

CONCLUSION

*Everything should be made as
simple as possible, but not
simpler*

Albert Einstein

This chapter summarizes the main achievements of this work while reflecting about the three main research questions. It discusses the strengths and limitations of the proposed approach and points to future work in a short term and research directions in a longer term.

8.1 MAIN ACHIEVEMENTS

As discussed in Chapter 1, the specific challenge here addressed is:

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

This thesis is a step in the direction of finding a solution to this challenge. The present research is not meant to propose one more formalism for Clinical Guidelines, but to investigate the knowledge underlying CGs that are necessary to address several tasks. The conceptual analysis phase, and in particular the use of top-level theories (when available), play an important role on guiding task-independent modeling choices. Moreover, the formalisation phase allows to express the required computing independently of a particular implementation technology. Therefore, the proposed approach is designed to be both task- and technology-independent, even though the evaluation is performed through specific CG-tasks and technology.

The Semantic Web provided a suitable environment for the implementation by fostering the reuse of large datasets as Linked Open Data, besides the provision of reusable knowledge. This feature is crucial to the goal of producing a sustainable solution for supporting tasks in the guideline life-cycle.

The evaluation for multimorbidity analysis and literature search showed relevant contributions with respect to the state of art, summarized as:

MULTIMORBIDITY ANALYSIS : a fixed number of generic rules for detecting several interaction types among any number of recommendations, enhanced by several external knowledge sources.

LITERATURE SEARCH : a knowledge-based systematic way to compose search-queries using the content from guidelines' conclusions and to enrich the queries with alternative descriptions that can be taken from external sources.

8.2 ANSWERS TO THE RESEARCH QUESTIONS

The investigation performed in each of the chapters concerning the main questions (RQ₁, 2 and 3) is summarized in Table 8.1. The evaluation is based on a selected task and the designed case-studies¹ (compared against related work or our previous results) and it is called 'conceptual evaluation' (theory) when it is not implemented or 'applied evaluation' (practice) otherwise.

Hereby we recapitulate the research questions with their obtained answers:

RQ₁ *What is a suitable conceptual model to address the CG tasks?*

This research proposes the TMR model as a potentially suitable conceptual model to address CG-tasks. It was evaluated for two tasks: multimorbidity analysis and literature search. In both cases, the TMR model was suitable and promoted improvements with respect to existing approaches.

This research question was investigated through several chapters of this thesis as subquestions:

RQ_{1.1} *What are the core concepts underlying clinical guideline recommendations? (Chapter 2)*

By investigating some top-level theories, clinical guidelines and CIG formalisms, we have proposed a first version of the TMR model containing the core concepts underlying clinical guidelines, namely: *recommendation, action type, transition and situation type*.

¹ Abbreviations in the Table: DU (Duodenal Ulcer), TIA (Transient Ischemic Attack), OA (Osteoarthritis), HT (Hypertension), DB (Diabetes) and BC (Breast Cancer).

Table 8.1: Summary of Research Questions Investigations

	Chapter 2 [111]	Chapter 3 [114]	Chapter 4 [117]	Chapters 5 & 6 [116, 118]	Chapter 7 [119]
Conceptual Analysis Model	Core Concepts for Clinical Guideline Core Concepts: Recommendation, Action, Transition and Situation	Generic Internal Rules + DrugBank	Systematic Analysis of Interaction Types - Hierarchy among Action Types, Causation Beliefs among Action Type and Transition, Trope and Derivative, Contribution - Qualitative aspects	Generic External Rules + LOD sources - Hierarchy among Types with grouping criteria, Causation and Incompatibility Beliefs among Event Types, Norms and Regulations - Reinterpretation of external data according to TMR model	Knowledge Driven Paper Retrieval
Task selection + requirements	Comorbidity Analysis	Multimorbidity Analysis	Multimorbidity Analysis	Multimorbidity Analysis + Polypharmacy	Paper Retrieval -> Composing Queries
Conceptual Analysis Task	Interaction Types: Contradiction, Repetition, Alternative and Repairable	Interaction Types: Contradiction, Repetition, Alternative and Repairable among recommendations and AlternativeDrugs and IncompatibleDrugs between recommendations and external data	Systematic Analysis of interaction Types: Divergent Causations, Contradiction, Repetition, Alternative, Repairable, Side-Effect, Compliance and Safety	Interaction Types: Contradiction, Repetition, Alternative and Repairable; External Alternative, Incompatible Action and Incompatible Effects (Side-Effect)	- Knowledge-based method for composing queries and adding alternative descriptions - Analysis of reasons why gold-standard papers were not retrieved by using original terms
Design of case study and experiments	adapted from literature on combining 2 guidelines DU + TIA	adapted from literature on combining 2 and 3 guidelines DU+TIA and OA+HT+DB	adapted from literature on combining 3 guidelines OA+HT+DB	adapted from literature on combining 3 guidelines OA+HT+DB	adapted from literature on retrieving paper for updating Dutch BC guideline

