PhD Thesis Title: "From Data to Decision. A Knowledge Engineering approach to individualize cancer therapy"

Author: Erik (Hendrik A.) Roelofs
Email: erik.roelofs@maastro.nl
Institution: MAASTRO CLINIC, GROW dept., Maastricht University, the Netherlands
Supervisors: Prof. Philippe Lambin, MD & Prof. Dr. André Dekker

Graduation Date: 2016-07-08
Available on line: http://doi.org/10.17195/candat.2016.07.1

ABSTRACT:

Radiation oncology is in trouble. In fact, healthcare in general is. Healthcare cost is rapidly rising, leading to a non-sustainable situation. An aging population can partly explain the increasing costs, but overtreatment plays a large role as well. This particularly happens due to defensive healthcare. On the other hand, overtreatment also occurs due to the longevity of clinical protocols, which can cause inefficiency, as does the technology-push from healthcare industry.

One of the options to reduce the costs is to apply cost-effectiveness analyses or Healthcare Technology Assessments to inform policy and decision makers about the rightful use of (new) technology. To this end, the development of personalized decision support systems was investigated, including cost-effectiveness in the field of proton therapy. The following hypotheses were addressed and affirmed:

- Rapid Learning Healthcare (RLHC) can improve clinical research and decision-making.
- Advances in Healthcare Information Technology (HIT) enable international data sharing while preserving patient privacy.
- Combining *in silico* planning comparison studies with multifactorial prediction models enables cost-effective application of limited, expensive resources, as introduced for proton therapy.

The basic concept of this thesis is the use of RLHC to complement evidence-based medicine with prediction modelling to provide clinical decision support for individualised cancer care (*Chapter 2*). RLHC applies advanced HIT to aggregate disparate clinical information sources into a data warehouse that presents a complete, integrated disease-oriented view of medical data for research purposes. This chapter presents a general overview of some of the techniques that were used in the following work.

Chapter 3 demonstrates how institutions can benefit from this improved presentation of patient information to improve efficiency and quality of data collection for clinical trials. Furthermore, RLHC uses machine learning principles to mine the data for clinically relevant relations to build data-driven complication prediction models.

Large amounts of data are needed for the modelling and validation of these models, which is unavailable from single institutions. Furthermore, to increase information heterogeneity, which improves the predictive value of models, and to allow for external validation, it is imperative that foreign data is available.

Chapter 4 presents the necessary data-sharing techniques that unlock the medical data silos in a privacy-preserving manner by installing 'connectors' inside the hospital's network firewalls. Furthermore, by transforming the data into internationally standardised semantic interoperable data models, computer systems are able to translate data into knowledge. Furthermore, medical

records in local languages are converted into international, unique codes. This means that the data can be reused in foreign countries with equal meaning and value, which allows for federated or distributed learning of complication prediction models.

In advance of a federated network, which is very advanced but not common practice yet, the framework of MISTIR (www.mistir.info) uses a centralised research database for international collaborative *in silico* clinical trials (*Chapter 5*). The system offers a secure data-exchange platform with high-quality datasets and trial protocols for multiple *in silico* planning comparisons of different treatment modalities. Quality assurance measures and automated data extraction procedures are applied for uniform analysis of the results.

Chapter 6 presents a ROCOCO (www.mistir.info/rococo) lung cancer trial that used the platform to investigate whether proton therapy can reduce dose to normal tissue, with equal or higher tumour dose. The trial participants downloaded de-identified clinical datasets to plan according to a strict protocol. The treatment planning results were returned to the server after which central analysis was performed. It showed that dose reduction is indeed possible, even when escalating the dose to the tumour.

Chapter 7 evaluates an online proton therapy decision support system (PRODECIS: www.prodecis.nl), reusing head and neck cancer data from another ROCOCO trial. The system assesses healthcare reimbursement eligibility for model-based tumour groups by comparison of photon and proton treatment plans, using publicly available toxicity and cost-prediction models from PredictCancer (www.predictcancer.org). The system proved successful in the assessment of 92% of the test cases and showed that, with proton therapy, 91% of the patients clinically benefit from reduced complications after one year, while 35% would be considered cost-effective.

References to author publications that relate specifically to the dissertation:

Chapter 2: Lambin P, **Roelofs E**, Reymen B, Velazquez ER, Buijsen J, *et al.*, 'Rapid Learning health care in oncology' - An approach towards decision support systems enabling customised radiotherapy.' Radiotherapy & Oncology. 2013 Oct;109[1]:159–64. doi:10.1016/j.radonc.2013.07.007

Chapter 3: **Roelofs E**, Dekker A, Meldolesi E, van Stiphout RGPM, Valentini V, and Lambin P., 'International data-sharing for radiotherapy research: An open-source based infrastructure for multicentric clinical data mining.' Radiotherapy & Oncology. 2014 Feb;110[2]:370–4. doi:10.1016/j.radonc.2013.11.001

Chapter 4: **Roelofs E**, Persoon L, Nijsten S, Wiessler W, Dekker A, and Lambin P, 'Benefits of a clinical data warehouse with data mining tools to collect data for a radiotherapy trial.' Radiotherapy & Oncology. 2013 Jul;108[1]:174–9. doi:10.1016/j.radonc.2012.09.019

Chapter 5: **Roelofs E**, Persoon L, Qamhiyeh S, Verhaegen F, De Ruysscher D, *et al.*, 'Design of and technical challenges involved in a framework for multicentric radiotherapy treatment planning studies.' Radiotherapy & Oncology. 2010 Dec;97[3]:567–71. doi:10.1016/j.radonc.2010.08.009

Chapter 6: **Roelofs E**, Engelsman M, Rasch C, Persoon L, Qamhiyeh S, *et al.*, "Results of a multicentric in silico clinical trial (ROCOCO): Comparing radiotherapy with photons and protons for non-small cell lung cancer.' Journal of Thoracic Oncology. 2012 Jan;7[1]:165–76. doi:10.1097/JTO.0b013e31823529fc

Chapter 7: Cheng Q, **Roelofs E**, Ramaekers BLT, Eekers D, van Soest J, *et al.*, 'Development and evaluation of an online three-level proton vs photon decision support prototype for head and neck cancer – Comparison of dose, toxicity and cost-effectiveness.' Radiotherapy & Oncology. 2016 Feb;118[2]:281–5. doi:10.1016/j.radonc.2015.12.029