

INTRODUCTION

*The mere formulation of a
problem is far more essential
than its solution*

Albert Einstein

1.1 MOTIVATION & RESEARCH QUESTIONS

Clinical guidelines (CGs) are important instruments to reduce variance in health care services by providing the health-care professionals with guidance for best-practices. In order to raise the quality of care service, another movement was to introduce a number of methodologies for guideline authoring where extensive literature review provides scientific evidence for the best-practice recommendations [76]. Although studies have shown that CGs adoption did improve the quality of care by reducing its variance [76], several issues still remain, for example: (i) the scientific evidence is often targeted to idealized single-common-diseases patients, which contributes to the exclusion of minorities such as patients that suffer from rare diseases or from multiple diseases (multimorbidity) [6, 23, 62]; and consequently, (ii) CGs are also not easily adaptable to patients' preferences; (iii) the process of updating CGs are often much slower with respect to the speed in which new evidence is made available [92]; and (iv) CGs are not easily adaptable to local circumstances (national or organizational) [33].

The current challenge in the medical field then is:

How to improve the current guideline-system in order to tackle its limitations and expand its benefits?

A schema for the current CG life-cycle is presented in Fig. 1.1, divided in four main phases (adapted from [11]): Authoring, Dissemination, Implementation and Evaluation. This can be briefly read as: in the Evaluation phase the outcomes of clinical practice are assessed, which produces input for developing or updating a CG in the Authoring phase, together with the results from Clinical Research; the produced guidelines are made public in the Dissemination phase to

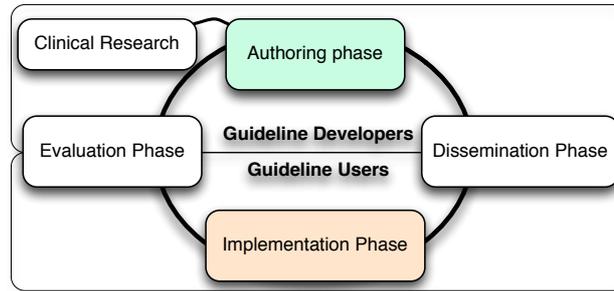


Figure 1.1: Clinical Guideline Life-Cycle

be finally applied in clinical practice in the Implementation phase, which will produce new outcomes to be assessed.

Within each phase, several tasks take place: for example, design of key-questions, literature search, comorbidity analysis (diseases that are influenced by others), guideline application to a patient. Hereby we refer to them as CG-tasks, or CG-subtasks when it is part of one or more tasks. Mostly those are knowledge-intensive tasks, which means they require a considerable amount of clinical knowledge to be handled by the health-care professionals. The guideline life-cycle is then a laborious process that could considerably benefit from computational support.

Several computer-based approaches have been proposed to support CG-tasks. For example, Peleg et al. [66] present a survey of approaches divided by the CG-tasks they address. However, this emphasizes that the guideline life-cycle is supported by fragmented computer-based approaches, rather than as a whole process. i.e. the existing approaches are designed for specific CG-tasks.

GENERAL PROBLEM: *The CG life-cycle is not supported by computer systems as a whole process, since existing computer-based approaches are designed to address specific CG-tasks.*

In particular, Computer Interpretable Guideline (CIG) formalisms or languages are designed so that the clinical knowledge can be represented and computed to produce desired information. Naturally, the CG-task mostly addressed is the application of a guideline to a particular patient (also called ‘execution of a CIG’) since this is the main purpose of CGs. As a consequence, most of the CIG languages express ‘plans’, which regards a number of actions/tasks to be performed in a certain order, like in a treatment (eventually expressing goals/intentions). However, the guideline is more than that:

it contains a number of recommendations, qualified with a deontic strength, based on justifications for certain actions/tasks to be performed or avoided, which is meant to drive the decisions about a plan. Based on the outcomes of our literature review, we identified two related issues: (i) the main CIG formalisms are primarily designed to support the execution of a CIG; (ii) formalisms proposed to address comorbidity analysis are either restricted by modeling decisions made for the CIG-execution task, or, again, designed for the comorbidity analysis task in particular.

SPECIFIC PROBLEM: *The main CIG formalisms are primarily designed to address CIG-execution task, which restricts its reuse for other purposes.*

A consequence of those problems is to have little or no reuse of computer-interpretable clinical knowledge across tasks and phases. We advocate that the ability to reuse knowledge is an essential feature for a ‘computationally-sustainable’ approach to support CG-tasks. Therefore we formulate the following hypothesis to guide our investigation:

HYPOTHESIS: *A task-independent knowledge representation can improve the ability to support several tasks while reusing knowledge within the guidelines and from external sources.*

This brings a specific challenge for Artificial Intelligence in Medicine:

Which clinical knowledge is needed to support several CG tasks while fostering knowledge reuse across those tasks?

