

Learning from routinely produced clinical data and Big Data technology in Radiation Oncology

Citation for published version (APA):

Lustberg, T. (2018). *Learning from routinely produced clinical data and Big Data technology in Radiation Oncology*. [Doctoral Thesis, Maastricht University]. Datawyse / Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.20181206tl>

Document status and date:

Published: 01/01/2018

DOI:

[10.26481/dis.20181206tl](https://doi.org/10.26481/dis.20181206tl)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Learning from routinely produced clinical data and Big Data technology in Radiation Oncology

Tim Lustberg

© copyright Tim Lustberg, Maastricht 2018

Printing: Datawyse | Universitaire Pers Maastricht

ISBN 978-94-6380-069-3



The research presented in this thesis was conducted at GROW-School for Oncology and Developmental Biology, Department FHML, of Maastricht University.

Learning from routinely produced clinical data and Big Data technology in Radiation Oncology

Dissertation

to obtain the degree of Doctor at the Maastrich University,
on the authority of the Rector Magnificus Prof.dr. Rianne M. Letschert
in accordance with the decision of the Board of Deans,
to be defended in public on Thursday 6th of December 2018 at 10.00 AM

by

Tim Lustberg

Supervisors

Prof. dr. ir. A.L.A.J. Dekker

Prof. dr. ir. F.J.W. Verhaegen

Assessment Committee

Prof. dr. D. De Ruyscher (Chair)

Prof. dr.ir. M.J. Dumontier

Dr. ir. S. Klein, Erasmus MC Rotterdam

Dr. ir. R. Monshouwer, Radboudumc Nijmegen

Hartelijk dank aan de mensen die geholpen hebben met mijn
promotieonderzoek

A special thanks to all the people involved in completing my
doctoral research

CONTENTS

| | |
|---|-----|
| Chapter 1 | |
| General introduction and outline of the thesis | 9 |
| Chapter 2 | |
| Big Data in Radiation Therapy: challenges and opportunities | 21 |
| Chapter 3 | |
| Towards a semantic PACS: Using Semantic Web technology to represent imaging data | 29 |
| Chapter 4 | |
| Radiation oncology terminology linker: A step towards a linked data knowledge base | 37 |
| Chapter 5 | |
| Set-up verification and 2D EPID dosimetry during breath hold compared to free breathing in breast cancer radiotherapy | 47 |
| Chapter 6 | |
| Implementation of a rapid learning platform: Predicting 2-year survival in laryngeal carcinoma patients in a clinical setting | 63 |
| Chapter 7 | |
| Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer | 81 |
| Chapter 8 | |
| Discussion and future prospective | 97 |
| Summary | 107 |
| Samenvatting | 111 |
| Valorisation Addendum | 115 |
| Curriculum Vitae | 119 |
| List of publications | 121 |

Chapter 1

General introduction and outline of the thesis

Tim Lustberg

In 2017, the International Telecommunications Union reported half of the world population has a mobile-broadband subscription, where in the developed world nearly everyone is connected to the internet through their smartphone [1]. Smartphones are playing such an important role in our lives that the younger generations cannot even imagine living without them. Getting anywhere with a paper map instead of using the GPS on your smartphone with google maps seems medieval. Because we are constantly connected to the internet, data is being generated as part of our daily lives: Put your smartphone under your pillow and record you sleep patterns; Go for a run with your phone and record pace, route or even your heart rhythm; Check-in at your favorite restaurant on social media during a diner with your friends. The list of examples is endless. Companies, usually asking for your consent when installing the app, use this data in many different ways [2]. For instance, by processing your search history, the products you bought online and the places you visited, Google can learn your preferences and predict which news items or advertisements you may appreciate and present these to you [3,4]. This scope of data collection and processing is referred to as Big Data [5]. A review of the large variety of Big Data definitions by De Mauro et al resulting in the following consensus: “Big Data is the Information asset characterised by such a High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value” [5].

While many businesses leverage Big Data technology to rapidly grow their business, it endangers others; a shining example is the numerous video-streaming services putting the video rental stores out of business. Overall there is a worry of how the use of Big Data will affect our economic and social systems [6], e.g. to driving insurance policies using people’s recorded lifestyle data. Privacy and data protection are topics of the day [7,8], the opportunities of cybercrime are increasing. Governments are aware of these rising issues and they are creating legislation to protect people’s privacy. In Europe, the General Data Protection Regulation act [9] will be enforced strict rules when collecting, processing and sharing confidential information punishing those who do not comply with large fines that could potentially bankrupt a company. An almost cliché quote used by many politicians over the last centuries rings very true for Big Data: “With great power, comes great responsibility.”

The success of Big Data in social-economic aspect of our lives motivates the question how this could be used to innovate health care. While Big Data technology is a hot topic in research, it has yet to make a big impact in clinical practice

[10–14]. An important reason is the problems of Big Data with quantity, quality and accessibility to data. Having large amounts of structured data is vital to come to a reliable solution without wasting time trying to find answers that are not in the available data. **Chapter 2** comments on the current state of Big Data in radiation oncology and the issues involved.

Radiation oncology

Radiation oncology is one of the disciplines involved with treating cancer patients, together with surgery and chemotherapy [15,16]. To understand the data problems a general understanding of the workflow of a radiotherapy clinic is required. When a patient is diagnosed with cancer the multi-disciplinary tumor board, consisting of specialists of each respective field, will decide the best treatment (combination) for the patient. Figure 1 shows a schematic simplified overview of the radiotherapy treatment process.

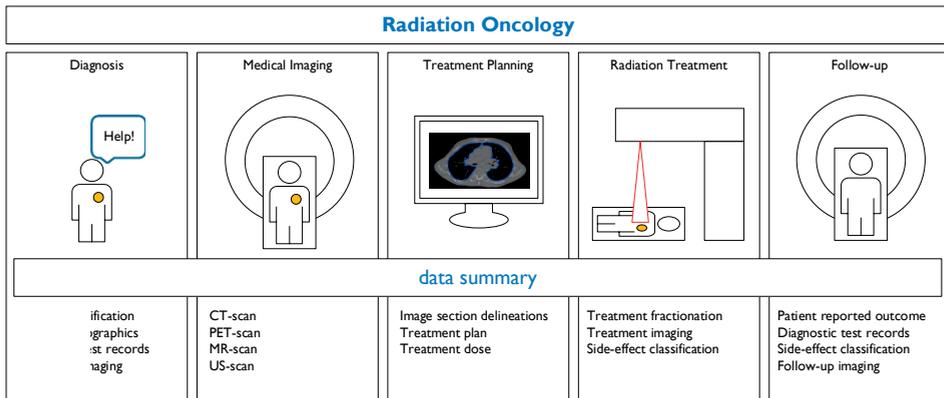


Figure 1: Radiation oncology overview with relevant data items.

Diagnosis

Part of the diagnosis process is classifying the disease. Terminologies are available to perform the classification [17–19]. However, a worldwide adoption of one of these terminologies has not been achieved [20]. A review by Stallinga et al. concluded that clinicians do not experience the recording of patient information as (native) ambiguous language as a problem. They suggest this to be the cause of low adoption of terminologies in clinical practice [21]. At the point of diagnosis, information about the general health of the patient is also gathered to estimate

how the negative side effects of radiotherapy will affect the patient. An example is the Forced Expiratory Volume measurement for lung cancer patients. During this test the amount of air that can be exhaled in one breath is measured, which is considered to be an indication of the general health of the lungs.

Medical Imaging

To create a radiotherapy plan, an anatomical model of the patient is needed. A digital representation of the anatomy is made where the patient is placed in the exact same position as he or she would be treated. The anatomical model will provide the treatment planning system with the necessary information to simulate the radiation treatment [16]. For some cancers, additional imaging is complementary to the planning process. Small tumors can be detected by creating a PET-scan [22] and soft tissue contrasts (e.g. in the brain and a brain-tumor) can be better displayed with an MRI-scan [23].

Treatment Planning

With the proper imaging as described above, a radiation technologist and oncologist delineate the organs at risk (OAR) and the tumor (Gross Tumor Volume, GTV). Margins are added around the tumor to account for sub-clinical disease spread (Clinical Target Volume, CTV) and treatment planning and delivery (Planning Target Volume, PTV) uncertainty [24]. An example of such an uncertainty is the patient placement under the treatment machine which can slightly differ from the simulation. Using these delineations, the treatment planning system is given a set of optimization rules to simulate the radiation treatment [16]. The simulation balances the radiation delivered to the tumor and healthy tissue, which results in a set of instructions for the radiotherapy treatment machine. The results are presented as Dose Volume Histograms (DVHs) [25,26]. These DVHs show the relation between the volume of the organ-at-risk or tumor against the amount of radiation dose (Figure 2). The radiation oncologist uses these graphs to determine if the prescribed radiation dose covers enough of the tumor volume (e.g. 95%) while the radiation dose to organs at risk stays below certain dose per volume thresholds (e.g. not more than 20% of the lungs can receive 25 Gy).

Radiation Treatment

The foundation of radiation treatment is based on the biology behind damaged cell repair, the healthy damaged cells recover partially while a larger fraction of

the deceased cells cannot. To recover, the healthy cells need time, this is why radiation treatment is divided into fractions [27]. Before starting the treatment, the patient position needs to be the same as during the treatment planning CT-scan. To verify this position the Electronic Portal Image Device (EPID) and/or Cone-Beam CT are used to create images of the patient [28]. The EPID is a high-energy photon imaging panel that can be used to measure the amount of radiation behind the patient. While it was originally intended as a setup verification device, it can be used to reconstruct the treatment dose delivered to the patient in [29–31].