				in collaboration with experts on combining Exercise therapy for BC with OA+HT+CHF			
RQ2	Formalization of Model Formalization of Task	FOL	Set Theory	Adapted Generic FOL Rules (v.3) for Internal Interactions	FOL	Algorithms for query composition	
		Generic FOL rules (v.1) for detecting internal interactions	Extended Generic FOL Rules (v.2 and v.4) for internal interactions	Generic FOL rules for detecting external interactions from any dataset			
		Specific rules for detecting external interactions using Drugbank-based vocabulary					
RQ3	Semantic Web Technology External Data and Vocabulary Implementation result	Cumulative interactions	Extended Cumulative interactions	Cumulative interactions			
		OWL/RDF + SRWL + SPARQL		RDF + SWI-Prolog	RDF + SPARQL		
		Dataset: Drugbank			Literature Repository: Pubmed		
		Web application to calculate and visualize interactions enriched with drugbank data		Prototype in SWISH environment to calculate and visualize interactions enriched with several datasets			- SPARQL code for composing term-queries
Type of Evaluation		Applied Evaluation	Conceptual Evaluation	Applied Evaluation	Applied Evaluation	Applied Evaluation	

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

Yes, the concepts underlying TMR model allowed to conceptually repeat an experiment from the literature on combining two guidelines, while supporting the rationale for generically identifying interactions.

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

Yes, the TMR model was extended to represent the interaction relation among several recommendations (internal interactions) or among recommendations and external knowledge (external interactions). The reification of this relation allows for (i) relating as many recommendations as needed and (ii) qualifying the interactions without the need for higher order modeling features.

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? (Chapter 4)*

The TMR model was extended to include refinements that increase the expressive power to bring it closer to the clinical knowledge: (i) *hierarchy* of action types; (ii) causation as *beliefs* endorsed by a source; (iii) expected *contribution* of an action type; (iv) *qualitative values* are also often adopted in the medical field; (v) property (*trope type*) affected by transition and *derivative*.

RQ1.5 *How do the refinements affect the detection of interactions? (Chapter 4)*

The refinements enhanced the detection of interactions allowing to detect more interactions (e.g. due to the hierarchy) and also different types of interactions (e.g. divergent beliefs or safety) via a systematic analysis.

RQ1.6 *How to measure the relevance of interactions? (Chapter 4)*

We have adopted a flexible way to measure the relevance of interactions. To this end, the interaction have as attributes several strengths that might have more or less importance in different contexts: modal, deontic, belief and causation strengths.

RQ1.7 *Does the approach stand up in a realistic case study developed with experts? (Chapter 4)*

A realistic case study was developed in collaboration with domain experts. The result was positive and lessons were learned that can allow the model to be improved in future cycles.

RQ1.8 *Can we make the model more generic so that other requirements could be addressed, potentially in other domains? (Chapter 5)*

Yes, the model was extended to be more general, allowing to represent: (i) hierarchy (subsumption) among *Types*; (ii) *Grouping Criteria* to justify the belonging to *Categories* based on expected *Transitions* (effects); (iii) *Causation Beliefs* among *Event Types*; (iv) *Incompatibility* as a *Belief* among *Event Types*, expressing any incompatible events; and finally (v) *Recommendations* and *Clinical Guidelines* as specializations of *Norms* and *Regulations*, since the general idea underlying the model is also the general idea of other regulations in other domains.