This thesis divides this challenge in three main research questions:

Research Question 1 (RQ1): *What is a suitable conceptual model to address the CG tasks?*

Research Question 2 (RQ2): *What are suitable formal languages to formalize the CG model and tasks?*

Research Question 3 (RQ3): *Is the Semantic Web paradigm suitable to support CG tasks and foster knowledge reuse?*

In particular, this research investigates how to support two sub-tasks within CG-tasks from the authoring phase, namely **development and update**. Hereby we briefly describe the tasks and how the sub-tasks are currently tackled, followed by a corresponding ‘idealized’ sub-task supported by a computer system.

GUIDELINE DEVELOPMENT : a single disease or condition is investigated in order to identify the best-practices recommended to tackle it.

Comorbidity Analysis: in the current process, this task regards investigating secondary diseases that often co-occur with the target one in order to identify possible recommendations that may be problematic and propose alternative ones. This task is traditionally limited to two diseases as the pairwise analysis is already hard enough to be manually performed.

Multimorbidity Analysis: this is an idealized task where a computer-system supports the analysis of several co-occurrent diseases or conditions to anticipate some typical or dangerous problems of common diseases or conditions co-occurring. For example, diabetes, hypertension and osteoarthritis commonly co-occur among elderly people.

GUIDELINE UPDATE : guidelines need to be frequently updated to accommodate the latest scientific evidence.

Literature Search: in the current process, a new literature review is performed by an expert committee by consulting literature repositories such as PubMed after a predefined time has passed from the guideline latest publication. There is not a consensus about how often the literature should be checked neither how many new evidences are needed to justify a new guideline to be published. In general it does not happen as often as the new evidences are released, and it can take a few years.

Automated Paper Retrieval: this is an idealized task where a system would automatically retrieve new papers that are relevant for a certain guideline based on its conclusions, i.e. the summary of the papers used as evidence.

Finally, the research problem investigated in this thesis is summarized in Table 1.1.

1.2 METHODOLOGY

Our goal is to develop a knowledge model that is not biased to a particular CG-task but that is applicable to several CG-tasks while fostering knowledge reuse. In order to achieve it we adopted an iterative methodology as depicted in Fig. 1.2. It comprises three phases that are performed in cycles with the purpose of finding partial answers

Table 1.1: Research Summary

Area	Knowledge Representation & Medical Informatics
Subject	Computer-Interpretable Clinical Guidelines (CIG)
General Problem	The CG life-cycle is not supported by computer systems as a whole process, since existing computer-based approaches are designed to address specific CG-tasks.
Specific Problem	The main CIG formalisms are primarily designed to address the CIG-execution task, which restricts its reuse for other purposes as the knowledge representation underlying them capture the knowledge in a “narrow” way.
Hypothesis	A task-independent knowledge representation can improve the ability to support several tasks while reusing knowledge within the guidelines and from external sources.
General Goal	To investigate cross-tasks computational support for CG that fosters knowledge reuse and maintenance in order to ultimately increase the adoption and efficacy of Computer-Interpretable Guidelines.
Specific Goal	To design a knowledge model and implementation for CIGs that can support multimorbidity analysis and literature search.
Research Questions	<p>RQ1: What is a suitable conceptual model to address the CG tasks?</p> <p>RQ2: What are suitable formal languages to formalize the CG model and tasks?</p> <p>RQ3: Is the Semantic Web paradigm suitable to support CG tasks and foster knowledge reuse?</p>

for each of the research questions: **Conceptual Analysis, Formalization and Implementation**

The cycle starts with the **Conceptual Analysis** phase:

1. Several inputs are used: (i) examples of clinical guidelines are studied; (ii) foundational ontologies or philosophical theories are used for supporting the conceptual analysis and modeling decisions; and (iii) existing CIG formalisms and approaches are reviewed to understand their benefits and limitations. The out-

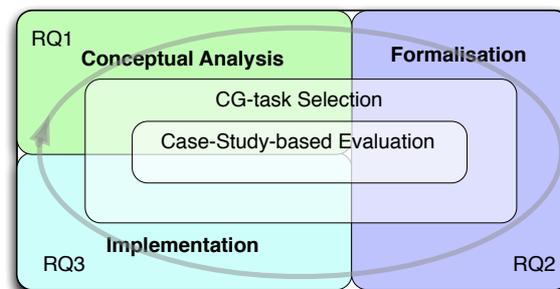


Figure 1.2: Schema for the methodology.

come is a conceptual model called TMR (Transition-based Medical Recommendation) model whose applicability needs to be evaluated for several tasks.