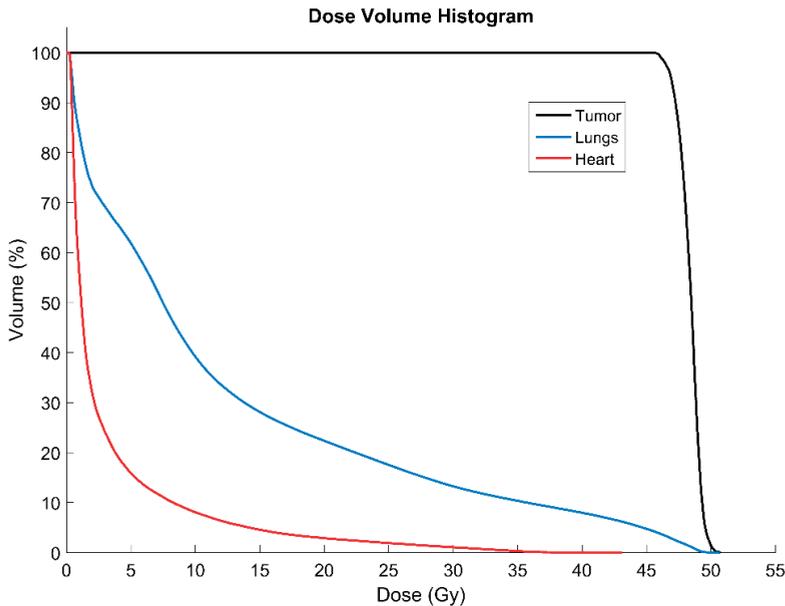


Figure 2: Example Dose Volume Histogram for the tumor, lungs and heart.

Follow-up

During and after completing the treatment, it is important to monitor the effect of the radiation treatment on the tumor and general health of the patient. As an outcome measure, the measurements described for the diagnosis of the patients are often repeated to compare against. Another important measurement is patient reported outcomes, questionnaires filled in by the patient (without the physician) about their health and quality of life after treatment [32,33].

Radiation oncology is highly dependent on the medical images as described above. Medical images are stored and transported using the Digital Imaging and Communications in Medicine (DICOM) standard. This basic external beam

radiation dataset is represented by the DICOM modalities CT, RTSTRUCT, RTPLAN and RTDOSE. The rest of the information generated by the treatment planning system and treatment machines is stored in an Electronic Medical Record (EMR). The optimization rules (e.g. the treatment dose and the dose constraints to the OAR) are often standardized in a treatment protocol, which can be stored in the EMR. However, other important information can also be stored in free text (e.g. the patient reported outcome) or spreadsheets managed by the clinicians outside EMR. Getting access to and understanding all these data is the start of most clinical research in radiation oncology.

Linked Data in radiation oncology

The DICOM objects created for radiation therapy, or information derived from them, are valuable to research. To apply the Big Data technologies to the data described in the previous paragraph this data needs to be a machine-readable and accessible format. However, getting access to the (meta-)data of the DICOM objects is often very restrictive. In most cases the only way to analyze the entire meta-data header, is to extract the DICOM file from the source system and analyze them with secondary software. In an effort to disclose the DICOM (meta-)data generated in radiation oncology, **Chapter 3** describes a software solution to tackle the problem of storing the DICOM (meta-)data in a system in such a manner that it can be linked to other data derived from the image, treatment information or patient outcomes using Linked Data technology. Linked Data is described as: “Linked Data is about using the Web to connect related data that wasn't previously linked, or using the Web to lower the barriers to linking data currently linked using other methods. More specifically, Wikipedia defines Linked Data as "a term used to describe a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF.” [34].

FAIR data in radiation oncology

A publication by Wilkinson et al. provides guidelines for data management and stewardship in research stating: “that all research objects should be Findable, Accessible, Interoperable and Reusable (FAIR) both for machines and for people” [35]. This builds upon the definition of Linked Data but does not limit itself

to Semantic Web technology because the guidelines are independent of technology implementations. Horizon 2020, the biggest research and innovation program of Europe, has adopted these guidelines as a framework to write a data management chapter for new grant proposals, acknowledging the necessity by the research community [36]. **Chapter 4** continues the work described in **Chapter 3** and describes a software platform that can analyze DICOM modalities and store them in a FAIR data layer together with the DICOM meta-data.

Applying data science to radiation oncology

After reviewing the technology available to improve health care using modern data science, these technologies were tested on clinical data to learn about the applicability of these concepts.

EPID dosimetry

As stated before, the EPID was originally intended as a position verification tool. Before starting treatment, the patient is irradiated for a short period from the beam angles defined during the treatment planning process. The simulated images can be compared to the actual position of the patient to determine if the patient was correctly set up and ready to start treatment. The method to use the EPID during the actual treatment to acquire an image and convert this image into a radiation dose is called EPID dosimetry [29–31]. EPID dosimetry is a form of quality assurance to determine if the radiation treatment is going according to plan. Some changes in the anatomy of the patient (e.g. the tumor shrinking rapidly or the lungs filling up with fluid) can greatly decrease the effectiveness of the treatment, detecting these changes early is vital for a positive outcome. **Chapter 5** is an application of Big Data image analysis applied to EPID dosimetry images created during routine breast cancer radiotherapy to provide evidence to validate a new breath hold technique by comparing the EPID dosimetry images of these patients to patients treated with a free-breathing protocol.

Decision support systems

An example of Big Data technology that can be applied to health care is a decision support system in the form of a prediction model [10,12,37]. Decision support systems (as the name implies) do not make decisions on their own but provide the professional with the information needed to support their decision. The

amount of data that a physician is confronted with is no longer interpretable by a human [38]. To support the physician in medical decisions, the machine-learning model processes the large amount of data and presents the physician with a comprehensible amount of information to support their medical decision. Combining predictions and their expertise can be used to engage the patient in shared decision-making. However, recommending the wrong item on Amazon has a far lower impact than predicting the wrong outcome of a treatment, the prediction influence decisions that can impact a person's health. The **Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD)** statement describes a machine-learning validation strategy that can measure the reliability of such a model [39]. As described by TRIDOD, it is important to perform an external validation to validate the reliability of a model. **Chapter 6** shows the value of external validation of a prediction model for the survival of laryngeal carcinoma patients using routine clinical data. It compares this validation to a second independent validation on a clinical trial dataset and elaborates on how this analysis method can complement the current state of the art in medical research.

Deep Learning image processing

Delineation of organs at risk can be a very time-consuming task for certain areas of the body. Applying Big Data technology to this task is an important subject [40–42]. Delineating organs at risk or tumors relies heavily on human interpretation of the image [43]. **Chapter 7** provides a clinical example of a Big Data learning approach and how it can be applied on medical images to delineate organs at risk for lung cancer patients. By adding a human component in the final step of the process, it demonstrates how imperfect models can still improve and support the clinic.

Finally, **Chapter 8** summarizes and discusses the work presented in this thesis.

References

1. Measuring the Information Society Report 2017. *ITU*. <http://www.itu.int:80/en/ITU-D/Statistics/Pages/publications/mis2017.aspx> [accessed January 10, 2018]
2. Your Personal Data and How Companies Use It. *DBS Backgr Checks Employ Screen Cbscreening*. <https://www.cbscreening.co.uk/news/post/your-personal-data-and-how-companies-use-it/> [accessed January 10, 2018]
3. Digital Advertising: A Hotbed for Machine Learning. *PubMatic*. <https://pubmatic.com/blog/digital-advertising-machine-learning/> [accessed January 10, 2018]
4. Perlich C, Dalessandro B, Raeder T, Stitelman O, Provost F. Machine learning for targeted display advertising: transfer learning in action. *Mach Learn* 2014;95(1):103–127. doi:10.1007/s10994-013-5375-2
5. Mauro AD, Greco M, Grimaldi M. A formal definition of Big Data based on its essential features. *Libr Rev* 2016;65(3):122–135. doi:10.1108/LR-06-2015-0061
6. Mittelstadt BD, Floridi L. The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts. *Sci Eng Ethics* 2016;22(2):303–341. doi:10.1007/s11948-015-9652-2
7. Dangers in Data Collection: How Much Is Too Much? *Insur J* 2016. <https://www.insurancejournal.com/magazines/mag-coverstory/2016/10/24/429763.htm> [accessed March 16, 2018]
8. Posted by Vincent Granville on September 14 2016 at 9:30am, Blog V. Why is Big Data so Dangerous? <https://www.datasciencecentral.com/profiles/blogs/why-is-big-data-so-dangerous> [accessed March 16, 2018]
9. Home Page of EU GDPR. *EU GDPR Portal*. <http://eugdpr.org/eugdpr.org.html> [accessed March 16, 2018]
10. Kohn MS, Sun J, Knoop S, Shabo A, Carmeli B, Sow D, et al. IBM's Health Analytics and Clinical Decision Support. *Yearb Med Inform* 2014;9(1):154–162. doi:10.15265/IY-2014-0002
11. Abernethy AP, Etheredge LM, Ganz PA, Wallace P, German RR, Neti C, et al. Rapid-learning system for cancer care. *J Clin Oncol Off J Am Soc Clin Oncol* 2010;28(27):4268–4274. doi:10.1200/JCO.2010.28.5478
12. Lambin P, Roelofs E, Reymen B, Velazquez ER, Buijssen J, Zegers CML, et al. 'Rapid Learning health care in oncology' – An approach towards decision support systems enabling customised radiotherapy'. *Radiother Oncol* 2013;109(1):159–164. doi:10.1016/j.radonc.2013.07.007
13. Shah A, Stewart AK, Kolacevski A, Michels D, Miller R. Building a Rapid Learning Health Care System for Oncology: Why CancerLinQ Collects Identifiable Health Information to Achieve Its Vision. *J Clin Oncol Off J Am Soc Clin Oncol* 2016;34(7):756–763. doi:10.1200/JCO.2015.65.0598
14. Benedict SH, Hoffman K, Martel MK, Abernethy AP, Asher AL, Capala J, et al. Overview of the American Society for Radiation Oncology–National Institutes of Health–American Association of Physicists in Medicine Workshop 2015: Exploring Opportunities for Radiation Oncology in the Era of Big Data. *Int J Radiat Oncol Biol Phys* 2016;95(3):873–879. doi:10.1016/j.ijrobp.2016.03.006
15. Delaney G, Jacob S, Featherstone C, Barton M. The role of radiotherapy in cancer treatment. *Cancer* 2005;104(6):1129–1137. doi:10.1002/cncr.21324
16. Baskar R, Lee KA, Yeo R, Yeoh KW. Cancer and Radiation Therapy: Current Advances and Future Directions. *Int J Med Sci* 2012;9(3):193–199. doi:10.7150/ijms.3635
17. Evaluation of the Content Coverage of SNOMED CT: Ability of SNOMED Clinical Terms to Represent Clinical Problem Lists - ScienceDirect. <http://www.sciencedirect.com.ezproxy.ub.unimaas.nl/science/article/pii/S002561961161728X> [accessed January 16, 2018]

