to Support Guideline-Based Care Management. In ten Teije et al. [37], chapter 8, pages 140–159. ISBN 978-1-58603-873-1.

[6] Abdulla Ali. Textual Similarity. Bachelor Thesis, Technical University of Denmark, Informatics and Mathematical Modelling, 2011.

[7] Riccardo Bellazzi, Ameen Abu-Hanna, and Jim Hunter, editors. *Artificial Intelligence in Medicine, 11th Conference on Artificial Intelligence in Medicine, AIME 2007, Amsterdam, The Netherlands, July 7-11, 2007, Proceedings*, volume 4594 of *Lecture Notes in Computer Science*. Springer, 2007. ISBN 978-3-540-73598-4.

[8] Ugur Cayoglu, Remco Dijkman, Marlon Dumas, Peter Fettke, Luciano García-Bañuelos, Philip Hake, Christopher Klinkmüller, Henrik Leopold, André Ludwig, Peter Loos, Jan Mendling, Andreas Oberweis, Andreas Schoknecht, Eitam Sheetrit, Tom Thaler, Meike Ullrich, Ingo Weber, and Matthias Weidlich. The Process Model Matching Contest 2013. 4th Int. Workshop on Process Model Collections (Management and Reuse), `http://processcollections.org/past/2013-2/matching-contest`, 2013. Accessed: June 22, 2015.

[9] Hao Chen. String Metrics and Word Similarity applied to Information Retrieval. Master Thesis, University of Eastern Finland, School of Computing, 2012.

[10] Peter Christen. *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection.* Springer, 2012. ISBN 978-3-642-31163-5.

[11] Paul De Clercq, Katharina Kaiser, and Arie Hasman. Computer-interpretable Guideline Formalisms. In ten Teije et al. [37], chapter 2, pages 22–43. ISBN 978-1-58603-873-1.

[12] Robert Dale, Hermann Moisl, and Harold Somers, editors. *Handbook of Natural Language Processing.* CRC Press, 2000. ISBN 978-0824790004.

[13] Kerstin Denecke. Using Semantic Links for Information Extraction and Semantic Representation.

[14] Kerstin Denecke and Jochen Bernauer. Extracting Specific Medical Data Using Semantic Structures. In Bellazzi et al. [7], pages 257–264. ISBN 978-3-540-73598-4.

[15] Kerstin Denecke, Inga Kohlhof, and Jochen Bernauer. Use Of Multiaxial Indexing for Information Extraction From Medical Texts.

[16] Remco Dijkman, Marlon Dumas, and Luciano García-Bañuelos. Graph Matching Algorithms for Business Process Model Similarity Search. In *Proceedings of the 7th International Conference on Business Process Management. BPM '09*, pages 48–63. Berlin, Heidelberg, Springer-Verlag, 2009.

[17] Remco Dijkman, Marlon Dumas, Boudewijn van Dongen, Reina Käärik, and Jan Mendling. Similarity of business process models: Metrics and evaluation. *Information Systems (IS)*, 36(2):498–516, 2011. Elsevier B.V.

[18] Marc Ehrig, Agnes Koschmider, and Andreas Oberweis. Measuring Similarity between Semantic Business Process Models. In *Proceedings of the 4th Asia-Pacific conference on Conceptual modelling*, volume 67, pages 71–80. Australian Computer Science Communications, 2007.

[19] John Fox and Robert Dunlop. OpenClinical knowledge management for medical care: PRO*forma*. ©Copyright OpenClinical 2002-2011, `http://www.openclinical.org/gmm_proforma.html`. Accessed: October 28, 2015.

[20] Avner Hatsek, Yuval Shahar, Meirav Taieb-Maimon, Erez Shalom, Denis Klimov, and Eitan Lunenfeld. A Scalable Architecture for Incremental Specification and Maintenance of Procedural and Declarative Clinical Decision-Support Knowledge. *The Open Medical Informatics Journal*, 4:255–277, 2010.

[21] Arjen Hommersom, Perry Groot, Michael Balser, and Peter Lucas. Formal Methods for Verification of Clinical Practice Guidelines. In ten Teije et al. [37], chapter 4, pages 63–80. ISBN 978-1-58603-873-1.

[22] David Isern and Antonio Moreno. Computer-based execution of clinical guidelines: A review. *International Journal of Medical Informatics*, 77(12):787–808, 2008.

[23] Katharina Kaiser. *LASSIE : Modeling Treatment Processes Using Information Extraction.* Doktorarbeit, Institute of Software Technology and Interactive Systems [E188], Vienna University of Technology, Faculty of Informatics, October 28, 2005.

[24] Katharina Kaiser, Cem Akkaya, and Silvia Miksch. How can information extraction ease formalizing treatment processes in clinical practice guidelines? A method and its evaluation. *Artificial Intelligence in Medicine*, 39(2):151–163, February 2007.

[25] Katharina Kaiser and Mar Marcos. Leveraging workflow control patterns in the domain of clinical practice guidelines. *BMC Medical Informatics and Decision Making*, 16(20), 2016.

[26] Christopher Klinkmüller, Ingo Weber, Jan Mendling, Henrik Leopold, and André Ludwig. Increasing Recall of Process Model Matching by Improved Activity Label Matching. In *Business Process Management, Lecture Notes in Computer Science*, volume 8094, pages 211–218, 2013.

[27] Christoph Kurek. Ein Vergleich der Methoden automatisierter Dublettenerkennung. Diplomarbeit, Universität Hamburg, Fachbereich Informatik, Arbeitsbereich Softwaretechnik, 2011.

[28] W. Lehnert, C. Cardie, D. Fisher, J. McCarthy, E. Riloff, and S. Soderland. Evaluating an Information Extraction System. *Journal of IntegratedComputer-Aided Engineering*, 1(6), 1994.

[29] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *An Introduction to Information Retrieval.* Online edition, Cambridge University Press, Cambridge, England, `http://www-nlp.stanford.edu/IR-book/`, 2009. Accessed: October 14, 2015.

[30] Martin Meitz, Henrik Leopold, and Jan Mendling. An Approach to Support Process Model Validation based on Text Generation. *EMISA Forum*, 33(2):7–21, August 2013.