2. A CG-task is selected as a use-case and studied to identify requirements to be addressed, which supports the evaluation. Some limitations are identified and can be used to improve the conceptual model right way or as input in a next iteration of the cycle. The outcomes are case-studies with the achievable results.
3. External datasets are investigated to foster uniform data reuse by reinterpreting their data according to the TMR model.

The next phase is the **Formalization**:

1. The input is the conceptual model and the inferences needed to support the use-case requirements. Formalisms are chosen to describe the model and how to compute the derivations needed to fulfill the requirements. The outcomes are formal models and algorithms.

The last phase is the **Implementation**:

1. The input is the formal models and the algorithms. Semantic Web technologies are used to implement and compute the knowledge in order to fulfill the requirements. External data can be used to enrich or support the task execution. The outcome is a prototype implementing the models and algorithms.
2. The evaluation is performed by verifying the results obtained by executing the designed case-studies and comparing the obtained results against the expected results and/or results from the literature.

1.3 OPERATIONALISATION OF THE RESEARCH QUESTIONS

The first CG-task selected was the multimorbidity analysis, more specifically on supporting the detection of interactions when combining several guidelines of different diseases/conditions. For this task, three cycles were performed, described in the chapters as follows:

CHAPTER 2: TMR: A Core Conceptual Model for Clinical Guidelines - Applied to Comorbidity Analysis

RQ1.1 *What are the core concepts underlying clinical guideline recommendations?*

RQ1.2 *Can the model support the detection of interactions among recommendations from two guidelines (comorbidity analysis)?*

Based on: Zamborlini, V.; da Silveira, M.; Pruski, C.; ten Teije, A.; van Harmelen, F. *“Towards a Conceptual Model for Enhancing Reasoning about Clinical Guidelines: A case-study on Comorbidity”*, in Knowledge Representation for Health-Care (KR4HC). Lecture Notes in Computer Science, vol. 8903, Vienna, Austria: Springer International Publishing, 2014, pp. 29-44. [111]

My contribution: We jointly reviewed the literature and I developed the core conceptual model and its evaluation, under the supervision of my co-authors with whom I have jointly written the referred publication.

CHAPTER 3: Generic Rules for Detecting Interactions and Use of the External Source Drugbank - Applied to Multimorbidity Analysis

RQ1.3: *Can the approach support the detection of interactions among recommendations from more than two guidelines to support multimorbidity analysis?*

RQ2.1 *How to formalize generic rules for detecting interactions among several guidelines?*

RQ3.1 *How can we implement the model and rules and enrich them with an external knowledge source?*

Based on [114]: Zamborlini, V.; Hoekstra, R.; da Silveira, M.; Pruski, C.; ten Teije, A.; van Harmelen, F. *“Inferring recommendation interactions in clinical guidelines,”* Semantic Web, vol. 7, no. 4, 2016, pp. 421–446.

Invited extension of [110]: Zamborlini, V.; Hoekstra, R.; da Silveira, M.; Pruski, C.; ten Teije, A.; van Harmelen, F. *“A Conceptual Model for Detecting Interactions among Medical Recommendations in Clinical Guidelines,”* in Knowledge Engineering and Knowledge Management (EKAW). Lecture Notes in Computer Science, vol. 8876, 2014, pp. 591–606.

My contribution: I proposed the extension of the model to cover the detection of interactions, its formalization and the case studies; and I jointly (i) implemented a prototype, applied to the case studies and evaluated the results; and (ii) reviewed the literature; all under the supervision of my co-authors with whom I have jointly written the referred publication.

CHAPTER 4: Systematic Analysis of Types of Interactions with Evaluation in a Realistic Multimorbidity Case Study

RQ1.4 *How can we refine the TMR model to be more flexible and express other knowledge types that are important in the clinical context?*

RQ1.5 *How do the refinements affect the detection of interactions?*

RQ1.6 *How to measure the relevance of interactions?*

RQ1.7 *Does the approach stand up in a realistic case study developed with experts?*

RQ2.2 *How to formalize/extend the generic rules for detecting interactions given the TMR refinements and interactions analysis?*

RQ2.3 *How to formalize the relevance of the interactions?*

Based on [117]: Zamborlini, V.; da Silveira, M.; Pruski, C.; ten Teije, A.; Geleijn, E.; van der Leeden, M.; Stuiver, M.; van Harmelen, F. “Analyzing Interactions on Combining Multiple Clinical Guidelines”. Artificial Intelligence in Medicine journal, 2017.