18. Bowman SE. Why ICD-10 Is Worth the Trouble. *J AHIMA* 2008;79(3):24–29
19. Sioutos N, Coronado S de, Haber MW, Hartel FW, Shaiu WL, Wright LW. NCI Thesaurus: A semantic model integrating cancer-related clinical and molecular information. *J Biomed Inform* 2007;40(1):30–43. doi:10.1016/j.jbi.2006.02.013
20. WHO | International Classification of Diseases (ICD) Information Sheet. WHO. <http://www.who.int/classifications/icd/factsheet/en/> [accessed January 16, 2018]
21. Stallinga HA, ten Napel H, Jansen GJ, Geertzen JHB, de Vries Robbé PF, Roodbol PF. Does language ambiguity in clinical practice justify the introduction of standard terminology? An integrative review. *J Clin Nurs* 2015;24(3–4):344–352. doi:10.1111/jocn.12624
22. Radiotherapy treatment planning for patients with non-small cell lung cancer using positron emission tomography (PET) - ScienceDirect. <http://www.sciencedirect.com.ezproxy.uv.unimaas.nl/science/article/pii/S0167814001004704> [accessed January 23, 2018]
23. Khoo VS, Dearnaley DP, Finnigan DJ, Padhani A, Tanner SF, Leach MO. Magnetic resonance imaging (MRI): considerations and applications in radiotherapy treatment planning. *Radiother Oncol* 1997;42(1):1–15. doi:10.1016/S0167-8140(96)01866-X
24. Defining the tumour and target volumes for radiotherapy. <https://www.ncbi.nlm.nih.gov.ezproxy.uv.unimaas.nl/pmc/articles/PMC1434601/> [accessed March 16, 2018]
25. Nelms B, Stambaugh C, Hunt D, Tonner B, Zhang G, Feygelman V. Methods, software and datasets to verify DVH calculations against analytical values: Twenty years late(r). *Med Phys* 2015;42(8):4435–4448. doi:10.1118/1.4923175
26. Drzymala RE, Mohan R, Brewster L, Chu J, Goitein M, Harms W, *et al.* Dose-volume histograms. *Int J Radiat Oncol* 1991;21(1):71–78. doi:10.1016/0360-3016(91)90168-4
27. Yang Y, Xing L. Optimization of radiotherapy dose-time fractionation with consideration of tumor specific biology. *Med Phys* 2005;32(12):3666–3677. doi:10.1118/1.2126167
28. Li H, Zhu XR, Zhang L, Dong L, Tung S, Ahamad A, *et al.* Comparison of 2D Radiographic Images and 3D Cone Beam Computed Tomography for Positioning Head-and-Neck Radiotherapy Patients. *Int J Radiat Oncol* 2008;71(3):916–925. doi:10.1016/j.ijrobp.2008.01.008
29. McDermott LN, Wendling M, Sonke JJ, van Herk M, Mijnheer BJ. Replacing Pretreatment Verification With In Vivo EPID Dosimetry for Prostate IMRT. *Int J Radiat Oncol* 2007;67(5):1568–1577. doi:10.1016/j.ijrobp.2006.11.047
30. van Elmpt W, Nijsten S, Petit S, Mijnheer B, Lambin P, Dekker A. 3D In Vivo Dosimetry Using Megavoltage Cone-Beam CT and EPID Dosimetry. *Int J Radiat Oncol* 2009;73(5):1580–1587. doi:10.1016/j.ijrobp.2008.11.051
31. Time dependent pre-treatment EPID dosimetry for standard and FFF VMAT - IOPscience. <http://iopscience.iop.org/article/10.1088/0031-9155/59/16/4749/meta> [accessed January 24, 2018]
32. Basch E, Snyder C, McNiff K, Brown R, Maddux S, Smith ML, *et al.* Patient-Reported Outcome Performance Measures in Oncology. *J Oncol Pract* 2014;10(3):209–211. doi:10.1200/JOP.2014.001423
33. Basch E, Iasonos A, McDonough T, Barz A, Culkin A, Kris MG, *et al.* Patient versus clinician symptom reporting using the National Cancer Institute Common Terminology Criteria for Adverse Events: results of a questionnaire-based study. *Lancet Oncol* 2006;7(11):903–909. doi:10.1016/S1470-2045(06)70910-X
34. Amit S. Linked Data, The Story So Far. *Semantic Services, Interoperability and Web Applications: Emerging Concepts*, IGI Global; 2011
35. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 2016;3:160018. doi:10.1038/sdata.2016.18

36. Guidelines on FAIR Data Management in Horizon 2020 | Agricultural Information Management Standards (AIMS). <http://aims.fao.org/activity/blog/guidelines-fair-data-management-horizon-2020> [accessed January 10, 2018]
37. Lambin P, Zindler J, Vanneste BGL, De Voorde LV, Eekers D, Compter I, *et al.* Decision support systems for personalized and participative radiation oncology. *Adv Drug Deliv Rev* 2017;109:131–153. doi:10.1016/j.addr.2016.01.006
38. Larichev OI, Moshkovich HM. Limits to decision-making ability in direct multiattribute alternative evaluation. *Organ Behav Hum Decis Process* 1988;42(2):217–233. doi:10.1016/0749-5978(88)90013-1
39. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Med* 2015;13(1):1. doi:10.1186/s12916-014-0241-z
40. van Baardwijk A, Bosmans G, Boersma L, Buijsen J, Wanders S, Hochstenbag M, *et al.* PET-CT-Based Auto-Contouring in Non-Small-Cell Lung Cancer Correlates With Pathology and Reduces Interobserver Variability in the Delineation of the Primary Tumor and Involved Nodal Volumes. *Int J Radiat Oncol* 2007;68(3):771–778. doi:10.1016/j.ijrobp.2006.12.067
41. Hu P, Wu F, Peng J, Liang P, Kong D. Automatic 3D liver segmentation based on deep learning and globally optimized surface evolution. *Phys Med Biol* 2016;61(24):8676–8698. doi:10.1088/1361-6560/61/24/8676
42. Mansoor A, Cerrolaza JJ, Perez G, Biggs E, Nino G, Linguraru MG. Marginal Shape Deep Learning: Applications to Pediatric Lung Field Segmentation. *Proc SPIE-- Int Soc Opt Eng* 2017;10133. doi:10.1117/12.2254412
43. Sharp G, Fritscher KD, Pekar V, Peroni M, Shusharina N, Veeraraghavan H, *et al.* Vision 20/20: Perspectives on automated image segmentation for radiotherapy. *Med Phys* 2014;41(5). doi:10.1118/1.4871620

Chapter 2

Big Data in Radiation Therapy: challenges and opportunities

Tim Lustberg, Johan van Soest, Arthur Jochems, Timo Deist, Yvonka van Wijk,
Sean Walsh, Philippe Lambin, Andre Dekker

Br J Radiol. January 2017; 90(1069): 20160689.

Abstract

Data collected and generated by radiation oncology can be classified by the 4Vs of Big Data (Volume, Variety, Velocity, and Veracity) because it is spread across different care providers and not easily shared due to patient privacy protection. The magnitude of the 4Vs is substantial in oncology, especially due to imaging modalities and unclear data definitions. To create useful models ideally all data of all care providers is understood and learned from, however this presents challenges in the guise of poor data quality, patient privacy concerns, geographical spread, interoperability, and the large volume.

In radiation oncology there are many efforts to collect data for research and innovation purposes. Clinical trials are the gold standard when proving any hypothesis that directly affects the patient. Collecting data in registries with strict predefined rules is also a common approach to find answers. A third approach is to develop data stores that can be used by modern machine learning techniques to provide new insights or answer hypotheses.

We believe, all three approaches have their strengths and weaknesses but they should all strive to create FAIR (Findable, Accessible, Interoperable, Reusable) data. To learn from this data, we need distributed learning techniques, sending machine learning algorithms to FAIR data stores around the world learning from trial data, registries and routine clinical data rather than trying to centralize all data. To improve and personalize medicine, rapid learning platforms must be able to process FAIR 'Big Data' to evaluate current clinical practice and to guide further innovation.

Primarily because of the ubiquity of imaging in oncology, as well as many other diagnostic and therapeutic procedures, cancer data is firmly in the realm of 'Big Data' [1]. To make an estimate, in the past 10 years approximately 140 million patients were diagnosed with cancer in about 100 thousand hospitals globally. If one assumes a data volume (depending on the hospital) of 0.1-10Gb of data per patient, the total volume of cancer patients' data in the world is estimated to be 14-1400 Petabyte of data. Specifically, data in radiation oncology can be classified as 'Big Data' because: the use of data intensive imaging modalities (Volume), the imaging archives are growing rapidly (Velocity), there is an increasing amount of imaging and diagnostic modalities available (Variety), and interpretation and quality differs between care providers (Veracity). With this deluge in data it becomes increasingly hard to translate all this data into knowledge and subsequently leverage this knowledge to guide clinical decisions [2]. The radiation oncologist is overwhelmed with scientific literature, swiftly evolving treatment techniques and the exponentially increasing amount of clinical data [2]. To provide high quality individualized treatments, radiation oncologists need help translating all this data into knowledge that supports decision making in routine clinical practice [3]. Collecting this data provides its own set of challenges. The data is spread over care providers around the world, difficult to share while protecting patient privacy, non-interoperable and varying in quality.

The gold standard to assess the utility of innovations that directly affect patients is clinical trials. However, clinical trials only provide information about a select patient population which often represents only a small percentage of the actual population. Also, clinical trials provide the radiation oncologist with little information when making clinical decisions for someone who does not (exactly) fit the trial population due to age, comorbidities etc. On the other hand, clinical trials do provide high quality reusable data due to the clear definitions that are provided by trial protocols. Initiatives such as IBM Watson attempt to simplify accessing knowledge garnered from scientific literature for physicians. Patient characteristics provided by the physician are used to find and retrieve relevant publications (and possible other sources) which can aid the physician in making precision decisions for that particular patient [4].

To fill the gap of evidence between clinical trials and the common patient (i.e., one who does not fit trial inclusion criteria), data registries are being created around the world [5]. In general, the goal is to register a selected set of parameters for all patients treated for a certain cancer. This results in a large patient population with high quality data [5]. However, this requires great effort of care

providers to collect this data, which limits the number of elements recorded as someone has to fill in a form, digital or paper, to provide the registries with the data specified in the registry protocol. There are some early initiatives [6] to automatically provide the registries with the data they require by data mining the Oncology Information Systems (OIS). In theory, this should work well for all structured data (e.g. the fractionation schema or age) but is challenging for data which is usually recorded in free text (e.g. smoking behavior or comorbidities). Registries in general give insights in practice but are not designed to guide decisions for individual patients. ASCO's CancerLinQ is the exception, it aims to create a "super" registry with a learning approach on routine health care data in medical oncology [7]. Cancer screening shows promising advancements in identifying high risk patients using data mining techniques [8], this particular example shows the power of centralizing data.