[31] J. Mendling, H.A. Reijers, and W.M.P. van der Aalst. Seven Process Modeling Guidelines (7PMG). *Information and Software Technology*, 52(2):127–136, 2010.

[32] Jan Mendling. Empirical Studies in Process Model Verification. *LNCS Transactions on Petri Nets and Other Models of Concurrency (ToPNoC). Special Issue on Concurrency in Process-Aware Information Systems*, 2:208–224, 2009. Springer-Verlag.

[33] Nataliya Mulyar, Wil M. P. van der Aalst, and Mor Peleg. Research Paper: A Pattern-based Analysis of Clinical Computer-interpretable Guideline Modeling Languages. *Journal of the American Medical Informatics Association*, 14(6):781–787, 2007.

[34] Mor Peleg, Samson Tu, Jonathan Bury, Paolo Ciccarese, John Fox, Robert A. Greenes, Richard Hall, Peter D. Johnson, Neill Jones, Anand Kumar, Silvia Miksch,

Silvana Quaglini, Andreas Seyfang, Edward H. Shortliffe, and Mario Stefanelli. Comparing Computer-interpretable Guideline Models: A Case-study Approach. *Journal of the American Medical Informatics Association (JAMIA)*, 10(1):52–68, Jan/Feb 2003.

[35] Mercedes Rigla, Raquel Tirado, Assumpta Caixàs, Belén Pons, and Jordi Costa. Gestational Diabetes Guideline CSPT. MobiGuide Guiding Patients Anytime Everywhere, 2012.

[36] Nick Russell, Arthur H.M. ter Hofstede, Wil M.P. van der Aalst, and Nataliya Mulyar. WORKFLOW CONTROL-FLOW PATTERNS: A Revised View. *BPM Center Report BPM-06-22*, BPMcenter.org, 2006.

[37] Annette ten Teije, Silvia Miksch, and Peter Lucas, editors. *Computer-based Medical Guidelines and Protocols: A Primer and Current Trends*, volume 139 of *Studies in Health Technology and Informatics*. IOS Press, 2008. ISBN 978-1-58603-873-1.

[38] Richard Thomson. OpenClinical knowledge management for medical care. ©Copyright OpenClinical 2002-2011, OpenClinical project launched April 2001, `http://www.openclinical.org/home.html`. Accessed: January 23, 2015.

[39] Camilo Thorne, Marco Montali, and Diego Calvanese. VeriCliG Automated Extraction and Verification of Careflows from Clinical Practice Guidelines. The VeriCliG project, ©Camilo Thorne, `http://www.inf.unibz.it/~cathorne/vericlig/`, 2012-13. Accessed: January 21, 2015.

[40] Camilo Thorne, Marco Montali, and Diego Calvanese. The VeriCliG Project (2012 Report). Technical report, KRDB Centre for Knowledge and Data, 2013.

[41] Camilo Thorne, Marco Montali, and Diego Calvanese. The VeriCliG Project (2013 Report). Technical report, KRDB Centre for Knowledge and Data, 2014.

[42] Camilo Thorne, Marco Montali, Diego Calvanese, Claudio Eccher, and Elena Cardillo. The VeriCliG Project: Extraction of Computer Interpretable Guidelines via Syntactic and Semantic Annotation. In *Proceedings of the 2013 Workshop on Computational Semantics for Clinical Text (CSCT 2013)*, 2013.

[43] Thomas Tschach. Kontrollflusstransformation von BPMN zu Asbru. Diplomarbeit, Fakultät für Informatik der Technischen Universität Wien, 2013.

[44] Princeton University. About WordNet. WordNet, Princeton University, `http://wordnet.princeton.edu/`, 2010. Accessed: April 28, 2015.

[45] W.M.P. van der Aalst. The Application of Petri Nets to Workflow Management. *The Journal of Circuits, Systems and Computers*, 8(1):21–66, 1998.

[46] W.M.P. van der Aalst, A.K. Alves de Medeiros, and A.J.M.M. Weijters. Process Equivalence: Comparing Two Process Models Based on Observed Behavior. In S. Dustdar, J.L. Faideiro, and A. Sheth, editors, *International Conference on Business Process Management (BPM 2006)*, volume 4102 of *Lecture Notes in Computer Science*, pages 129–144. Springer-Verlag, Berlin, 2006.

[47] W.M.P. van der Aalst and A.H.M. ter Hofstede. Workflow Patterns home page. ©Workflow Patterns Initiative, `http://www.workflowpatterns.com/`, 2010-2011. Accessed: January 6, 2015.

[48] W.M.P. van der Aalst, A.H.M. ter Hofstede, B. Kiepuszewski, and A.P. Barros. Workflow Patterns. *Distributed and Parallel Databases*, 14(3):5–51, July 2003.

[49] Peter Votruba. Document Exploration and Linking Tool (with Addons). `http://www.ifs.tuwien.ac.at/~votruba/DELTA/`, 2006. Accessed: January 21, 2015.

[50] Ingo Weber, Jörg Hoffmann, and Jan Mendling. Beyond soundness: on the verification of semantic business process models. *Distributed and Parallel Databases (DPD)*, 27(3):271–343, 2010. Springer-Verlag.

[51] Matthias Weidlich, Remco Dijkman, and Jan Mendling. The ICoP Framework: Identification of Correspondences between Process Models. In *Proceedings of the 22nd International Conference on Advanced Information Systems Engineering (CAiSE'10)*, volume 6051 of *Lecture Notes in Computer Science (LNCS)*, pages 483–498. Springer-Verlag, 2010.

[52] Matthias Weidlich, Jan Mendling, and Mathias Weske. Efficient Consistency Measurement based on Behavioural Profiles of Process Models. *IEEE Transactions on Software Engineering (TSE)*, 37(3):410–429, 2011.

[53] Matthias Weidlich, Eitam Sheetrit, Moisés C. Branco, and Avigdor Gal. Matching Business Process Models Using Positional Passage-based Language Models. In *Proceedings of the 32nd International Conference on Conceptual Modeling (ER'13)*, volume 8217 of *Lecture Notes in Computer Science (LNCS)*, pages 130–137. Springer, 2013.