RQ1.9 *How to uniformly exploit clinical knowledge from several external sources? (Chapter 5)*

In order to reuse knowledge from external sources we have reinterpreted them in terms of the TMR model. For example, drug categories from Drugbank were reinterpreted as causation beliefs. By using this reinterpretation strategy for each dataset, the clinical knowledge can be uniformly used for several tasks.

RQ1.10 *Can the TMR model support the term-based query composition as part of the literature search for the task of updating a guideline? (Chapter 7)*

Yes, a knowledge-driven method for composing term-based queries is developed taking into account the semantic-roles *Action* and *Effect* that the terms can assume in a guideline conclusion. This method was applied to an experiment from the literature and was shown to perform at least as good as the related work. Moreover, the model allows to easily extend the original description with alternative ones, which was shown to considerably improve the recall without increasing much the number of retrieved papers.

RQ2 *What are suitable formal languages to formalize the CG model and tasks? Based on the conceptual analysis, FOL and Set Theory were*

enough to formalize the concepts and relations within the TMR model as well as the inferences relevant for the purposes of this work. Regarding the tasks, FOL was enough to achieve satisfactory results for detection of interactions, while algorithms were used to describe the composition of conclusions' terms into queries.

This research question was investigated through several chapters of this thesis as subquestions:

RQ2.1 *How to formalize generic rules for detecting interactions among several guidelines?* (Chapter 3)

A small set of rules is formalized using First Order Logic in a generic way, i.e. no specific action, drug or effect are mentioned in the rules. In order to have a finite set of rules that combines as many recommendations as needed, the two-step solution was: first, designing rules that infer a unique interaction of a certain type among any pair of recommendations; and second, designing rules that accumulate interactions of the same type relating common recommendations.

RQ2.2 *How to formalize/extend the generic rules for detecting interactions given the TMR refinements and interactions analysis?* (Chapter 4)

The rules are derived from the systematic analysis (RQ1.5) so that they are known to cover all the possibilities analyzed.

RQ2.3 *How to formalize the relevance of the interactions?* (Chapter 4)

We proposed a way of calculating interaction strengths using either average or product of several features, such as deontic, belief or causation strengths (normalized from $[-1,1]$).

RQ2.4 *How to formalize generic rules for detecting external interactions?* (Chapter 5)

When the external knowledge is reinterpreted according to the TMR model (RQ1.9), generic rules for detecting external interactions can be written independently of specific vocabularies from external sources, in a similar way as we did for the internal interactions.

RQ2.5 *How to formalize the proposed method for composing queries?* (Chapter 7)

We designed algorithms for describing how to compose term-based queries using the TMR model.

RQ3 *Is the Semantic Web paradigm suitable to support CG tasks and foster knowledge reuse?*

Yes. The Semantic Web paradigm is known to foster knowledge/-data sharing and interoperability. It indeed favored the reuse of several clinical datasets available as Linked Open Data, e.g. Drugbank, Sider, AERS, LIDDI. This allowed for taking advantage of medical background knowledge for the detection of interactions at a larger scale than the related work on multimorbidity, and can easily be extended with new knowledge sources that are increasingly being published as LOD². Although the literature search task is not yet directly benefiting from the Semantic Web, we have paved the way to exploring existing medical vocabularies (see Sect. 8.3). Finally, the reasoning requirements scoped in this thesis for the CG-tasks were also fulfilled.

This research question was investigated through several chapters of this thesis as subquestions:

RQ3.1 *How can we implement the model and rules and enrich them with an external knowledge source? (Chapter 3)*

In the first experiment of prototype implementation we have adopted standard Semantic Web Technologies OWL/RDF, SWRL and SPARQL. Via this prototype we were able to demonstrate the use of Drugbank dataset in order to enrich the detection of interactions with alternative and incompatible drugs.

RQ3.2 *How can the rules be implemented in a more maintainable way? (Chapter 5)*

In the second prototype implementation, we have adopted the SWI-Prolog language to implement the inference rules. Its greater expressive power favored understandability and maintenance.