A different approach is to use routine clinical data from around the world to transform data into knowledge. As a proof of concept, the euroCAT project created data stores at several cancer centers which can be accessed using web technologies. The local data is mined, pseudo-anonymized, translated, mapped to standard concepts, and made available to trusted partners in the network. The trusted partners do not have direct access to the data, but they can send a machine learning algorithm to the different data stores to learn from the data without sharing it (i.e., knowledge sharing, not data sharing). To demonstrate the power of this technique an existing model was improved by combining the data of five centers (www.eurocat.info). However, a lot of time and effort is still required to access and utilize all data generated in clinical routine and translate this into knowledge. The end result of distributed learning are prediction models which can support physicians when making patient specific choices (www.predictcancer.org). A different successful data mining was started by Public Health England. Data from all linear accelerators in England were collected automatically using the Radiotherapy Dataset (RTDS) tools (http://www.ncin.org.uk/collecting_and_using_data/rtds). This dataset was analyzed to examine the variation in given treatments for different regions of the country.

An important topic when discussing collection of healthcare data is patient privacy. All three approaches handle privacy issues differently. Registries are usually hosted in some central location by professional societies or government-related entities which are often authorized to collect identifiable patient data and securely store it while giving researchers an anonymized view on this data. Clinical trials work with informed consent and with pseudo-anonymized data.

Distributed systems are privacy-by-design systems as they simply do not allow data to leave the site where it was collected.

It should be noted, that there is a perception among healthcare providers that data must be kept in isolation due to privacy issues. The fact is that there are existing solutions for these issues. The real barrier to learning from (Big) data in healthcare is that it requires willingness, resources, and expertise [9].

Healthcare data is not yet “Big” enough to apply purely data driven machine learning approaches and clinical expertise is needed to create useful models that make sense to the clinicians. Clinical trials, clinical registries and routine clinical data all provide unique evidence which is currently utilized separately. Combining the three evidence sources into interoperable data stores makes them complementary to each other and will enable healthcare to move forward. However, data quality (4Vs) and sharing issues are hindering progress. To achieve ‘Big Data’ in healthcare the data has to be Findable, Accessible, Interoperable and Re-usable (FAIR). The FAIR Guiding Principles [10] can be applied to achieve good data management and stewardship which will enable knowledge discovery and innovation. Eventually, when data driven machine learning approaches have matured it will provide a large knowledge base and clinical trials will only be used for a small subset of studies that requires to specific setup of a trial to prove (i.e., a new experimental treatment).

IBM have stated that there are numerous ways to improve healthcare using their technology and they provide a conclusion of utmost importance: “*Information technology cannot drive change*” [4]. ‘Big Data’ can be a powerful tool to move healthcare forward, but healthcare providers need to invest resources to make this happen. Consequently, industry leaders in radiation oncology are already exploring this horizon market in anticipation of the opportunities and challenges that ‘Big Data’ in healthcare represents, both in terms of efficiency and efficacy. Our experience is that the technical limitations of sharing data are minimal. Practical reasons are that healthcare providers are not willing, don’t have the resources and/or knowledge to share their data. Sharing data can have an effect on the reputation of the healthcare provider because it allows their performance to be compared to others. Furthermore, this data can be used in research that a competing institute is working on as well, possibly creating unwanted competitors. By limiting the access to data to a machine and only sharing the model learned from the data, these issues are eliminated or largely negated. Despite the conflicting interests of healthcare providers, change may be driven by pressure from external institutions (such as government and health insurance

companies) to ensure that the highest standard and most cost effective care is delivered to the patient.

Many world leaders throughout history have referenced the 17th century poem by John Donne – *“No man is an island”*, when illustrating the need for collective responsibility and action towards a brighter future for all. This maxim rings as true in healthcare as it does in all other areas of life. We believe that utilizing patient privacy preserving distributed machine learning to translate and combine all data sources into knowledge will enable healthcare to move to individualized, high quality, affordable and safe cancer treatments. Ensuring the sustainability of healthcare. This will also allow to move further towards participative medicine with customized patient decision aids.

References

1. Zarrouk M. Delivering Excellence in Patient Care with Ready Access to Clinical Data
2. Oberije C, Nalbantov G, Dekker A, Boersma L, Borger J, Reymen B, *et al.* A prospective study comparing the predictions of doctors versus models for treatment outcome of lung cancer patients: a step towards individualized care and shared decision making. *Radiother Oncol* 2014;112(1):37–43. doi:10.1016/j.radonc.2014.04.012
3. Benedict SH, Hoffman K, Martel MK, Abernethy AP, Asher AL, Capala J, *et al.* Overview of the American Society for Radiation Oncology–National Institutes of Health–American Association of Physicists in Medicine Workshop 2015: Exploring Opportunities for Radiation Oncology in the Era of Big Data. *Int J Radiat Oncol Biol Phys* 2016;95(3):873–879. doi:10.1016/j.ijrobp.2016.03.006
4. Kohn MS, Sun J, Knoop S, Shabo A, Carmeli B, Sow D, *et al.* IBM's Health Analytics and Clinical Decision Support. *Yearb Med Inform* 2014;9(1):154–162. doi:10.15265/IY-2014-0002
5. Bilimoria KY, Stewart AK, Winchester DP, Ko CY. The National Cancer Data Base: A Powerful Initiative to Improve Cancer Care in the United States. *Ann Surg Oncol* 2008;15(3):683–690. doi:10.1245/s10434-007-9747-3
6. Efstathiou JA, Nassif DS, McNutt TR, Bogardus CB, Bosch W, Carlin J, *et al.* Practice-based evidence to evidence-based practice: building the National Radiation Oncology Registry. *J Oncol Pract* 2013;9(3):e90-95. doi:10.1200/JOP.2013.001003
7. Shah A, Stewart AK, Kolacevski A, Michels D, Miller R. Building a Rapid Learning Health Care System for Oncology: Why CancerLinQ Collects Identifiable Health Information to Achieve Its Vision. *J Clin Oncol* 2016;34(7):756–763. doi:10.1200/JCO.2015.65.0598
8. Liao LJ, Chou HL, Lo WC, Wang CT, Chou HW, Chen CD, *et al.* Initial outcomes of an integrated outpatient-based screening program for oral cancers. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2015;119(1):101–106. doi:10.1016/j.oooo.2014.09.020
9. Sullivan R, Peppercorn J, Sikora K, Zalberg J, Meropol NJ, Amir E, *et al.* Delivering affordable cancer care in high-income countries. *Lancet Oncol* 2011;12(10):933–980. doi:10.1016/S1470-2045(11)70141-3
10. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 2016;3:160018. doi:10.1038/sdata.2016.18

Chapter 3

Towards a semantic PACS: Using Semantic Web technology to represent imaging data

Johan van Soest, Tim Lustberg, Detlef Grittner, M. Scott Marschall, Lucas Persoon, Bas Nijsten, Peter Feltens, Andre Dekker

Stud Health Technol Inform. 2014;205:166-70.

Abstract

The DICOM standard is ubiquitous within medicine. However, improved DICOM semantics would significantly enhance search operations. Furthermore, databases of current PACS systems are not flexible enough for the demands within image analysis research. In this paper, we investigated if we can use Semantic Web technology, to store and represent metadata of DICOM image files, as well as linking additional computational results to image metadata. Therefore, we developed a proof of concept containing two applications: one to store commonly used DICOM metadata in an RDF repository, and one to calculate imaging biomarkers based on DICOM images, and store the biomarker values in an RDF repository. This enabled us to search for all patients with a gross tumor volume calculated to be larger than 50 cc. We have shown that we can successfully store the DICOM metadata in an RDF repository and are refining our proof of concept with regards to volume naming, value representation, and the applications themselves.

Keywords: DICOM, Semantic Web, knowledge representation, semantic interoperability, SPARQL, Imaging

Introduction

The DICOM standard [1] is ubiquitous within medicine, especially in diagnostic-related departments (e.g. radiology, radiation therapy or nuclear medicine). The standard fulfils a twofold function, describing the image (meta)data using a key-value pair system (often called DICOM tags), and also as a communication protocol for transmitting this kind of data.

Over the years, the DICOM standard has added support for imaging modalities and steadily more metadata. One of these metadata features is the referencing structure based on unique identifiers (UIDs) to indicate relations between modalities, for instance, to state that a PET-CT image contains two separate image series that were scanned at the same time. Within radiotherapy, the DICOM standard is extended to record the tumor delineations of a physician (an RTSTRUCT), or to define the actual radiation treatment plan (RTPLAN) including dosimetry information (RTDOSE).

The relationships between RTSTRUCT, RTDOSE, and RTPLAN are unidirectional: RTSTRUCT refers to CT/PET/MRI, RTPLAN refers to RTSTRUCT, and RTDOSE refers to RTPLAN. This directionality leads to difficulties when traversing the DICOM modalities, especially in the opposite direction. For example, if a user has a CT scan, and wants to retrieve the radiation treatment plan (RTPLAN), he has to search for the RTSTRUCT object based on the specific CT scan, and from there search for the RTPLAN object based on the RTSTRUCT object. This is an inefficient operation because all RTSTRUCT files, and all RTPLAN files for the patient need to be processed to find the correct treatment plan.

Another issue is related to imaging biomarkers. Imaging biomarkers are essential to developing prognostic and predictive models for radiotherapy [2,3]. As this field is evolving rapidly, we need a flexible solution that can be easily extended with new features in order to store biomarkers. We need to store imaging biomarkers while maintaining a link between the biomarker and the images/modalities from which the biomarker was calculated [4]. A direct link between image and biomarker can be established based on the UID of the DICOM file(s). However, additional metadata of the image should then be processed before storage, which is not common practice in a picture archiving and communication system (PACS).