[54] Zhibiao Wu and Martha Palmer. Verb Semantics and Lexical Selection. In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics. Las Cruces, New Mexico*, 1994.

APPENDIX A

GESHER

Listing A.1: GESHER plan #19922.

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<hybrid-asbru-plan id="19922" name="Patient not compliant for 3 or more
    different (OR SAME) meals in PAST week" xsi:noNamespaceSchemaLocation="
    C:\Users\erezsh\Documents\My Dropbox\AyeletTraining\Asbru\schema\
    Current\hybrid-asbru-plan_revised.V4.xsd" xmlns:xsi="http://www.w3.org
    /2001/XMLSchema-instance">
    <plan-body>
      -<semi-formal-plan-body>
         -<single-action isMandatory="true" is-projected="false">
            -<if-then-else comments="">
               -<simple-expression name="Patient not compliance for 3 or
                     more different (OR SAME) meals in PAST week">
                  <concept vocabulary="" name="Patient not compliance for 3
                        or more different (OR SAME) meals in PAST week" key=
                        "5238" path="" concept-type="pattern"/>
                  <text/>
               </simple-expression>
               -<then-branch>
                  -<plan-activation comments="">
                     -<plan-schema>
                        <plan-pointer id="19925" name="Reference: consider
                            insulin notification" DeGeLSite=""/>
                     </plan-schema>
                  </plan-activation>
               </then-branch>
               -<else-branch>
                  -<plan-activation comments="">
                     -<plan-schema>
                        <plan-pointer id="20178" name="change nutritional
                            prescription for patient" DeGeLSite=""/>
                     </plan-schema>
                  </plan-activation>
```

```
                </else-branch>
            </if-then-else>
        </single-action>
    </semi-formal-plan-body>
</plan-body>
</hybrid-asbru-plan>
```

Listing A.2: GESHER plan #19969.

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<hybrid-asbru-plan id="19969" name="if patient eating enought carbohydrates=Yes" xsi:noNamespaceSchemaLocation="C:\Users\erezsh\Documents\My Dropbox\AyeletTraining\Asbru\schema\Current\hybrid-asbru-plan_revised.V4.xsd" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
    <plan-body>
    -<semi-formal-plan-body>
        -<single-action isMandatory="true" is-projected="false">
            -<if-then-else comments="">
                -<simple-expression name="Eating enough carbohydorates=yes">
                    <concept vocabulary="" name="Eating enough carbohydorates=yes" key="5042" path="" concept-type="pattern"/>
                    <text/>
                </simple-expression>
                -<then-branch>
                    -<plan-activation comments="">
                        -<plan-schema>
                            <plan-pointer id="19972" name="Call back patient eating enought carbohydrates=Yes" DeGeLSite=""/>
                        </plan-schema>
                    </plan-activation>
                </then-branch>
                -<else-branch>
                    -<plan-activation comments="">
                        -<plan-schema>
                            <plan-pointer id="19973" name="Insist on the importance of eating the recommended amount of carbohydrates" DeGeLSite=""/>
                        </plan-schema>
                    </plan-activation>
                </else-branch>
            </if-then-else>
        </single-action>
    </semi-formal-plan-body>
</plan-body>
</hybrid-asbru-plan>
```

Listing A.3: GESHER plan #19973.

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
```

```
<hybrid-asbru-plan id="19973" name="Insist on the importance of eating the
    recommended amount of carbohydrates" xsi:noNamespaceSchemaLocation="C:\
    Users\erezsh\Documents\My Dropbox\AyeletTraining\Asbru\schema\Current\
    hybrid-asbru-plan_revised.V4.xsd" xmlns:xsi="http://www.w3.org/2001/
    XMLSchema-instance">
  <plan-body>
    -<semi-formal-plan-body>
      -<single-action isMandatory="true" is-projected="false">
        -<plan-activation comments="">
          -<procedure is-personalized="false" text="" name="Insist on
              the importance of eating the recommended amount of
              carbohydrates" level-of-evidence="" strength-of-
              recommendation="" comments="" glID="19973" concept-key="
              5048" concept-name="insist on the importance of
              following diet prescription">
            <optional-term/>
          </procedure>
        </plan-activation>
      </single-action>
    </semi-formal-plan-body>
  </plan-body>
</hybrid-asbru-plan>
```

APPENDIX B

LASSIE

Listing B.1: LASSIE plan #90689.

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<hybrid-asbru-plan id="90689" name="the nurse insists the patient on the
    importance of eating enough carbohydrates"
    xsi:noNamespaceSchemaLocation="C:\Users\erezsh\Documents\My Dropbox\
    AyeletTraining\Asbru\schema\Current\hybrid-asbru-plan_revised.V4.xsd"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
    <settings>
        <semi-formal-settings>
            <actors>nurse</actors>
            <clinical-settings></clinical-settings>
        </semi-formal-settings>
    </settings>
    <conditions>
        <filter-precondition>
            <structured-text>If the patient was NOT COMPLIANT with the
                prescribed diet</structured-text>
            <semi-formal-filter-precondition confirmation-required="false"
                overridable="false">
                <simple-expression name="If the patient was NOT COMPLIANT
                    with the prescribed diet">
                    <concept concept-type="state-parameter" vocabulary=""
                        name="If the patient was NOT COMPLIANT with the
                        prescribed diet" key=""/>
                    <text>If the patient was NOT COMPLIANT with the
                        prescribed diet</text>
                </simple-expression>
            </semi-formal-filter-precondition>
        </filter-precondition>
    </conditions>
    <plan-body>
        <structured-text>the nurse insists the patient on the importance of
             eating enough carbohydrates</structured-text>
        <semi-formal-plan-body>
            <single-action isMandatory="false">
                <plan-activation>
                    <notification name="the nurse insists the patient on the importance of eating enough carbohydrates" glID="" concept-key=""/>
                </plan-activation>
            </single-action>
        </semi-formal-plan-body>
    </plan-body>
</hybrid-asbru-plan>
```

Listing B.2: LASSIE plan #90697.

