Kerstin Denecke* and Claire Chalopin

A generic concept for the development of model-guided clinical decision support systems

Abstract: Disease development and progression are very complex processes which make clinical decision making non-trivial. On the one hand, examination results that are stored in multiple formats and data types in clinical information systems need to be considered. Beyond, biological or molecular-biological processes can influence clinical decision making. So far, biological knowledge and patient data is separated from each other. This complicates inclusion of all relevant knowledge and information into the decision making. In this paper, we describe a concept of model-based decision support that links knowledge about biological processes, treatment decisions and clinical data. It consists of three models: 1) a biological model, 2) a decision model encompassing medical knowledge about the treatment workflow and decision parameters, and 3) a patient data model generated from clinical data. Requirements and future steps for realizing the concept will be presented and it will be shown how the concept can support the clinical decision making.

Keywords: Model-based Therapy; Clinical Decision Support; Disease Modelling

DOI: 10.1515/CDBME-2015-0080

1 Introduction

Individualized patient treatment requires considering all relevant clinical information items and weighting them according to the patient-specific situation. Additionally, knowledge and information on the development of diseases as well as on influence factors including the underlying biological processes are required. Further, knowledge and experiences about treatment and treatment plans, risk factors, problems and complications etc. need to be accessible. However, the challenges we face nowadays in clinical settings are that (1) clinical data is distributed in several systems and available in multiple formats, and (2) best practices and information about treatment plans and options are captured in biomedical literature or clinical practice guidelines, but are not formalized. Not only accessing the relevant data during diagnosis and treatment among the large set of available patient data is a challenge, also its mental integration, analysis and interpretation. Understanding the dependencies among examination results, biological processes and medical conditions is crucial for making proper decisions. For these reasons, support in evidence-based medicine is important to ensure high quality treatment in the different pre-, intra and perioperative stages, and also beyond the domain of surgery.

The concept of model-based therapy [1, 2] addresses the challenges outlined before by relying upon knowledge and data models that provide an evidence-based, reproducible basis for decision making. Among others biological processes can be modelled such as the effect of hormones on disease development. A digital patient model integrates various patient-specific information entities that need to be considered within therapy planning and clinical decision making. It further describes relations between clinical parameters and sociological factors related to a specific clinical pathology. In this work, we will introduce a concept of model-based decision support that links knowledge about biological processes, treatment decisions and clinical data. It consists of three models: 1) a biological model, 2) a decision model encompassing medical knowledge about the treatment workflow and decision parameters and 3) a patient data model generated from clinical data. We will collect the requirements for the realization of the concept.

2 Three-part modelling concept

Our concept of model-based decision support relies upon three types of models (see Fig. 1):
- A biological model describes the disease mechanisms and underlying biological processes and parameters (e.g. hormone-disease effect, eye functions, latitude of a tumor).
- A decision model encompasses the medical knowledge on the diagnostic of a disease, the therapy decision including the therapy options, their risks for the patient and their relations to single information items relevant in decision making.
A patient specific data model includes and integrates the relevant patient data (laboratory tests, examination results, medical report, signal and image data).

The decision model reflects the intellectual reasoning performed by clinicians during diagnosis, treatment and monitoring of the disease. It includes decision parameters and their relations to therapy options. Further, the patient model integrates the relevant clinical patient data of different types. The biological model can also contain information on the efficacy of drugs given some biological circumstances of patients for a personalized therapy plan. The models clearly contain information relevant for each other. For example, predictions about cell growth in the context of tumor diseases impact the predictions on treatment options or expectancy of life. However, to reduce complexity and to enable their re-use (in particular re-use of models on biological processes), it is necessary to separate biological models from the other model types even though they are related to the same disease.

Figure 1: By integrating information on biological processes, decision processes and clinical data, a model-based decision support system can be built. Information from one model is relevant for calculations in another one.

We suggest building a clinical decision support system on top of a model that integrates these three model types in the context of a specific medical condition. All relevant information items of a decision making process together with the relevant decision parameters and the dependencies among biological and clinical aspects of a disease are linked in an integrated model. The resulting decision support system will be capable of providing a prognosis of treatment outcome, providing hints to missing examination results or just providing an integrated view on data and therapeutic options with the relationships among them. Further, it helps to learn dependencies between biological processes and their influence on factors relevant for patient treatment.

3 Use case

In this section, we want to describe a use case for our three-model concept and specify the concrete requirements based on the general collection from the section before. To demonstrate the application of the three models as basis for a model-based clinical decision support system, consider the following use case of treatment and monitoring of patients suffering from pituitary adenoma.