To address the issues described above, we need a flexible storage solution for (additional) metadata, while still being able to define a formal structure to represent data already available in the DICOM standard. One of these flexible storage

solutions, different from traditional databases is based on the Semantic Web technology [5]. Our definition of the Semantic Web is that it uses linked data and standardized terminologies (ontologies), therefore making it possible for everyone to contribute to a web of knowledge [6]. This linked data can be stored in a Resource Description Framework (RDF) repository, which can be viewed as a single table with three columns (subject, predicate and object) and where every row is called a triple. When we unify the object of a triple with its equal in another triple's subject, we build a graph of the data. This RDF repository can be queried using a specific query language: SPARQL (SPARQL Protocol and RDF Query Language), which performs pattern matching on the available graph of data [7].

In this study, we hypothesize that we can use Semantic Web technology to store and represent metadata of DICOM image files and additional computational results linked to the image metadata.

1. Methods

To test our hypothesis, we separated the actions into three different tasks. The first task addressed the storage and representation of DICOM metadata in an RDF repository. The second task targeted the storage of additional computational results, related to the images. The third task was to implement a proof of concept.

1.1. Storage and representation of DICOM metadata

To store the data in an RDF repository (and make it accessible via a SPARQL endpoint), we first developed an ontology representing the most common DICOM metadata elements. This ontology serves as a definition of our data domain, and as a contract regarding the structure and data elements if anyone wants to contribute or query our data [8]. Our ontology can be found at the NCBO Bioportal [9] ontology browser as "Semantic DICOM Ontology" [10]. Although the ontology is still under development, the basic concepts/structure is stable, as it is based on the DICOM standard itself.

This ontology is used by our Semantic DICOM (SeDI) conversion service. This service is written in Java, using the dcm4che library [11]. For every DICOM file, the application will check whether specific tags defined in the ontology are present in the DICOM file, and creates RDF triples based on the rules defined in the ontology. As our conversion service keeps the basic ontological rules (e.g. regarding domain and range) in mind, the data representation will follow our

ontological structure. After creating RDF triples, the triples are exposed via a SPARQL endpoint.

1.2. Storage of additional computation results

To calculate additional imaging biomarkers, we developed a pipeline application, which also accepts data using the DICOM communication protocol and uses a plug-in architecture for different computational modules. This pipeline application is developed in Java using the dcm4che [11] and Spring Dependency Injection [12] libraries.

Computational modules can be developed outside the direct scope of our pipeline application as with an API for calculation modules, every type of calculation can be plugged into the pipeline (when wrapped with Java code). This enhances the flexibility to add computations when needed.

Afterwards the results can be exported to different types of storage, including an RDF repository. When using the RDF export module (which triggers a SPARQL update/insert query) the application tries to execute a pre-defined query (containing placeholders for calculated values) on a given RDF repository, and can insert/update triples. Users can create their own insert/update query and are not bound to a specific ontology, which leaves room for end-users to commit the results to the ontology of their choice.

1.3. Proof of concept: query patients and their tumor volume

To test our hypothesis, we implemented a proof of concept setup (Figure 1) where the results of two systems were stored in separate RDF repositories (Figure 1). We used 10 randomly selected patients from our PACS (with different tumor locations) which were sent to both the SeDI conversion service, and the computation pipeline. Both of the applications and triple stores resided on computers in different countries.

To structure the output of our computational pipeline, we used the Radiation Oncology Ontology (under development), which is able to extend the Semantic DICOM ontology for the radiotherapy domain. For our hypothesis test, we linked our calculation output to the identification of the delineated volume of interest (representing the gross tumor volume), stored in the RDF repository of the SeDI conversion service.

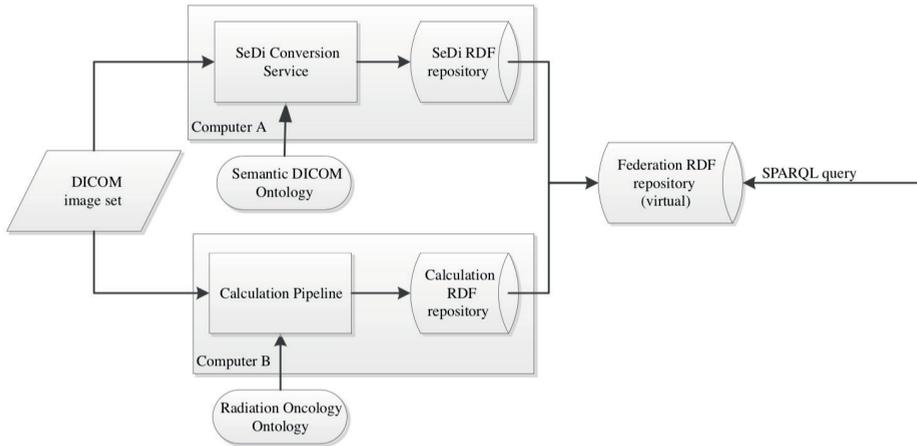


Figure 1. Setup proof of concept

2. Results

We successfully parsed the DICOM tags specified in the ontology and stored the resulting triples in our RDF repository. The created concepts were structured according to the DICOM hierarchy of patients, studies, series and images, as specified in the SeDI ontology.

For 6 out of 10 patients, our computation pipeline could calculate the gross tumor volume. 4 out of 10 patients resulted in no calculation result, because the gross tumor volume was not named correctly using our local clinical standard for volume descriptions.

When performing a query on the federated endpoint for patients with a tumor volume larger than 50 cc, 5 patients were returned. When querying for all patients and tumor volumes, the query resulted in all 6 patients, where one patient had a volume of 2 cc, and was therefore not returned in our primary query results.

3. Discussion

We successfully implemented a proof of concept to store the semantics of imaging data important to biomarker research, and were able to use it to retrieve data which was more extensive and enriched in comparison to information available in a PACS database. Therefore, our results confirmed our hypothesis.

The storage of DICOM information in an RDF repository has an advantage over generic solutions where PACS databases store the entire (or very small parts of the) header. Such PACS relational databases are less flexible in changing data structures (and data), where the Semantic Web technology is more flexible in nature. For instance, within an RDF repository, adding extra information at a later point in time does not alter or change the existing data. Furthermore, it creates the possibility for everyone to link their own additional findings to the already available datasets. Therefore, we expect that the Semantic Web technology will provide a more suitable framework for research purposes, where questions regarding data (and/or calculated features) are still changing. Such a framework can increase the possibilities for data mining, especially when resources are linked to eCRF or EHR data stored in other RDF repositories which are accessible as linked data and through SPARQL endpoints.

We still plan to address several issues. At first, 4 out of 10 patients did not have a proper name describing the gross tumor volume (within the DICOM metadata). This description is a free-text field in many delineation applications; therefore, we cannot expect that delineation descriptions are always correct. We need a proper (ontological) mapping to determine relevant delineation descriptions. The RTOG made a first attempt to standardize delineation naming [13], however it would take several years to reach compliance to this standard. Second, we have created two different applications, which operate without coordination. In future work, we plan to implement a messaging system where the SeDI conversion service will notify the computation pipeline that the DICOM metadata has been stored in a RDF repository, and that computation on this image set can start. Third, although we are able to store the data in an RDF repository, we have not made use of the Value Representation (VR) types available in the DICOM metadata. Currently, all values are stored as text, instead of using the Value Representation (VR) type. This will be solved in our ontology, and incorporated into our SeDI conversion service.

We have shown that a semantic representation of DICOM related information enables us to query for image features that are important to biomarkers in radiotherapy. Our work towards a semantic PACS for radiotherapy can eventually be generalized to store and retrieve images using the semantics of other domains.

References

1. The National Electrical Manufacturers Association, Digital Imaging and Communications in Medicine (DICOM), NEMA Publications, PS3.1-PS3.12, 2011.
2. A.P. Abernethy, L.M. Etheredge, P.A. Ganz, P. Wallace, R.R. German, C. Neti, et al., Rapid-learning system for cancer care, *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* 28 (2010) 4268–4274.
3. P. Lambin, E. Roelofs, B. Reymen, E.R. Velazquez, J. Buijsen, C.M.L. Zegers, et al., Rapid Learning health care in oncology – An approach towards decision support systems enabling customised radiotherapy, *Radiother. Oncol.* (2013).
4. M.A. Levy, J.B. Freymann, J.S. Kirby, A. Fedorov, F.M. Fennessy, S.A. Eschrich, et al., Informatics methods to enable sharing of quantitative imaging research data, *Magn. Reson. Imaging.* 30 (2012) 1249–1256.
5. T. Berners-Lee, J. Hendler, O. Lassila, The semantic web, *Sci. Am.* 284 (2001) 28–37.
6. T. Berners-Lee, Linked data-the story so far, *Int. J. Semantic Web Inf. Syst.* 5 (2009) 1–22.
7. D. Allemang, *Semantic Web for the working ontologist effective modeling in RDFS and OWL*, Morgan Kaufmann/Elsevier, Waltham, MA, 2011.
8. N.F. de Keizer, A. Abu-Hanna, J.H. Zwetsloot-Schonk, Understanding terminological systems. I: Terminology and typology, *Methods Inf. Med.* 39 (2000) 16–21.
9. D.L. Rubin, D.A. Moreira, P. Kanjamala, M.A. Musen, BioPortal: A Web Portal to Biomedical Ontologies., in: *AAAI Spring Symp. Symbiotic Relatsh. Semantic Web Knowl. Eng.*, 2008: pp. 74–77.
10. D. Grittner, J. van Soest, T. Lustberg, M.S. Marshall, P. Feltens, A. Dekker, *Semantic DICOM Ontology*, <http://biportal.bioontology.org/ontologies/SEDI>. (2014).
11. dcm4che, <http://www.dcm4che.org/>. (2014).
12. Spring Framework, <Http://spring.io/>. (2014).
13. L. Santanam, C. Hurkmans, S. Mutic, C. van Vliet-Vroegindeweij, S. Brame, W. Straube, et al., Standardizing Naming Conventions in Radiation Oncology, *Int. J. Radiat. Oncol.* 83 (2012) 1344–1349.

Chapter 4

Radiation oncology terminology linker: A step towards a linked data knowledge base

Tim Lustberg, Johan van Soest, Peter Fick, Rianne Fijten, Tim Hendriks, Sander
Puts and Andre Dekker

Stud Health Technol Inform. 2018;247:855-859.

Abstract.

Performing image feature extraction in radiation oncology is often dependent on the organ and tumor delineations provided by clinical staff. These delineation names are free text DICOM metadata fields resulting in undefined information, which requires effort to use in large-scale image feature extraction efforts. In this work we present a scale-able solution to overcome these naming convention challenges with a REST service using Semantic Web technology to convert this information to linked data. As a proof of concept an open source software is used to compute radiation oncology image features. The results of this work can be found in a public Bitbucket repository.