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<hybrid-asbru-plan id="90697" name="insulin therapy" xsi:noNamespaceSchemaLocation="C:\Users\erezsh\Documents\My Dropbox\AyeletTraining\Asbru\schema\Current\hybrid-asbru-plan_revised.V4.xsd" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
    <settings>
        <semi-formal-settings>
            <actors>physician</actors>
            <clinical-settings></clinical-settings>
        </semi-formal-settings>
    </settings>
    <conditions>
        <filter-precondition>
            <structured-text>If the patient was not compliant for three or more different meals in a period of one week</structured-text>
            <semi-formal-filter-precondition confirmation-required="false" overridable="false">
                <simple-expression name="If the patient was not compliant for three or more different meals in a period of one week">
                    <concept concept-type="state-parameter" vocabulary="" name="If the patient was not compliant for three or more different meals in a period of one week" key=""/>
                    <text>If the patient was not compliant for three or more different meals in a period of one week</text>
                </simple-expression>
            </semi-formal-filter-precondition>
        </filter-precondition>
    </conditions>
    <plan-body>
        <structured-text>insulin therapy should be started</structured-text>
        <semi-formal-plan-body>
            <single-action isMandatory="false">
                <plan-activation>
```

```
                    <drug-prescription action="start" name="insulin therapy
                        should be started" glID="" concept-key="90922"
                       concept-name="insulin therapy"/>
                </plan-activation>
            </single-action>
        </semi-formal-plan-body>
    </plan-body>
</hybrid-asbru-plan>
```

APPENDIX C

GDM guideline

The guideline for gestational diabetes mellitus is based on the consensus versions of three guidelines (see References in [35]) and was adapted for Corporacio Sanitaria Parc Tauli de Sabadell (CSPT) as part of the MobiGuide project [4].

The aim is to support patients outside the clinic. The GDM CIG is composed of five parallel plans that monitor the patient for 1) blood glucose, 2) ketonuria, 3) diet, 4) exercise and 5) blood pressure. The patient can be notified and asked questions. The caregiver can also be notified when an intervention is required [4].

The GDM guideline [35] is presented on the following pages. The original manually created markups ‘activity’ are displayed in yellow and ‘condition’ in blue.

Introduction

Pregnancy is characterized by insulin resistance and hyperinsulinemia, thus it may predispose some women to develop diabetes. The resistance stems from placental secretion of diabetogenic hormones including growth hormone, corticotropin releasing hormone, placental lactogen, and progesterone, as well as increased maternal adipose deposition, decreased exercise, and increased caloric intake. These and other endocrinologic and metabolic changes ensure that the fetus has an ample supply of fuel and nutrients at all times. Gestational diabetes occurs when pancreatic function is not sufficient to overcome the insulin resistance created by changes in diabetogenic hormones during pregnancy.

Several adverse outcomes have been associated with diabetes during pregnancy. Importantly, the risk of these outcomes increases continuously as maternal fasting plasma glucose levels increase. However, there is no clear threshold that defines patients at increased risk. Adverse outcomes include:

- Preeclampsia
- Hydramnios
- Fetal macrosomia
- Fetal organomegaly (hepatomegaly, cardiomegaly)
- Birth trauma
- Operative delivery
- Perinatal mortality
- Neonatal respiratory problems and metabolic complications (hypoglycemia, hyperbilirubinemia, hypocalcemia, erythremia)

If maternal hyperglycemia is present during organogenesis because of overt (also termed pregestational) diabetes, there is an increased risk of miscarriage and congenital anomalies.

Definition and diagnostic criteria

Gestational diabetes is defined as carbohydrate intolerance that begins or is first recognized during pregnancy. In 2010, the International Association of Diabetes and Pregnancy Study Group (IADPSG), recommended to classified it in two subtypes:

a. overt diabetes, which is defined with the same criteria than in general population but during the initial prenatal visit (Note: These patients will not be considered in MobiGuide):
    - Fasting plasma glucose ≥ 126 mg/dL [7.0 mmol/L], or
    - A1C ≥ 6.5 percent using a standardized assay, or
    - Random plasma glucose ≥ 200 mg/dL [11.1 mmol/] that is subsequently confirmed by elevated fasting plasma glucose or A1C, as noted above

b. Gestational diabetes.

    The diagnostic criteria for gestational diabetes have been a matter of debate since the publication of the IADPSG recommendations in 2010. The recommendation is based on outcome data reported in the HAPO study.

    The Hyperglycemia and Adverse Pregnancy Outcome (HAPO) study included more than 23,000 pregnant women. After a 75-g oral two-hour glucose tolerance test, the risk of macrosomia increased as much as five-fold as fasting blood glucose concentration increased above 75 mg/dL (4.2 mmol/L), or one-hour glucose concentration increased above 105 mg/dL (5.8 mmol/L), or two-hour glucose concentration increased above 90 mg/dL (5.0 mmol/L), and the risk increased continuously across the spectrum of glucose results. There was also a positive, but weaker, correlation between increasing glucose concentration and maternal complications (eg, preeclampsia) and neonatal metabolic morbidity (eg, hypoglycemia, hyperbilirubinemia), but not long-term childhood morbidity, such as obesity at age two years in a small subset of offspring. Of note, women with significant hyperglycemia were excluded from the HAPO analysis (exclusion criteria: fasting glucose concentration greater than 105 mg/dL [5.8 mmol/L], two-hour glucose concentration greater than 200 mg/dL [11.1 mmol/L], or a random glucose concentration later in gestation greater than 160 mg/dL [8.9 mmol/L]).

On the light of the results of the HAPO study, the International Association of Diabetes and Pregnancy Study Groups (IADPSG) and the American Diabetes Association proposed criteria for Gestational Diabetes based on the results of the 75-gram Glucose tolerance test

Plasma glucose	**mg/dl**	**mmol/l**
Fasting	≥ 92	≥ 5.1
One hour	≥ 180	≥ 10.0
Two hours	≥ 153	≥ 8.5

The thresholds represent the glucose values at which the odds of infant birth weight, cord C-peptide (proxy for fetal insulin level), and percent body fat >90 percentile were 1.75 times the estimated odds of these outcomes at mean glucose levels, based on fully adjusted logistic regression models. Compared to women in the HAPO study with all glucose values below the thresholds, women who exceeded one or more of these thresholds had a two-fold higher frequency of large for gestational age infants and preeclampsia, and >45 percent increase in preterm delivery and primary cesarean delivery. Using an odds ratio of 2 for the thresholds defined a population with further increased frequencies of these outcomes, but the increase was modest and resulted in failure to identify many women who were at almost comparable risk.