Invited extension of [112]: Zamborlini V.; da Silveira, M.; Pruski, C.; ten Teije, A.; van Harmelen, F. “Analyzing Recommendations Interactions in Clinical Guidelines: Impact of action type hierarchies and causation beliefs” in Artificial Intelligence in Medicine AIME, 2015, pp. 317–326.

My contribution: We jointly reviewed the literature and I proposed (i) the extension of the model by adding concepts that increase expressivity and flexibility, such as actions hierarchy, causation belief and contribution to patient wellbeing, (ii) the systematic analysis of the interaction types and (iii) the formalization. The realistic case study was developed in collaboration with domain-experts. All under the supervision of my advisors with whom I have jointly written the referred publication.

CHAPTER 5: Generic Rules for Detecting Interactions Enhanced with Several External Sources¹

RQ1.8 *Can we make the model more generic so that other requirements could be addressed, potentially in other domains?*

RQ1.9 *How to uniformly exploit clinical knowledge from several external sources?*

RQ2.4 *How to formalize generic rules for detecting external interactions?*

¹ Chapters 4 and 5 can not be chronologically organized because the extended version of the first came after the extended version of the second, which means one uses (part of) the contributions of the other.

RQ3.2 *How can the rules be implemented in a more maintainable way?*

RQ3.3 *How can we implement the model in a way that knowledge provenance can be tracked?*

RQ3.4 *How to implement the reinterpreted clinical knowledge from external datasets?*

Based on [118]: Zamborlini, V.; Hoekstra, R.; da Silveira, M.; Pruski, C.; ten Teije, A.; van Harmelen, F. “Generalizing the Detection of Clinical Guideline Interactions enhanced with LOD”. *Communications in Computer and Information Science*, 2017.

Invited extension of [113]: Zamborlini, V.; Hoekstra, R.; da Silveira, M.; Pruski, C.; ten Teije, A.; van Harmelen, F. “Generalizing the Detection of Internal and External Interactions in Clinical Guidelines,” in *Proceedings of the 9th International Conference on Health Informatics (HEALTHINF)*, 2016, pp. 105-116. BEST STUDENT PAPER AWARD

My contribution: I extended the conceptual model to allow for (i) reasoning over the propagation of causation beliefs through the hierarchy; (ii) addressing causation and drug incompatibility as beliefs among events. I jointly extended the model with general concepts from norms theories. I extended the formalization and added interactions among several drugs due to polypharmacy; I jointly reviewed the literature, developed the nanopublication-based framework and acquired and analyzed the external datasets; I implemented the model and rules and I’ve performed the evaluation. All under the supervision of my advisors with whom I have jointly written the referred publication.

CHAPTER 6: Prototyping the Detection of Interactions using SWISH infrastructure

RQ3.5 *How to produce a functional prototype with reduced overhead on combining Storage, Application Logic and Presentation tiers?*

This is an application paper meant for motivating and describing the implementation of the prototype for detection of interaction using the SWISH framework. The purpose is to encourage the (re)use of the prototype by other researchers.

Based on [116]: Zamborlini, V.; Wielemaker, J.; da Silveira, M.; Pruski, C.; ten Teije, A.; van Harmelen, F. “SWISH for prototyping Clinical Guideline Interactions Theory” in *Proceedings of the 9th International Conference Semantic Web Applications and Tools for Life Sciences*, Amsterdam, Netherlands, 2016.

My contribution: I have jointly written Sections 1 and 4. I wrote 2.2 explaining the guideline theory and Section 3 explaining how SWISH was used. Jan Wielemaker extended the prototype with graphical visualization. All under the supervision of my co-authors with whom I have jointly written the referred publication.

The second task selected was the medical literature retrieval to support updating of clinical guidelines. For this task, one cycle (conceptualization, formalization and implementation) was performed.