Keywords. Linked data, Semantic Web, Radiation oncology, DICOM, ontology, semantic interoperability

1. Introduction

Outcome registration is of vital importance to move towards a rapid learning healthcare system for cancer patients [1]. It takes months to see the effect of the radiation or chemotherapy treatment on cancer and years to estimate the severity of some side effects. Together with the outcome data, the radiation oncology treatment features provide the knowledge which is needed to innovate. However, in daily clinical practice, this information is not stored as structured data, they need to be computed using the DICOM objects created during radiation treatment planning.

The DICOM standard does not enforce a terminology for creating delineations of organs in the radiation treatment planning process because the labels are free text. There are efforts in the field of radiation oncology to standardize the delineation practice [2]. However, clinics often have their own implementation of such a standard. Earlier studies have shown that the routine clinical data is a valuable source of information if you can get through the issues of missing, ambiguous and contradicting data [3–5].

To overcome the challenges of messy clinical data and automatically generate radiation oncology treatment features, a software platform is needed. The local naming terminology for the organ and tumor delineation needs to be linked to a terminology understood by the computational platform. Linking this data to other data sources, such as Semantic DICOM (SeDI) [6] and the Radiation Oncology Ontology (ROO) [7] creates a knowledge base to facilitate future studies to predict treatment outcome for cancer patients treated with radiation.

The aim of this study is to provide such a platform as an expandable open source software library that can be used to generate and store radiation oncology treatment features as linked data with minimal user interaction.

2. Methods

In previous work, Semantic Web technology was used to enhance the query-ability of the information stored in the DICOM metadata, using the Semantic DICOM software and ontology [<http://semantic-dicom.org>]. Linking the DICOM header data to radiation oncology treatment features enabled quick access to for instance CT-scans with a certain tumor size quickly. One of the limitations of this study was the manual work required to pre-process the tumor volumes. In this work,

we added three features to eliminate these limitations: Terminology mapping service to link delineation naming terminologies, an ontology to link all the results together and a flexible computational platform for calculating imaging features.

2.1. Terminology mapping service

In the radiation oncology community, there are efforts to standardize the organ and tumor delineation naming conventions. However, when investigating routine clinical data, the delineation names will often deviate from the convention due to for instance: typos, interpretation of a convention or the preference to use a native language in non-English speaking countries. To make this data machine-interpretable these local terminologies need to be mapped to a terminology that the computational platform understands

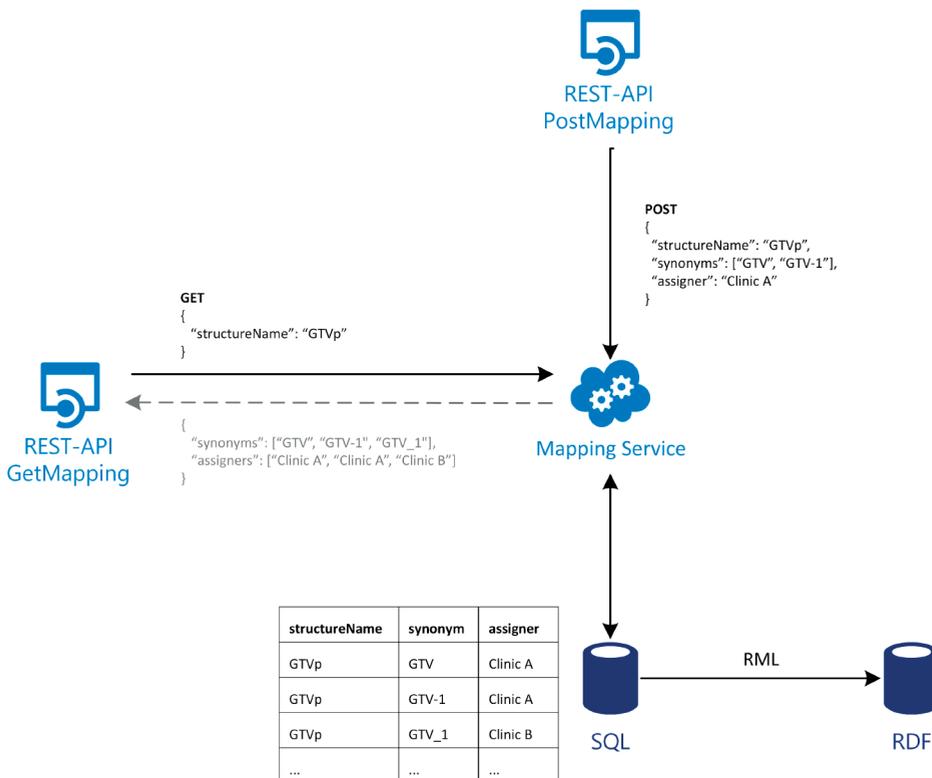


Figure 1. Mapping Service overview. Communicating with the REST API using JSON data transfer objects. The mappings are converted into linked data using RML

The terminology Mapping Service (Figure 1) provides a REST API to solve this problem. The REST API uses an SQL database to store the terminology mappings and the governance of these mappings, recording the user and institute that created the mapping. Any application can use this service to store and retrieve synonyms to enable an image analysis application to choose the correct delineations when computing image features.

2.2. Medical Image Analysis Platform

As a continuation of the feature calculation software used in the Semantic DICOM project, the Medical Image Analysis platform was created. The goal was to make the software more scalable and easier to use. To accomplish this, the software was re-implemented with a micro service design, based on the Netflix OSS framework [<https://netflix.github.io/>], where every functional part is its own REST service (Figure 2). The example calculations (worker) are implemented in Matlab 2015b with an object-oriented design to extract image features from radiation oncology DICOM objects. However, the framework was designed to compute features using any image feature extraction software in any programming language.

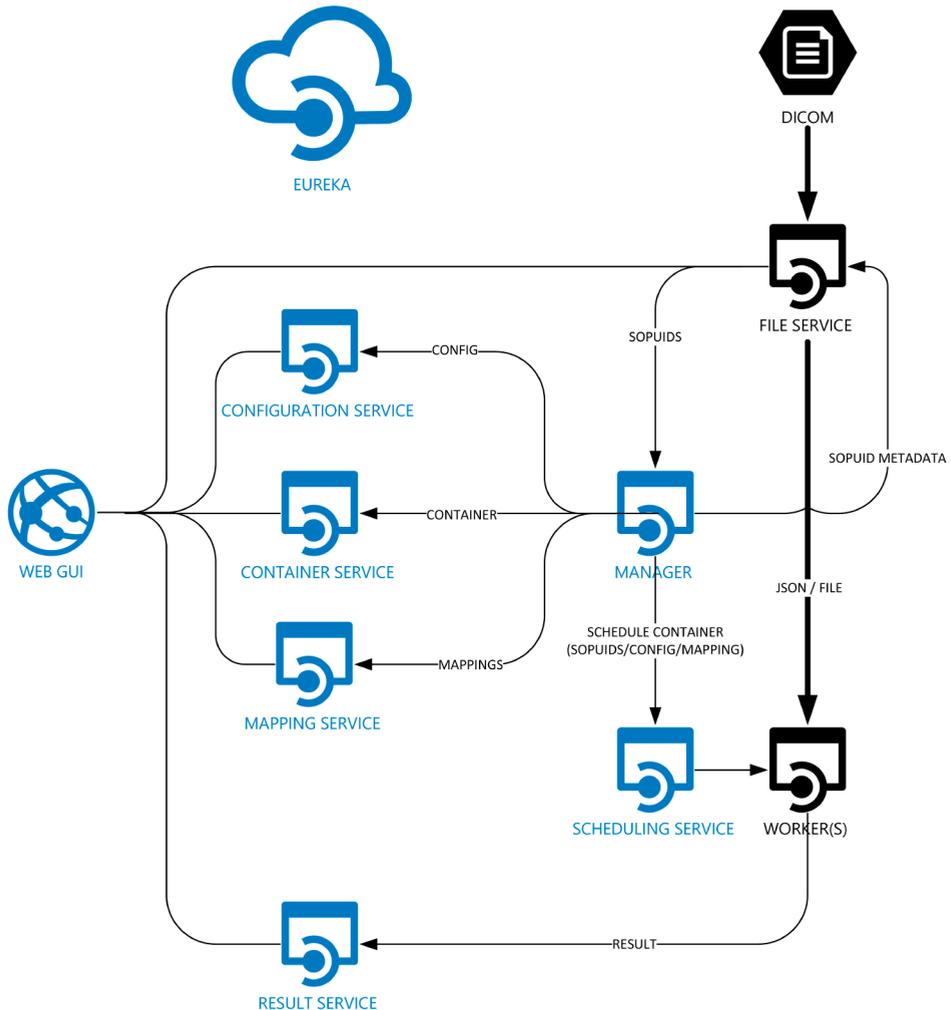


Figure 2. MIA overview. The micro-service platform to manage the steps (services) needed for image feature extraction using a web-based user-interface.

2.3. Linked data representation of image features using Semantic Web technology

To create the linked data knowledge base needed to answer complex questions the data generated by the software components described in this work, we added RDF Mapping Language [www.rml.io] conversions scripts to create triples from the SQL data (Figure 3). This reproduces the result of the SeDI project where the image features are linked to the DICOM metadata to answer more complex questions a regular PACS system cannot.