However, although the ADA has assumed the IADPSG recommendations, the American College of Obstetrics and gynecologists (ACOG) does not recommend this approach "because there is no evidence that diagnosis using these criteria leads to clinically significant improvements in maternal or newborn outcomes and it would lead to a significant increase in health care costs".

The National Institute of Health is planning a Consensus Development Conference to determine the optimal approach to screening and diagnosis in the United States.

Thus, in this guideline we are going to maintain the recommendations for diagnosis made by the Spanish Group for the study of Diabetes and Pregnancy which agree with the recommendations from the ACOG:

1. All pregnant women should be screened for GDM using the O'Sullivan test, that consist on a 50-g, 1-hour loading test.
    a. First trimester (high risk for GDM)
        i. Age ≥ 35
        ii. Body Mass Index (BMI)= mass (Kg)/ (Height (m))^2 ≥ 30 kg/m2
        iii. Previous GDM confirmed or suspected on the light of the perinatal outcomes, i.e. macrosomia.
        iv. Familiar history of diabetes in first degree relatives.

b. Second trimester (24-28 weeks)
    i. All the pregnant women non previously diagnosed of GDM

c. Third trimester
    i. All the pregnant women non previously screened
    ii. Women with normal O'Sullivan test but who develop complications suggesting GDM (macrosomia, polihydramnios). In those cases it would be performed directly a 100 gr OGTT test

2. The diagnosis of GDM can be made based on the result of the 100-g, 3-hour oral glucose tolerance test, for which there is evidence that treatment improves outcome. The plasma levels designated by the National Diabetes Data Group are appropriate to use (see table).

Status	**Plasma level**	
	mg/dL	mmol/L
Fasting	105	5.8
1 hour	190	10.6
2 hour	165	9.2
3 hour	145	8.0

3. A positive diagnosis requires that two or more thresholds be met or exceeded. If only one point is exceeded, then fasting and 1h postprandial glucose control are adviced. In case of normality of all the glucose results during at least the first week of follow-up, random and sporadic tests would be enough to guarantee an appropriate carbohydrate tolerance. The patient receives nutritional recommendations and exercise recommendations. Note: These patients will be enrolled in MobiGuide.

The highly concentrated hyperosmolar glucose solution can cause gastric irritation, delayed emptying, and gastrointestinal osmotic imbalance, leading to nausea and vomiting. In case of patients unable to tolerate oral hyperosmolar glucose, the recommended option is:

Serial glucose monitoring -- Periodic random fasting and one-hour postprandial blood glucose testing -- is a monitoring option for women at high risk for gestational diabetes who are unable to take an oral glucose load. HbA1c ≤ 6% could help to guarantee that this option is a reasonable way to start with. The patient receives nutritional recommendations and exercise recommendations. Note: These patients will be enrolled in MobiGuide.

Rationale for treatment

Identifying women with GDM is important because meta-analysis of randomized trials has shown that appropriate therapy can decrease maternal and fetal morbidity, particularly macrosomia. The value of treatment of GDM was best illustrated in two large trials that randomly assigned women with mild GDM to a regimen of either diet/blood glucose monitoring/insulin as needed or routine obstetrical care. One trial was performed in Australia and the other in the United States. Both trials observed that treatment resulted in a significant decrease in the prevalence of macrosomia (US trial: 6 versus 14 percent with usual care; Australian trial: 10 versus 21 percent with usual care), and the reduction in macrosomia was associated with a significant decrease in shoulder dystocia in one trial and the composite outcome of shoulder dystocia, nerve palsy, bone fracture, and death in the other. The rate of cesarean delivery was significantly reduced in the treatment group of one trial, and not increased in the other trial, although an increase is often observed in women with gestational diabetes. Treatment did not lead to a significant reduction in the prevalence of neonatal metabolic abnormalities (eg, hypoglycemia, hyperbilirubinemia) in either trial. Fetal/neonatal death was rare in all groups in both trials.

Both trials reported lower pregnancy weight gain in the treated group than in the control group. The US trial also reported a significant decrease in the rates of preeclampsia and gestational hypertension in the treatment group (8.6 versus 13.6 percent with usual care) and the Australian trial noted a significantly reduced rate of gestational hypertension (12 versus 18 percent with usual care).

Monitoring

Glucose monitoring

Women with GDM should measure their blood glucose concentration at least four times daily (fasting and one hour after the first bite of each meal) to determine whether hyperglycemia severe enough to increase fetal risk is occurring. Results should be recorded in a glucose log, along with dietary information. This facilitates recognition of glycemic patterns and helps immeasurably in interpreting results stored in the memory of modern meters. Multiple daily measurements allow recognition of women who should begin insulin therapy and appear to decrease the risk of macrosomia. The decision to start insulin therapy (or not to start it) is taken by the endocrinologist. Note: in some not very frequent cases, the insulin therapy is not started after considering the specific patient personal context (e.g. gestational age = 39 week and fetus percentile < 60, low family support, analphabetism and post-prandial BG < 160 mg/dl and fasting BG < 120 mg/dl).

Although there are no data on the duration of good control sufficient to reduce the frequency of self-monitoring or the appropriate frequency of testing in GDM that is well controlled (see section 5.1.1 Glucose target) with nutritional therapy, we recommend decreasing the frequency of glucose monitoring when good glycemic control is accomplished with medical nutritional therapy for more than 1 month. Decrease the frequency of glucose monitoring means to measure blood glucose (fasting and 3 postprandial measurements) twice a week instead of every day.

The advent of continuous glucose monitoring has the potential to allow determination of peak postprandial glucose levels; future research should determine whether using CGM in management of GDM will improve outcomes.

Glucose target