RQ3.3 *How can we implement the model in a way that knowledge provenance can be tracked? (Chapter 5)*

We have proposed a framework based on use of Semantic-Web Open Vocabularies such as Nanopublication, Prov and

² The number of Life-Science datasets in the LOD-cloud reported in 2014 was 83 http://lod-cloud.net/state/state_2014/. The current version <http://lod-cloud.net/> does not report a number, but the graph shows that it clearly more than doubled, from which most are biomedical datasets.

OpenAnnotation. Although it allows for a very detailed provenance tracking, it is also a very verbose solution that uses three named graphs per assertion.

RQ3.4 *How to implement the reinterpreted clinical knowledge from external datasets?* (Chapter 5)

This was achieved using a few Prolog rules per dataset to import the data combined with the Nanopublication structure for tracking provenance.

RQ3.5 *How to produce a functional prototype with reduced overhead on combining Storage, Application Logic and Presentation tiers?* (Chapter 6)

The SWISH environment was adopted to implement a prototype for interaction detection. Several functionalities from SWISH favored the development and maintenance of the prototype.

RQ3.6 *How to implement the composition of queries for literature search?* (Chapter 7)

We have developed SPARQL codes to select and compose the terms related to the semantic-roles Action and Effect. The queries were successfully used in an existing prototype that submits them to a Pubmed API and retrieves the corresponding medical papers.

8.3 LIMITATIONS, LESSONS LEARNED & FUTURE WORK

This section presents limitations and lessons learned with respect to general aspects and specifically about the two investigated CG-tasks, together with possible future work.

Operational issues

MANUAL CG-DATA ENTRY Our results rely on the assumption that the guideline knowledge is available according to the TMR format, i.e. in practice we manually instantiate the model. On the one hand this can be seen as a limitation but on the other hand, this is exactly what allows us to extrapolate the limitations imposed by both the plain-text paradigm and by the modeling choices biased for particular tasks. It means that the proposed approach is not yet scalable

in this aspect, but we consider this price worth to pay when it allows an alternative-possibly-successful path for addressing the posed challenge.

In order to address this issue we consider three possibilities: (i) semi-automatically extracting knowledge from the guideline text; which can speed up the process but is known to require a validation step for correcting or completing the missing/implicit data; (ii) investigating if/how knowledge from guidelines already formalized using existing CG formalisms (meant for execution) can be reused; (iii) the ideal scenario when the guideline development process includes the creation of some structured knowledge (in the line of [26, 82, 91]).

FEW CASE-STUDIES Only a few case studies were used for evaluation. The lack of easily accessible case studies (like a benchmark), together with the difficulties on developing the case studies from scratch or adapting them from the literature make the design of case studies a very laborious task, particularly for multimorbidity. We are currently working with scientists from the VU Medical Center in Amsterdam and from Beijing University for improving/developing case studies. Extending the model with new features (e.g. time) would allow the existing but also new case-studies to become more and more realistic. We plan to make the developed case studies available to be reused by other scientists, as we did for the case studies presented in this thesis.

FEW CG-TASKS Although we aim to produce a model that can support several CG-tasks, the proposed approach is evaluated for only two CG-tasks multimorbidity analysis for CG development and literature search for CG update). However, the two addressed sub-tasks can be applied to other CG-tasks, e.g. literature search is useful for guideline development too. The applicability to other CG-tasks will be investigated, such as the adaptation of guidelines to specific requirements in certain countries/regions [65, 84].

Conceptual issues

FOCUS ON TREATMENT For the sake of scope restriction we have focused on the treatment part of the guidelines, rather than diagnosis, prognosis or prevention. A dedicated investigation is necessary

to avoid the modeling choices for these issues to be biased by the current modeling choices evaluated for treatment.

TIME AND QUANTITY The present approach is neutral with respect to time and quantity aspects, as we needed to restrict scope and they were not essential to current stage of the research. We have already started an investigation on merging our approach with the one proposed in [5, 70], which uses a time-ontology and implements temporal reasoning techniques such as planning for detecting temporally-expanded interactions when applying the guideline to a particular patient. The causation structure implicitly reflects one time aspect: the action starts before its effect. This and other elements in the model can be extended and expressed in terms of the well known Allen's relations for time intervals [2].