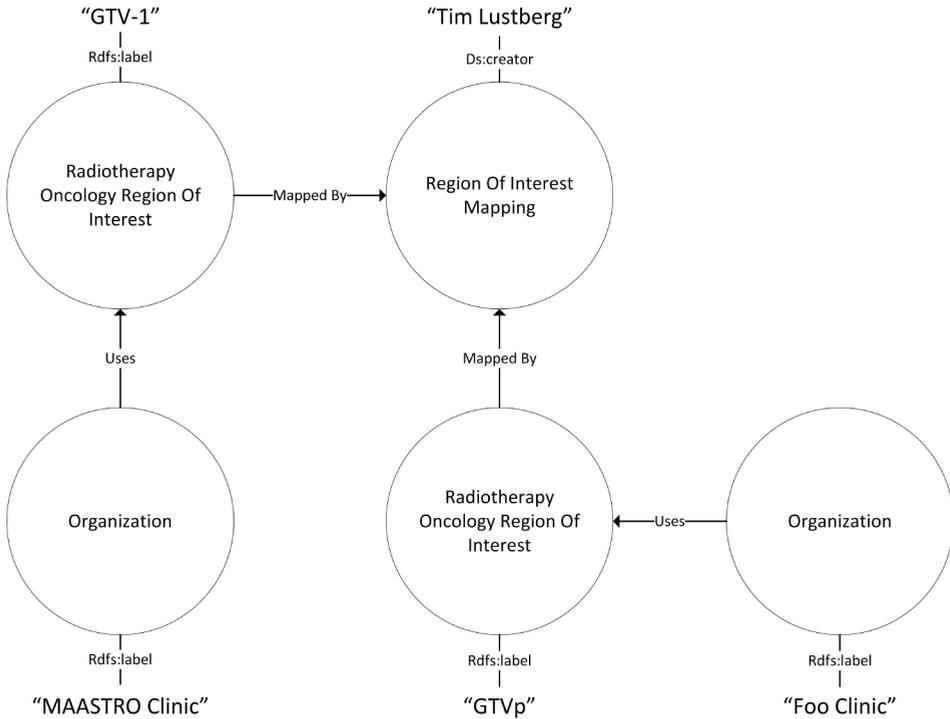


Figure 3. Structure name mapping ontology

3. Results

The software is available as an Open Source project on a public Bitbucket repository [<https://bitbucket.org/account/user/maastrosdt/projects/MIA>]. The current release of the MIA and the Mapping Service works as described in the methods and is able to handle large volumes of DICOM data to quickly extract image features.

4. Discussion

The Mapping Service provides an easy API for software developers to work with when creating a terminology mapping application. Because the problem we wanted to solve in this work was to compute clinical image delineations labeled with free text, the Mapping Service was used to achieve this goal. However, in future implementation efforts this service could be reused to map any terminology. Using new RML scripts to link these mappings to an ontology could be of interest in other fields.

On top of the analysis uncertainties, there is a far bigger concern when computing image features in multiple centers based on human delineations. It has been reported that delineation practices vary between radiation oncologists in the same hospital using the same guidelines and even more so when comparing delineations from multiple centers.

In the future, we would like to move to a system where the Mapping Service will become obsolete, because the delineation labels will be based on a proper terminology and every clinician in the world used the same guidelines. However, we have to acknowledge that this might never be achieved because it requires a level of consensus which is rarely reached in medicine. The linked data systems presented in this work provide a scalable platform to deal with these imperfections to create a knowledge base for further research.

The Mapping Interface included with the MIA could be improved in numerous ways. Using the actual imaging data attached to the delineation names could provide a knowledge base for the software to learn what an organ looks like and how organs are positioned relative to each other. This information could be used to for instance identify the heart if the two delineations of the two lungs are known. This would further decrease the need for human intervention. Another improvement would be to show the images with the delineations on the Mapping Interface. This would make it easier for the user to identify the correct delineation when mapping it to a terminology, as opposed to guessing based on the names provided.

References

1. P. Lambin, E. Roelofs, B. Reymen, E.R. Velazquez, J. Buijsen, C.M.L. Zegers, S. Carvalho, R.T.H. Leijenaar, G. Nalbantov, C. Oberije, M. Scott Marshall, F. Hoebbers, E.G.C. Troost, V. Stiphout, R. G.p.m, W. van Elmpt, T. van der Weijden, L. Boersma, V. Valentini, A. Dekker, "Rapid Learning health care in oncology" – An approach towards decision support systems enabling customised radiotherapy', *Radiother. Oncol.* 109 (2013) 159–164. doi:10.1016/j.radonc.2013.07.007.
2. L. Santanam, C. Hurkmans, S. Mutic, C. van Vliet-Vroegindeweij, S. Brame, W. Straube, J. Galvin, P. Tripuraneni, J. Michalski, W. Bosch, Standardizing Naming Conventions in Radiation Oncology, *Int. J. Radiat. Oncol.* 83 (2012) 1344–1349. doi:10.1016/j.ijrobp.2011.09.054.
3. A. Dekker, S. Vinod, L. Holloway, A. George, G. Goozee, G. Delaney, et al, Rapid learning in practice: A lung cancer survival decision support system in routine patient care data, *Radiat. Oncol.* 2014;111(Suppl1):35 (n.d.).
4. T. Lustberg, M. Bailey, D.I. Thwaites, A. Miller, M. Carolan, L. Holloway, E.R. Velazquez, F. Hoebbers, A. Dekker, T. Lustberg, M. Bailey, D.I. Thwaites, A. Miller, M. Carolan, L. Holloway, E. Rios Velazquez, F. Hoebbers, A. Dekker, Implementation of a rapid learning platform: Predicting 2-year survival in laryngeal carcinoma patients in a clinical setting, *Oncotarget.* 7 (2016) 37288–37296. doi:10.18632/oncotarget.8755.
5. T.M. Deist, A. Jochems, J. van Soest, G. Nalbantov, C. Oberije, S. Walsh, M. Eble, P. Bulens, P. Coucke, W. Dries, A. Dekker, P. Lambin, Infrastructure and distributed learning methodology for privacy-preserving multi-centric rapid learning health care: euroCAT, *Clin. Transl. Radiat. Oncol.* 4 (2017) 24–31. doi:10.1016/j.ctro.2016.12.004.
6. J. Van Soest, T. Lustberg, D. Grittner, M.S. Marshall, L. Persoon, B. Nijsten, P. Feltens, A. Dekker, Towards a semantic PACS: Using Semantic Web technology to represent imaging data, *Stud. Health Technol. Inform.* 205 (2014) 166–170.
7. Radiation Oncology Ontology - ROO | CancerData.org, (n.d.). <https://www.cancerdata.org/roo> (accessed November 18, 2015).

Chapter 5

Set-up verification and 2D EPID dosimetry during breath hold compared to free breathing in breast cancer radiotherapy

Patricia Brouwers, Tim Lustberg, Jacques Borger, Angela van Baardwijk,
Jos Jager, Lars Murrer, Bas Nijsten, Bart Reymen, Judith van Loon,
Liesbeth Boersma

Pract Radiat Oncol. 2015 May-Jun;5(3):e135-41.

Abstract

Purpose:

To compare set-up and 2D electronic portal imaging device (EPID) dosimetry data of breast cancer patients treated during voluntary moderately Deep Inspiration Breath Hold (vmDIBH) and free breathing (FB). |

Patients and methods:

Set-up data were analyzed for 29 and 51 consecutively treated patients, irradiated during FB and vmDIBH, respectively. Of the 51 vmDIBH patients, the first 25 had undergone an extra trained CT scan and used an additional “breathing stick” (vmDIBH_trained). The last 26 patients did not use the breathing stick and did not undergo a trained CT (vmDIBH_untrained). The delivered 2D transit dose was measured with EPIDs in 15 FB and 28 vmDIBH patients, and compared to a 2D predicted dose by calculating global gamma values γ using 5% and 5 mm as dose difference and distance-to-agreement criteria, respectively. Measurements with a percentage of pixels $|\gamma| > 1$ over 10% were classified as deviating.

Results:

Only small, sub mm differences were seen in the set-up data between the different patient groups. The mean of means, systematic error and random error ranged from -0.6 mm to 3.3 mm. The percentage of pixels with $|\gamma| > 1$ for all patients was 9.8% (2-25.8). No statistically significant differences were observed between the patient groups. In total, 38% of the gamma images were classified as deviating: 43.6% in vmDIBH_untrained patients, compared to 38.0% in vmDIBH_trained patients and 33.3% in FB patients ($p > 0.05$).

Conclusion:

Both set-up and 2D EPID dosimetry data indicate that reproducibility of radiotherapy for patients treated during FB and vmDIBH is similar. Small but not significant differences in 2D EPID dosimetry were observed. Further investigation with 3D EPID dosimetry is recommended to investigate the clinical relevance of deviant gamma images.

Introduction

The conventional technique of breast and thoracic wall irradiation is by tangential fields, implicating that the heart and lungs can be partially located within the radiation field. Several reports have shown that heart irradiation may lead to late cardiac toxicity[1]. In addition, several papers report that lung injury occurs due to breast irradiation, although the incidence of clinically relevant radiation pneumonitis is fortunately quite low [2].

In the literature, a significantly increased risk of cardiac death has been observed for patients treated with left sided breast cancer 20-30 years ago [3]. The incidence of cardiac injury using modern radiotherapy techniques however is not completely clear. Although Offersen *et al* [4] described several uncertainties with respect to parameters related to radiation-induced heart injury, it seems clear that both radiation dose and volume play an important role in the development of both heart and lung toxicity. Consequently, reducing both irradiated volume and dose to the heart and lung is expected to reduce heart and lung toxicity.

Multiple respiratory techniques have been described to spare both heart and lung, taking into account that during deep inspiration the heart moves out of the radiation field and the relative volume of irradiated lung is reduced. This can be achieved using voluntary breath hold techniques [5–8], an Active Breathing Control (ABC) device [9] or gating [7,10]. For the latter two additional equipment is required, while the voluntary method seems to be easy and inexpensive. However, concerning voluntary breath hold, the reproducibility is often questioned. Only a limited number of studies [5,11–13] analyzed set-up in breath hold, but none of these studies reported actual measured transit dose for verification.

The aim of this paper is to investigate whether the reproducibility of voluntary moderately Deep Inspiration Breath Hold (vmDIBH) is similar to free breathing (FB) by reporting set-up and 2D EPID dosimetry data, acquired during the development of the vmDIBH technique in our institute.

Patients and Methods

Patients

In our institute, vmDIBH was implemented in 2005. In 2008 we started a step by step process to simplify the logistics, i.e. two CT scans at one day, instead of three CT-scans at two days, as described below. Data were obtained in 80 consecutive

breast cancer patients, treated in a fixed period between October 2008 and October 2009 (Figure 1). During the period of the simplification process three treatment groups could be distinguished:

1. Control group, i.e. FB patients (N=29)
2. vmDIBH patients with three CT scans and using an additional aid called the "breathing stick" (N=25), developed to assist patients to achieve a reproducible breath hold (Figure 2). These patients first underwent two planning CT scans: one during FB, and one during untrained vmDIBH [9]. Both scans were compared, and if the maximum heart distance (MHD, Figure 3). was ≥ 1 cm in the FB scan, and < 1 cm in the vmDIBH, it was decided to treat the patient during breath hold. The patient was then phoned and asked to practice the breath hold at home, using written breath hold instructions. After practicing breath hold at home, a third, trained vmDIBH CT scan was obtained, which was used for treatment planning.
3. vmDIBH patients with only 2 untrained CT scans and without the "breathing stick" (N=26). To investigate whether omitting the third CT scan would be safe, we performed an "interim in-silico" comparison in 10 patients of group 2: the Digitally Reconstructed Radiographs (DRRs) for the tangential fields of the untrained vmDIBH CT scan were compared with the DRRs of the trained vmDIBH CT scan. We found a small set-up difference, manageable by our set-up protocol, but no clinically relevant difference was seen in the MHD (mean difference 0.3 mm +/- 1.8). Based upon this analysis, we decided to omit the third trained vmDIBH CT scan, and used the untrained vmDIBH CT scan for treatment planning. At the same time we omitted the "breathing stick".