The ACOG recommends the following targets, with insulin therapy initiated if they are exceeded.

- Fasting blood glucose concentration ≤ 95 mg/dL (5.3 mmol/L)
- One hour-postprandial glucose < 130 to 140 mg/dL (7.2 to 7.8 mmol/L) or
- Two hour-postprandial blood concentration ≤ 120 mg/dL (6.7 mmol/L)

In 'Hospital de Sabadell', we recommend the patients to measure postprandial blood glucose one hour after the first bite of each meal. For one-hour postprandial glucose, 140 mg/dL is considered the bound to detect anomalous glycemic control.

Little guidance is available as to what proportion of measurements exceeding these thresholds should trigger intervention. Insulin should be considered if elevated blood glucose is observed (see 5.3.1 Insulin).

Once insulin has been started, the glucose values considered as a goal of treatment will be those recommended by the 5th Workshop-Conference on Gestational Diabetes Mellitus:

- Fasting blood glucose concentration < 95 mg/dL (5.3 mmol/L)
- Sixty-90 minutes-postprandial glucose ≤ 120 mg/dL (6.7 mmol/L)

Glycated hemoglobin

Glycated hemoglobin (A1C) may be a helpful test in assessing glycemic control during pregnancy, particullarly when overt diabetes is suspected. It should be noted that A1C values tend to be lower in pregnant compared to nonpregnant women because the average blood glucose concentration is about 20 percent lower in pregnant women and, in the first half of pregnancy, there is a rise in red cell mass and a slight decrease in red blood cell life span.

Ketonuria

Although there is conflicting evidence as to whether ketonuria is associated with an adverse effect on cognitive development of the fetus, we routinely monitor fasting urinary ketones in women with GDM. The patient measures ketonuria using urine strips. Monitor ketonuria routinely means to measure ketones in the urine every day at fasting conditions. If ketonuria is suspected by the patient along the day, she could decide to additionally measure ketonuria before lunch or dinner. The results of ketonuria could be: a) positive (++); b) positive (+); c) negative (+/-); d) negative (-); e) negative (--).

Ketonuria indicates that the person is in a catabolic state and is breaking down fat, and can occur in anyone who has a negative caloric balance. Pregnant women develop elevated Î²-hydroxybutyrate levels more rapidly than nonpregnant individuals during a 12- to 18-hour fast, and it is not known whether such elevated levels have an adverse impact on fetal development. Several studies have suggested that early maternal malnutrition can affect neurobehavioral development in children of women with diabetes. In one such study, plasma beta-hydroxybutyrate levels independent of glucose levels had an adverse association with cognitive development in pregnancies in women with prepregnancy diabetes, GDM, and in normal pregnancies.

If ketonuria is negative for two weeks, the patient is recommended to measure ketonuria 2 or 3 times a week. In case of ketonuria detection (the number of ketonuria measurements with result "positive" is equal or higher than 3 in a period of time of one week):

- If the patient was COMPLIANT with the prescribed diet, the nurse decides to increase the carbohydrates intake either at dinner or at bedtime: the amount of carbohydrates at dinner or at bedtime is increased by 1 unit (10 grams). If post-dinner or fasting glucose values achieve the threshold for insulin treatment as a consequence of increasing the amount of carbohydrates, starting insulin would be the best option.
- If the patient was NOT COMPLIANT with the prescribed diet, the nurse insists the patient on the importance of eating enough carbohydrates. If the situation is kept for more than 1 additional week, insulin therapy should be started.

Therapy

Nutrition

Patients with GDM should receive nutritional counseling upon diagnosis and be placed on an appropriate diet. The goals of medical nutritional therapy are to:

- Achieve normoglycemia (see ACOG recommendation in 4.1.1 Glucose target)
- Prevent ketosis
- Provide adequate weight gain
- Contribute to fetal well-being

There is scanty level evidence to support most aspects of the nutritional prescription for GDM.

In clinical practice, women often require 1600 to 2200 kcal per day.

Weight during pregnancy	Caloric requirement (kcal/kg/day)
Underweight	40
Ideal body weight	30
Overweight	22 to 25
Morbidly obese	12 to 14

Once the caloric needs are calculated, carbohydrate intake needs to be distributed across meals and snacks to blunt postprandial hyperglycemia. Carbohydrate intake is limited to less than 55 percent of total calories and should be distributed in 3 main meals and 2-3 snacks. In order to avoid fasting ketonuria and facilitate post-dinner glucose control, a bed-time snack may be needed. As a general recommendation, the patients are informed that complex carbohydrates, such as those in starches and vegetables, are more nutrient dense and raise postprandial blood glucose concentrations less than simple sugars, which should be avoided.

Close follow-up is important to ensure nutritional adequacy. Individual assessment and self blood glucose monitoring are used to determine and modify specific nutrition/food recommendations. If the patient acknowledges that she is not following nutritional prescription more than once, she receives specific recommendations about the importance of following nutritional prescription.

If insulin therapy is added to nutrition therapy, a primary goal is to maintain carbohydrate consistency at meals and snacks to facilitate insulin adjustments. This decision is taken by nurse + physician.

Exercise

The value of exercise in women with GDM requires further exploration to determine the potential range of benefits. Nevertheless, based on the data available in pregnant and in nonpregnant individuals, we recommend the regular practice (≥ 4days/week) of light or moderate exercise (16 - 28 METs hours per week) as part of the treatment plan for women with GDM and no medical or obstetrical contraindications to this level of physical activity.

If the patient does not follow recommendations related to physical activity, she is insisted by the nurse on the importance of following the recommended practice of physical activity.

This section of the general guideline will be further developed in more detail.

Pharmacologic therapy

If normoglycemia cannot be maintained by medical nutritional therapy, then anti-hyperglycemic agents should be initiated. There is only one option in pregnant patients who require medical therapy aimed at controlling blood glucose: insulin (and some insulin analogs), which is the only recommended approach in Spain.

Insulin

Women with GDM are placed on insulin therapy when target glucose levels are exceeded despite dietary therapy.

In the two randomized trials in which diagnosis and treatment of mild GDM improved outcomes, only 20 and 8 percent of women, respectively, required insulin, while 80 and 92 percent of women, respectively, were treated satisfactorily with diet.