Pituitary adenomas are common benign tumours of the pituitary gland. Some tumours secrete one or more hormones in excess. Such so-called secretory pituitary adenomas are usually found due to hormonal imbalances that affect bodily functions. The treatment of these tumours requires the competences of interdisciplinary clinicians, endocrinologists, neurosurgeons and radiologists. The treatment decision is usually performed by considering a broad range of patient data such as clinical symptoms (infertility, acromegaly, deterioration of the visual functions) or results from clinical examinations (hormone test, image data acquisition, ophthalmological examination).

A model-based decision support system for supporting in pituitary adenoma treatment comprises a biological model reflecting the healthy biological processes of the pituitary gland, a decision model on the treatment options and their dependencies to clinical and biological parameter. When applied to a specific patient record, the patient data model provides the patient-specific data. Based on the decision model, our model-based decision support system sends notifications when relevant patient data for deciding for a treatment is missing. Once all relevant data is available to the system, it compares the healthy processes to the given pathological data of patient, suggests possible treatments (surgery, medical treatment, radiotherapy or simple observation) and predicts the outcome and risks for the single treatment options. The calculation considers the severity of the clinical symptoms, the findings of the clinical examination (size and location of the tumours). Together with the general state of the patient this leads to the choice of the most relevant therapy for the patient.

From this use case, it becomes clear that for its realisation among others a biological model on hormone effects is required as well as a therapy decision model for pituitary adenoma, and a standardized representation and storage of the patient-specific data.
4 Requirements

More specifically, in order to realize a clinical decision support system based on the three models, methods and challenges in three main areas need to be considered:

- **Modelling of biological processes, diseases and clinical decision making**: Methods for semi-automatic modelling of diseases, biological processes and clinical decision processes need to be developed.

- **Semantic linking**: Methods for linking data and biological models with patient models through ontologies including methods for integrating the underlying ontologies are necessary. Semantic annotations can be produced by extracting relevant terms from texts or classifying image data automatically.

- **Ontology developing**: Ontologies and classification systems for describing the concepts of the models and providing the relations for inferencing are required.

More details on these requirements are provided in the following.

**Requirement 1: Model development.** For generating biological and decision model, mathematical modelling techniques such as Multi-Entity Bayesian Networks (MEBN) can be exploited. MEBNs are based on the Bayesian probability theory with first-order logic [3]. In general, MEBN are similar to Bayesian Networks (BN) with nodes, edges and conditional probabilities. MEBNs have been used to model and simulate abstractions of real-life situations and processes [3]. They are based on both, mathematical correlations between random variables (representing uncertainty and what is known with reasonable certainty) and a graph, denoting the conditional dependency structure between these random variables. Information entities from the domain of interest, e.g. results of clinical examinations, medical imaging, patient behavior and patient characteristics (e.g. age, gender, and tobacco and alcohol consumption) are integrated in the model as random variables [4].

However, developing decision models or biological models requires expertise in the areas to be modelled which results in a very time-consuming process. In our experience, it took a team of two persons one year to build a therapy decision model for laryngeal carcinoma [5]. Methods need to be developed for learning relevant information items and their correlations from clinical practice guidelines, biomedical literature and learned from patient data to build semi-automatically a graph-like structure that connects the relevant information items for a concrete disease or organ. In this context methods for information and relation extraction can be exploited to collect relevant information items and their relations from textual documents (see requirement 2). Given the fact, that not all decision relevant information is described in literature and other sources (e.g. experiences), methods are necessary to support physicians in validating the generated model and in easily adding missing information items or relations. For example, starting from an ontology restricted to a specific subdomain, concepts could be suggested to the biologist or physician to be selected and included into the model.

For generating patient data models, relevant clinical data have to be collected, the semantic information extracted and stored in appropriate data bases. The set of relevant patient data for the treatment of a given disease is generally determined from observations in the clinic and from interviews with physicians. In order to be able to process the patient data, the semantic information has to be extracted (see requirement 2), using automatic techniques if possible, and stored in appropriate databases. These last ones enable to represent the information in a structured and standard way in order to facilitate the information access and interpretation. For example, the openEHR Foundation proposes appropriate tools for the patient data modelling [6, 7].