CHAPTER 7: TMR-based Knowledge-driven Paper Retrieval to support updating of Clinical Guidelines

RQ1.10 *Can the TMR model support term-based query composition as part of the literature search for the task of updating a guideline?*

RQ2.5 *How to formalize the proposed method for composing queries?*

RQ3.6 *How to implement the composition of queries for literature search?*

Based on [119]: Zamborlini, V.; Hu, Q.; Huang, Z.; da Silveira, M.; Pruski, C.; ten Teije, A.; van Harmelen, F. “*Knowledge-driven Paper Retrieval to support updating of Clinical Guidelines: A use case on PubMed*”, Lecture Notes in Artificial Intelligence LNAI 10096, 2017.

Invited extension of [115]: Zamborlini, V.; Hu, Q.; Huang, Z.; da Silveira, M.; Pruski, C.; ten Teije, A.; van Harmelen, F. “*Knowledge-driven Paper Retrieval to support updating of Clinical Guidelines: A use case on PubMed*”, Proceedings of International Joint Workshop KR4HC - ProHealth, 2016, pp. 57-70.

Additional: Section 7.4.2 includes a semi-formal description of the proposed approach.

My contribution: I proposed how the TMR model can be applied to support the updating of a guideline and discussed with Qing Hu how the experiments should be designed. I’ve performed the part of the experiment that is new with respect to the previous experiment. I wrote the Sections 3 to 5 and 7, and we jointly wrote Sections 1 and 6. All under the supervision of my co-authors with whom I have jointly written the referred publication.

1.4 CONTRIBUTIONS

This section briefly presents the overview of contributions achieved in this thesis. We consider the main contribution an approach that on

the one hand is designed to be independent of a particular guideline-task but on the other hand is applicable to several tasks. The approach provides:

1. The TMR *conceptual model* allows for expressing important aspects of clinical recommendations inspired on general top-level theories, existing clinical guidelines and CIG formalisms. The model is evaluated for two CG-tasks: multimorbidity analysis and paper retrieval for guideline update.
2. The *formalization* allows for different implementations by using several technologies for: (i) the TMR model and its general inferences; (ii) the generic rules for detection of interactions while exploiting external knowledge from several datasets; (iii) algorithms to support the paper retrieval for guideline update.
3. The Semantic Web-based *implementations* provided as (i) a *prototype* that allows for detecting and visualizing 7 types of interactions among several recommendations using several external knowledge sources (ii) *code* that allows for extracting term-based queries from TMR-based guideline conclusions according to 4 query patterns. (iii) the *nanopublication-based framework* that allows for tracking provenance and changes of data from causation beliefs and recommendations.

In particular for the two tasks covered in this work the improvements with respect to the state of art are:

MULTIMORBIDITY ANALYSIS :

1. The number of detection rules does not increase with the number of guidelines, actions, drugs or effects;
2. The number of detection rules does not increase with the number of external datasets (only one or two importing rules per dataset);
3. In summary, new detection rules are sporadically needed only to enrich the approach with new types of interactions;
4. The generic and cumulative aspects of the rules allow for combining any number of recommendations and consequently guidelines;
5. The reuse of general purpose clinical knowledge datasets (or medical background) considerably increases the ability to cover all important interactions;

6. Although the external clinical datasets are mostly about drugs, and therefore also the examples used in this thesis, the proposed approach is not restricted to drugs. It means that if an external dataset would describe incompatibilities among examinations or other therapies, this data can be seamlessly used in our approach.

LITERATURE SEARCH - GUIDELINE UPDATE :

1. The selection of keywords and their composition into a query is supported by a method that takes into account the semantic role of terms in a CG conclusion according to the TMR model, as an alternative to calculating the importance of medical terms;
2. TMR-based method supports the addition of alternative descriptions to the original terms and their composition into a query;
3. An experiment shows that the proposed approach can be at least as good as the compared method, and that the addition of alternative descriptions allowed to retrieve more of the gold standard papers (recall) without increasing much the total number of retrieved papers;
4. Although the addition of alternative descriptions was performed manually and biased to the gold standard, an analysis of the papers that were not retrieved was conducted in order to understand the kind of alternative descriptions that could improve the recall. We believe that the results of this analysis is a step on the direction of pursuing generic automated acquisition of alternative descriptions for enriching the queries.

The structure of this thesis is: Chapters 2 to 7 present the development of the proposed research as previously discussed; and Chapter 8 presents a discussion of the achievements, strengths and limitations of this thesis, as well as future work that is necessary to thoroughly address the mentioned challenges. Because the chapters are based on separate publications and to keep them self contained, some content is repeated from one chapter to another, particularly regarding motivation, related work and background of case studies. In order to favor readability, we have harmonised the terminology and figures across different chapters.