Insulin therapy is started if two or more elevated values are observed in a week period in the same interval (fasting, postprandial Breakfast, postprandial LUNCH, postprandial DINNER). The endocrinologist is the person who takes the decision to start insulin therapy.

If only two blood glucose targets are exceeded in one specific measurement point (fasting, 1-hour postprandial breakfast, 1-hour postprandial lunch, 1-hour postprandial dinner) and the threshold is exceeded by less than 15 mg/dL, then two situations are considered (only for the first time these situations are detected):

- If the patient was NOT COMPLIANT with the prescribed diet (diet intake higher than the recommended amount of carbohydrates or eating not recommended food (such as ice cream, buns, cakes, packaged juice, plain sugar, or chocolate) or eating snacks not included in the prescribed diet) then the nurse or the doctor motivate the patient to be compliant with the diet. If the patient was not compliant for three or more different meals in a period of one week, and blood glucose is elevated as a result of non- compliance to nutritional prescription, insulin therapy should be started.
- If the patient was COMPLIANT with the prescribed diet then the nurse or the doctor can postpone starting insulin therapy by changing the nutritional prescription at lunch or dinner. E.g. two 1h post-dinner values of 142; 145 in a women eating more than or equal to 50 carbohydrate grams for dinner, with persistant negative fasting ketonuria. In this example the reduction at dinner of 10 grams of carbohydrates could be enough for glucose normality restoration. See table for other situations:

Breakfast?	**Lunch**	**Dinner**
Diet 1600 kcal	60 g CARBS --≥ 50 g CARBS	50 g CARBS --≥ 40 g CARBS
Diet 1800 kcal	85 g CARBS --≥ 75 g CARBS	55 g CARBS --≥ 45 g CARBS
Diet 2000 kcal	85 g CARBS --≥ 75 g CARBS	75 g CARBS --≥ 65 g CARBS

Furthermore, the newborn weight prognosis based on ultrasoud measurements may be taken into account in the decision of starting insulin therapy; For fetus considered large for gestational age, one hour-postprandial glucose values above 120 mg/dL are enough for starting insulin. On the contrary, when the fetus is considered small for gestational age, 140 mg/dl should be the threshold value for considering insulin treatment. Definition of Small Gestational Age (percentile ≤ 10) and Large Gestational Age (percentile > 90).

The dose and type of insulin used is calculated based upon the specific abnormality of blood glucose noted during monitoring. One principle we have found useful is to start with the simplest regimen and increase the complexity as needed to address the particular situation. If the post-dinner glucose level is elevated, then an injection of rapid acting insulin is given just prior to dinner. Additional doses of rapid acting insulin can be necessary to maintain euglycemia after breakfast or lunch. If fasting glucose is elevated, intermediate acting insulin can be given at bedtime, then a total of four injections per day are needed.

In summary:

- If insulin is required because the fasting blood glucose concentration is high, an intermediate-acting insulin (NPH insulin), is given before bedtime; an initial dose of 0.1-0.15 unit/kg body weight is utilized.
- If postprandial blood glucose concentrations are high, insulin aspart or insulin lispro is given before meals at an initial dose of 4-6 units.
- In a severely obese woman, the initial doses of insulin may need to be increased to overcome the combined insulin resistance of pregnancy and obesity. This should be personalized by the endocrinologist according to each patientâ€™s characteristics.

Subsequent (weekly) adjustments in the various components of the insulin regimen are made based upon the corresponding glucose levels. Because any insulin regimen requires serial readjustment of dosage in response to specific fasting or postprandial glucose levels, the starting dose should be considered just that, a starting point. Adjustments in insulin dosage may be done to achieve the above mentioned objectives.

Hypoglycemia remote from meal or snack time is rare in women with GDM, and is treated by administering 10 to 20 g of a simple sugar like orange juice or non-fat milk. The patient is instructed to take this decision after getting a measure of blood glucose. If low glucose values are encountered more than once at the same time of day, insulin doses are adjusted downward accordingly.

Long-acting insulin analogs (insulin glargine, insulin detemir) have not been studied extensively in pregnancy. However, insulin detemir has been recently approved by the EMEA for the treatment of women with diabetes and pregnancy. In vitro perfusion studies have demonstrated that insulin glargine does not cross the placenta, however, concern about transplacental transfer of glargine in vivo remains. Based on available data, we prefer use of human NPH insulin as part of a multiple injection regimen in pregnant women with GDM.

Oral anti-hyperglycemic agents

A systematic review by the Johns Hopkins University Evidence-based Practice Center for the Agency for Healthcare Research and Quality presented evidence from randomized trials and observational studies that:

1. maternal glucose levels did not differ substantially between gravidae treated with insulin versus those treated with oral glucose-lowering agents, and
2. there was no consistent evidence of an increase in any adverse maternal or neonatal outcome with use of glyburide, acarbose, or metformin compared with use of insulin.

The ADA and ACOG do not endorse the use of oral anti-hyperglycemic agents during pregnancy and such therapy has not been approved by the Unites States Food and Drug Administration for treatment of GDM. We concur, but metformin may be a good complement to diet (plus insulin if required) in obese women. This should be personalized according to each patient's characteristics. However, metformin crosses the placenta and, in one study, cord arterial levels were twice as high as maternal venous levels.

Obstetric management

Obstetric management will be similar to the usually applied for non GDM women. However, it is advised to add an additional US study during 28-32 gestational week to detect macrosomia.

Peripartum management

Maternal hyperglycemia should be avoided during labor to reduce the risk of fetal acidosis and neonatal hypoglycemia. The risk of adverse neonatal metabolic outcomes (hypoglycemia, hyperbilirubinemia, hypocalcemia, erythremia) is related to both antepartum and intrapartum maternal hyperglycemia and appears to increase with the degree of maternal hyperglycemia. A protocol for intrapartum metabolic control is being applied which includes e.v. insulin perfusion/ dextrose serum. However, for almost all women with gestational diabetes, no insulin treatment is needed during labour and we are going to review this part of the protocol in order to simplify it.

Future risks