**Requirement 2: Methods for collecting information items and extracting information.** For disease modelling, parameter, diagnoses, and other important information normally hidden in free text need to be extracted, analysed automatically and mapped to an underlying ontology. For these purposes, natural language processing techniques are required which typically involve tokenization and sentence splitting, stop-list filtering, stemming, lemmatization, POS tagging, chunking and shallow or deep parsing. For these basic tasks open source software is available, for example natural language processing tools provided by the Stanford language processing group (http://nlp.stanford.edu/software/index.shtml, accessed: 22.02.2015). Named-entity recognition aims at identifying within a collection of text all of the instances of a name for a specific type of thing [8]. Examples of named entity categories in the medical domain include diseases and illnesses, symptoms, procedures or drugs. Domain-specific named entity recognition can be enabled through tools such as MetaMap (http://metamap.nlm.nih.gov/, accessed: 22.02.2015) or OpenCalais (http://www.opencalais.com/, accessed: 22.02.2015). Moreover, segmentation methods are appropriate to extract the semantic information from image data, in general represented by the anatomical structure to be examined such as a tumour [9, 10].
The biological and clinical data relevant for model-generation ranges from text-based content, to medical images, numerical data, and genomic data whose processing requires a variety of methods due to different data characteristics and content types. An optimized data analysis system for the target domain is necessary that provides structured data or information surrogates with semantic annotations. Besides the relevant information items, it is crucial for semi-automatic model generation, to determine the dependencies of entities as described in texts. For this purpose, relationship extraction methods are required that make use of semantic knowledge and natural language processing.

**Requirement 3: Ontologies.** Mapping of extracted terms to concepts of ontologies is necessary to get a standardized representation and to assign semantic categories to terms. Further, the use of ontologies and formal concepts of a domain is necessary for adapting inference functionalities to various situations and application scenarios, but also to make unstructured text automatically processable within the context of reasoning. In the domain of medicine, there are already ontologies available that describe medical concepts such as disease or agents (viruses) including MeSH, SNOMED CT or UMLS. They can be considered as basis to identify and formally describe relevant clinical concepts.

**Requirement 4: Data and model linking.** For realizing the concept, methods are required to link the models or to exploit the output from one model in another one. In particular in the patient model, data linking is necessary. Patient information is currently stored in multiple information systems. Semantics are unavailable, i.e. data items are not semantically annotated which complicates annotation, analysis and linking.

Integration of the patient specific model with the biological and decision model is crucial for interpreting patient data, making predictions, evaluating risks and of course for decision making. Remember that the decision model comprises the information entities that should be considered in decision making. Instantiating these entities with real patient values would lead to a patient-individual decision model. A semantically integrated model is necessary to answer questions such as:

- Which patient data are pathological? How do the parameters influence each other?
- Which therapies may be envisaged with a given patient data set?
- Which are the risks and success rates of the therapies for a particular patient?

Answers to these questions allow to better understand the disease mechanisms.

The most straight forward approach for integrating the models is to identify output from one model and exploit it for calculation of the other model. To realize this, it is essential to base upon the same semantics and ontologies (see requirement 2). On the basis of the integrated models, a clinical decision support system can be developed. The system can provide answers to the questions mentioned. Further, clinicians can be assisted in diagnosing and categorizing a disease or even in the choice of an optimal therapy. The integration of the three model types also supports in controlling the success of the therapy for a given patient: For example, questions can be answered such as which is the success of each therapy given the individual patient condition? In order to suggest therapies, methods are required for reasoning and inferencing. Integration of statistical methods into the system will allow for epidemiological studies or analysis of treatment outcome and thus goes beyond clinical decision support by supporting also research and quality management.

5 Discussion and conclusion

The main benefit of the three-model based approach is that functionalities of current information or clinical decision support systems are extended by:

1. bridging the gap between system biological models and patient and clinical data (patient model) as well as medical knowledge data (decision model) – in order to make justified, well-grounded therapeutic decisions, and by
2. improved semantic annotation and linking of the heterogeneous data – in order to support users with an integrated view on patient data and on knowledge about disease development and treatment.

There are still many research problems to address to realize the concept completely, for example: (1) How can we build semi-automatically a digital patient model, including biological models and therapy decision models? and (2) How can the models be linked and interpreted together in order to provide a decision support system? We are currently focusing on developing tools to support expert-based modelling of decision processes.

**Funding:** This research is part of the Digital Patient Modelling group partially funded by the BMBF.
Author's Statement
Conflict of interest: Authors state no conflict of interest.

Material and Methods: Informed consent: Informed consent has been obtained from all individuals included in this study. Ethical approval: The research related to human use has been complied with all the relevant national regulations, institutional policies and in accordance the tenets of the Helsinki Declaration, and has been approved by the authors’ institutional review board or equivalent committee.

References