The quantity aspect, often addressed as the amount of drug, also has some interesting traits to consider: e.g. it is intrinsically related to the time aspect, when for example repeating a certain action within a time interval is equivalent to another action with higher 'quantity' (e.g. *biking twice for 15km can be equivalent to biking 30km.*)

Those improvements would allow, for example, to calculate the specific time and quantity interval setting in which an interaction occur, or, alternatively, if it can be avoided by adjusting those values.

SITUATION AND GOAL A deeper investigation about approaches/theories for representing and reasoning over situations and goals is needed to support, for example, addressing non-trivial matching of situations (due to hierarchy, overlapping or part-hood) but also calculating benefits and harms of interventions (as contributions to goals) [18, 25]. Other important aspects to consider are (i) when a situation type should be expressed as the pre-condition for a transition to happen or for a recommendation to hold; and (ii) the temporal extension of a situation (e.g. intermittent fever). It would allow, for example, to avoid interactions when recommendations' pre-conditions do not overlap (e.g. *if age < 10 don't give aspirin × if age > 15 give aspirin*).

QUALITATIVE ASPECTS One qualitative aspect for transitions is currently expressed, namely derivative (*increase, decrease or maintain*). However, other aspects such as rate and magnitude are also important for comparing the similar transitions and choosing the most suitable one. These aspects might reveal the efficiency and efficacy of the

actions believed to promote similar transitions. For example, the clinical studies and guidelines often compare actions as being ‘better than’ or ‘as good as’ another action or with placebo.

FOUNDATIONAL ONTOLOGIES Unfortunately, not all the modeling choices could be founded in some top-level theory. Several discussions with experts in this field supported the modeling choices made in this thesis, but they are not formally grounded in an existing theory. The difficulties found for grounding the present research can serve as input for extending existing theories.

Formalization issues

OTHER FORMALISMS First Order Logic has shown to be enough to handle the issues addressed in the present research. More expressive formalisms can be investigated to provide more powerful reasoning features. However, the implementation might also become more computationally expensive. These trade-offs need to be investigated together with the requirements of each CG-task.

Implementation issues

SEMANTIC WEB The Semantic Web provides a rich environment that can definitely be further explored. The integration with external knowledge sources brought about various open questions: (i) how to deal with inferring reified relationships; and (ii) how to select the appropriate identity criteria [9] and handle contextualized semantics for ‘owl:same_as’ [12].

EXISTING CLINICAL VOCABULARIES/ONTOLOGIES The standardized medical vocabularies do play an important role in the integration with external sources. For example, we do use UMLS³ in our experiments reported in Chapters 5 and 7 respectively to map situations or actions to external sources and (manually) acquire alternative descriptions for medical terms. However, more investigation is needed to thoroughly explore them within the TMR approach, e.g. importing relevant relationships among medical concepts from SNOMED-CT⁴, such as hierarchy or parthood. It could allow for enriching both

³ <https://www.nlm.nih.gov/research/umls/>

⁴ <http://www.snomed.org/snomed-ct>

the detection of interactions and the query composition for literature search by (semi)automatically acquiring alternative descriptions.

EXPLORING EVIDENCE CHAINS The implementation proposed in this work allows to keep track of provenance data powered by the Nanopublication framework. This feature can be further explored by: (i) tracking more fine grained provenance, allowing easy access to a specific piece of text of the evidence document [37]; (ii) automatically calculating or verifying the evidence level for a certain recommendation [43]; and (iii) having access to metadata about the evidence, such as the eligibility criteria of the clinical studies [34, 42]. Those features are considered relevant by some healthcare professionals that worry about blindly following the recommendations within a guideline [23].

Hereby we highlight some improvements specif for *Multimorbidity Analysis* and *Literature Search* tasks.

Multimorbidity analysis

MEASURES FOR FILTERING INTERACTIONS The excess of interactions is not well received in the medical community, as current alert systems are accused to cause ‘alert fatigue’ [52]. Avoiding it requires good mechanisms for selection/prioritization. This mechanism should not be fixed but adaptable to each circumstance, for example, in an emergency situation, the contradiction with a recommendation to not give aspirin to avoid bleeding could be ignored. In other circumstances, interactions derived from recommendations with low evidence level could be ignored, or with low probability. Some measurements were introduced in Chapter 4, whose calculations can be further investigated through more case-studies in collaboration with experts.