Most women with GDM are normoglycemic after delivery. However, they are at high risk for recurrent GDM, prediabetes (impaired glucose tolerance or impaired fasting glucose), and overt diabetes over the subsequent five years.

Recurrence

One-third to two-thirds of women with GDM will have GDM in a subsequent pregnancy. Women who have a recurrence tend to be older, more parous, and have a greater increase in weight between their pregnancies than women without a recurrence. Higher infant birth weight in the index pregnancy and higher maternal prepregnancy weight have also been associated with recurrent GDM.

Long-term risk of diabetes

A history of GDM is predictive of an increased risk of developing type 2 diabetes, type 1 diabetes, and cardiovascular disease.

- As many as 20 percent of women with GDM have impaired glucose tolerance during the early postpartum period.
- A systematic review and meta-analysis found that women with GDM were at significantly higher risk of developing subsequent type 2 diabetes than women with normoglycemic pregnancies (RR 7.43, 95% CI 4.79-11.51; 20 cohort studies including 675,455 women of who 10,859 had type 2 diabetes). The relative risk was 4.69 within the first five years after delivery and 9.34 more than five years after delivery. In fact, 10 to 31 percent of parous nonpregnant women with diabetes have experienced a pregnancy complicated by GDM prior to their diagnosis.
- Waist circumference and BMI are the strongest anthropometric measures associated with development of type 2 diabetes in women with GDM. Type 2 diabetes develops in 50 to 75 percent of obese (BMI ≥ 30 kg/m2) women with a history of GDM versus fewer than 25 percent of women with GDM who achieve normal body weight after delivery. Other major risk factors are gestational requirement for insulin and early gestational age at the time of diagnosis (ie, less than 24 weeks of gestation). Additional risk factors for impaired glucose tolerance and overt diabetes later in life include autoantibodies (eg, glutamic acid decarboxylase, insulinoma antigen-2), high fasting blood glucose concentrations during pregnancy and early postpartum, higher fasting blood glucose at diagnosis of GDM and high glucose levels in oral glucose tolerance testing, neonatal hypoglycemia, and GDM in more than one pregnancy. Parity, large birth weight, and diabetes in a first-degree relative are less correlated with later diabetes.

- GDM is also a risk factor for the development of type 1 diabetes, particularly in populations with a high prevalence of this disorder. In our country there is no recent data but one study found that more than 13% of women with GDM have positivity to islet cell antibodies (ICA). Thus, if there is suspected the possibility of type 1 diabetes, determination of plasma autoimmune markers are mandatory (GAD65 Ab; IAA).
- Women with GDM are at higher risk of developing cardiovascular disease and developing it at a younger age than women with no history of GDM.

Follow-up and prevention of type 2 diabetes

ACOG, the ADA and the Fifth International Workshop Conference on Gestational Diabetes recommend long-term follow-up of women with GDM:

All women with previous GDM should undergo an oral glucose tolerance test 6 to 12 weeks after delivery, using a two-hour 75 gram oral glucose tolerance test. An abnormal fasting blood glucose level is diagnostic (diabetes if ≥ 126 mg/dL, impaired fasting glucose (IFG) if 100 to 125 mg/dL); however, sensitivity for diagnosis of diabetes is low. Impaired glucose tolerance (IGT) is diagnosed if the two-hour value is 140 to 199 mg/dL. Collectively, IFG and IGT are known as "prediabetes."

- Women with an abnormal oral glucose tolerance test are then classified as having prediabetes or overt diabetes mellitus
- Those with prediabetes should be counseled about their subsequent risk for developing overt diabetes and referred for discussion of management options (eg, lifetime modification such as medical nutritional therapy, indications for metformin). They should try to achieve their ideal body weight through diet and exercise and, if possible, they should avoid drugs that may adversely affect glucose tolerance (eg, glucocorticoids). They should have yearly assessment of glycemic status.
- A woman who has overt diabetes mellitus should receive appropriate education and treatment. She should also be given advice regarding contraception and the planning of future pregnancies.
- Women with prediabetes or overt diabetes should be counseled regarding the importance of good metabolic control prior to any future pregnancies.
- Women with normal glucose tolerance should be counseled regarding their risk of developing GDM in subsequent pregnancies and type 2 diabetes in the future. Lifestyle interventions (weight loss, exercise) are clearly beneficial for reducing the incidence of these disorders. Drug therapy (eg, metformin) also may have a role in preventing future type 2 diabetes.

References

Adapted from

- Screening and Diagnosis of diabetes mellitus during pregnancy. Treatment and course of gestational diabetes mellitus. Up to Date. Authors: Donald R Coustan, MD; Lois Jovanovic, MD
- Guia asistencial de diabetes mellitus y embarazo. Grupo Espanol de Diabetes y Embarazo. Av.Diabetol 2006;22:73-87
- Screening and diagnosis of gestational diabetes mellitus. Obstet Gynecol 2011;118:751-2. Committee opinion on Obstetric Practice

Glossary

Asbru A task-specific and intention-based plan representation language to embody clinical guidelines and protocols as time-oriented skeletal plans [34]. 1, 3, 4, 6, 8, 9, 19, 25, 26, 33, 37, 77

clinical practice guideline A set of general rules and policies for the management of patients who have a particular clinical condition. 1

GATE General Architecture for Text Engineering - an open-source free software for text processing [1, 3]. 33

GESHER A knowledge acquisition application for specifying procedural and declarative knowledge of CPGs [3]. xvii, xviii, 4, 26, 31, 32, 34, 35, 40, 42–47, 49, 57, 58

gestational diabetes mellitus A condition in which women without previously diagnosed diabetes exhibit high blood glucose levels during pregnancy [2]. 34, 65

LASSIE A method using IE techniques to semi-automatically model parts of CIGs from textual documents in a CIG formalism such as Asbru [3]. xvii, xviii, 4, 9, 27, 31, 33, 34, 40–47, 49, 61, 62

Acronyms

BPM Business Process Management. 17

BPMN Business Process Model and Notation. xvi, xvii, 10, 11, 25, 26

CGP clinical guideline and protocol. 33

CIG computer-interpretable guideline. 1, 5, 6, 9, 10, 25, 31, 43, 65, 77

CPG clinical practice guideline. 1, 3, 17, 25–27, 31, 33, 49, 77

GDM gestational diabetes mellitus. xvii, 34, 35, 40, 42, 43, 49, 65

IE Information Extraction. 2, 3, 9, 10, 33, 77

NLP Natural Language Processing. 4, 6, 9, 33

UMLS Unified Medical Language System. 10, 11, 33

XML Extensible Markup Language. 8, 33, 37

XSLT Extensible Stylesheet Language Transformation. xviii, 37, 39, 40, 43, 49