NEW TYPES OF INTERACTIONS Adding new features in the model might allow for detecting more interactions or even new types of interactions, as happened in Chapter 4. Moreover, we have started investigating a number of interactions presented in [62] that are considered relevant by the domain experts.

Literature Search:

PICO AND PIPOH Besides exploring the conclusions, the key questions designed to guide the literature review could also be a source for automatically composing the queries in a similar fashion to the proposed approach. Two existing methods for designing the key questions, called PICO (**P**opulation, **I**ntervention, **C**omparison, **O**utcome) and PIPOH (**P**atient, **I**ntervention, **P**rofessionals, **O**utcome, **H**ealthcare settings), partially overlap with the current version of the TMR model.

LITERATURE REPOSITORY Currently we use only one literature repository API, from PubMed, which allows only for term-based queries. We plan to investigate the use of APIs that allow for semantic-search, such as the one proposed in [80] that relies on RDF structured knowledge extracted from papers' abstracts on PubMed.

In summary, new iterations of the proposed methodology are still needed to better address the investigated CG-tasks but also to address new ones and ultimately evaluating the contribution of the proposed knowledge model to improve the current guideline-system. However, interesting results were already obtained.

8.4 OUTLOOK

Enormous efforts are being invested on producing clinical knowledge and making it available in clinical practice. However, the increasing amount of knowledge and its complexity make it difficult to efficiently apply it in benefit of the patient. So, retaking the big challenge in the medical informatics discussed in Chapter 1:

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

We believe the answer inevitably goes towards fostering smart clinical knowledge reuse. Or, in other words, the system needs to be 'sustainable' in the sense that new knowledge does not make the system obsolete or increase the complexity of its tasks but enhances them and existing knowledge can be reused for other (unforeseeable) purposes [22]. This challenge is then twofold: (i) healthcare professionals should invest on providing structured reusable clinical knowledge;

and (ii) computer scientists should invest on providing solutions that favor smooth knowledge acquisition and reuse.

This scenario imposes a socio-technical trade-off to knowledge representation models meant to address this challenge:

the model should be as simple as possible: it is important from both a scientific point of view and from a practical point of view, to avoid the model to become unnecessarily expensive to construct, complicated to use and expensive to compute;

but the model should not be simpler: it is important for the model to be sufficiently general to be applicable in several tasks, and sufficiently expressive to properly address the tasks.

It is true that not all types of knowledge might be relevant for all the tasks. But this means we need to know which type of knowledge need to be shared among which tasks, and also where in the cycle it can be included. In this line, we plan to investigate other CG-tasks in order to improve the model to comprehend the whole guideline life-cycle while fostering the knowledge reuse. This includes investigating the overlap with existing CG formalisms and how the knowledge can be shared.

Although medical vocabularies have been standardized (e.g. SNOMED-CT⁵, ICD⁶), no standards were successfully adopted for the representation of clinical knowledge underlying guidelines. One may claim that not even the guidelines are successfully adopted regardless to computer representation standards. But on the other hand, these standards can be the key to promote adherence to guidelines, to speed up the update process, to check for compliance and so on.

In addition, providing special views of the clinical knowledge for each type of stakeholders is essential for effectively supporting high quality care. Both healthcare professionals and layman can/should be empowered by contextualized knowledge in a suitable format. In this line, we plan to investigate if/how we can phrase the clinical support in the context of ICT4D (Information and Communication Technology for Development) as a special clinical guideline cycle. Briefly, this issue regards providing suitable advices to community agents or patients in countries or regions where access to healthcare facilities

⁵ <http://www.snomed.org/snomed-ct>

⁶ <http://www.who.int/classifications/icd/en/>

is difficult or non-existent. Of course, ICT4D issues involve much more than ICT, e.g. cultural aspects.

We understand this is an ambitious outlook, but we believe its success would promote important results and ultimately support well-informed healthcare decisions to benefit people's lives around the world.