Breath hold compared to free breathing in breast cancer radiotherapy

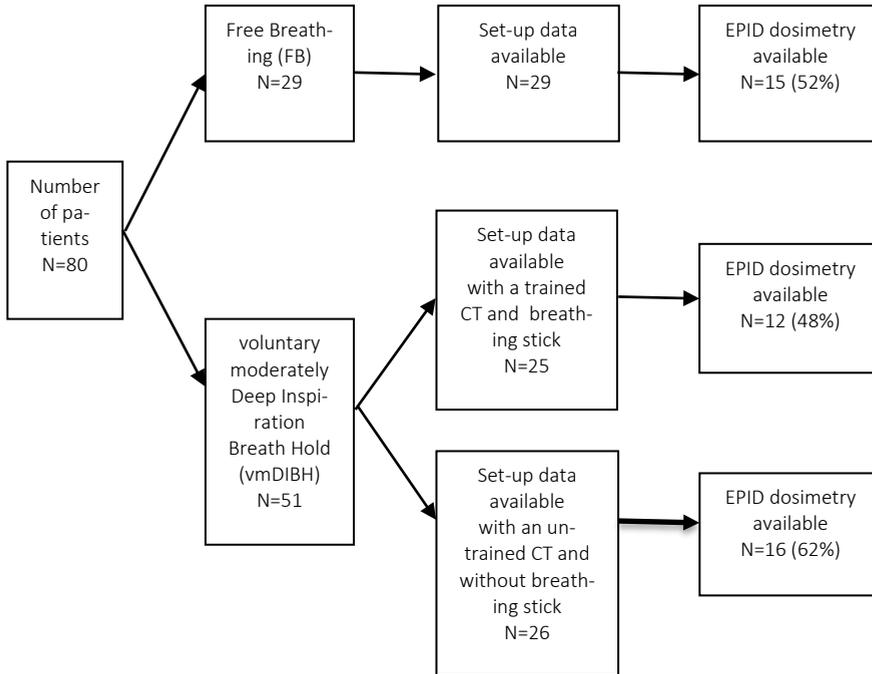


Figure 1. Schematic overview of number of patients and available data.

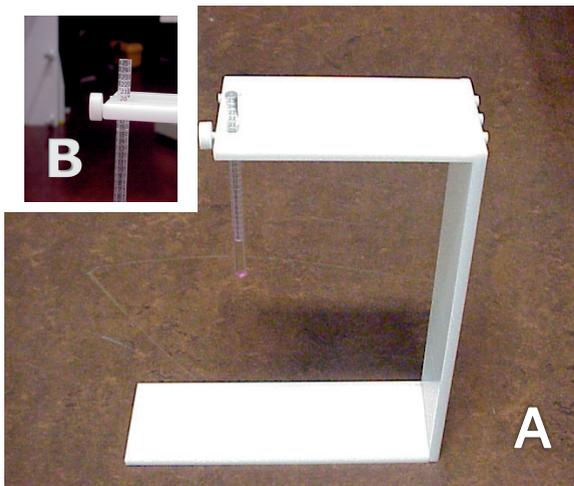


Figure 2. (A) The “breathing stick” which was placed on the skin of the patient outside the radiation fields in the epigastric area during breath hold and (B) an integrated ruler. The contact point of the tip of the ruler was marked by a dot on the skin during the third CT scan in vmDIBH. During treatment the radiation technician could check visually, and the patient could feel, whether the tip of the stick made contact at the dot on the skin, ensuring a correct breath hold.

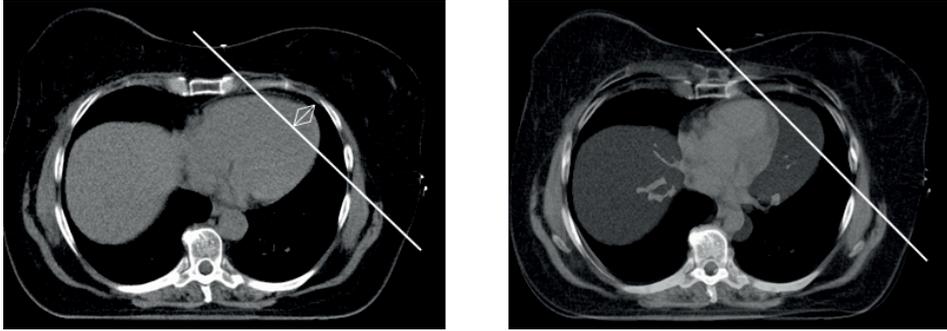


Figure 3. The MHD is the maximum distance between the heart contour and the posterior field border of a tangential treatment beam. A: Free breathing, the heart is partially located within the radiation field. B: in vmDIBH the heart moves out the radiation field.

Methods

CT scanning and treatment planning

Patients were scanned in supine position with the arms above the head in an arm support (Civco, Posirest-2, USA), and the legs resting on a Kneefix (Civco, USA). All CT scans were obtained with 3 mm slice thickness from the level of the mandible down to the diaphragm (Siemens Somatom Sensation). Treatment planning was performed using forward IMRT planning as described earlier [14]. The target volume consisted of the breast or thoracic wall with or without regional lymph nodes.

Set-up verification

A Shrinking Action Level (SAL) protocol was used for set-up verification with $\alpha=10$ and $n=3$ [15] for all patients. Four skin markers were placed onto the skin at the medial, lateral, cranial and caudal edge of the breast or thoracic wall. Both lateral and AP electronic portal images (EPIs) were matched to the DRRs using both anatomy and skin markers [16].

Differences between the EPIs and the DRRs were analyzed in three directions: left-right, cranio-caudal and anterior-posterior. The mean of means (μ) was determined by calculating the average of the individual systematic set-up errors for all patients. The systematic set-up error for the population (Σ) was calculated by taking the standard deviation of the individual systematic set-up errors for all patients. The random set-up error of the population (σ) was calculated by taking the average of the individual random set-up errors for all patients [17].

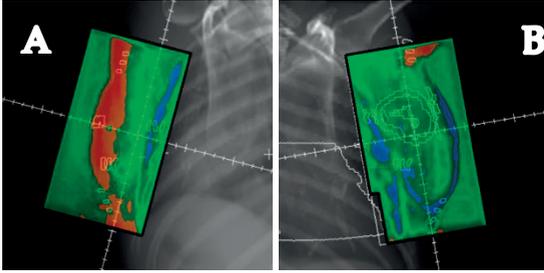


Figure 4. Examples of gamma images of a tangential breast field. Red and blue colors represent regions where the measured dose is higher or lower than planned, respectively, while green represents regions where the planned and measured dose are in agreement. (A) a typical simple translation, and (B) an example of rotation or changes in breast shape.

Dose verification

According to the clinical protocol, two-dimensional transit dose distributions were measured during the first three fractions, thereafter weekly, using electronic portal images (EPIs). OptiVue 500/1000/1000 ST amorphous silicon flat panel portal imagers (Siemens Medical Solutions, Concord, CA), attached to OncoRay medical linear accelerators (Siemens Medical Solutions, Concord, CA), were used for these measurements. The measured delivered transit dose was compared to a predicted dose by calculating global gamma values using 5% and 5 mm as dose difference and distance-to-agreement criteria [18]. The percentage of pixels with an absolute gamma value >1 ($|\gamma|>1$) was determined only for latero-medial and medio-lateral beams. Median values of this parameter were compared between the three patient groups. Measurements with a percentage of pixels $|\gamma|>1$ over 10% were classified as deviating. Deviant gamma images were inspected visually. Only if one part of the image showed clear underdosage and the other part showed clear overdosage, the dose difference was considered to be due to a simple translation (set-up error / organ motion); if the gamma image showed other patterns of over- and/or underdosage, the dose difference was considered to be due to rotation, change in breast shape or a combination of both (Figure 4).

Due to the fact that treatment fields did not always fit within the field of view of the flat panel portal imagers, EPID dosimetry data could not be evaluated for all patients. In total, 241, 590, and 565 measurements were obtained for 15 FB, 12 vmDIBH patients with trained CT and 16 vmDIBH patients with an untrained CT, respectively (Figure 1).

Analysis and statistics

Set-up data were analyzed and compared for the three treatment groups: 25 vmDIBH patients treated with trained CT (vmDIBH_trained), 26 vmDIBH patients without trained CT (vmDIBH_untrained) and 29 FB patients. Mean of means, random error, and systematic error of the vmDIBH patients were calculated for this purpose. In addition, the number of measurements and applied set-up corrections was counted for the three treatment groups. The percentage of deviant measurements and of dose differences interpreted as due to a translation were compared between the patient groups. We were particularly interested in translations, since we assumed that a change in breath hold would likely show up as translations. In all analyses, the Kruskal-Wallis test was used with $p < 0.05$ as level of significance. In case of a significant difference, a Mann-Whitney test was used for detailed analysis. Median results are noted with range in parentheses.

Results

Patient characteristics are listed in Table 1. All vmDIBH patients had left sided breast cancer, with an assumed benefit from vmDIBH, based upon a reduction of the maximum heart distance (Figure 3). FB patients had right or left sided breast cancer, assuming that laterality does not influence set-up accuracy. In one patient vmDIBH was not feasible because of shortness of breath; in all other patients vmDIBH had no added value.

Table 1. Patient characteristics.

| | Median age and range | Type of surgery | |
|------------------|-------------------------|-----------------|------------------------------|
| | | Mastectomy | Breast Conserving Surgery |
| Free Breathing | 64 (42-83) | 9 | 20 |
| vmDIBH_trained | 53 (41-68) | 4 | 21 |
| vmDIBH_untrained | 51 (37-70) | 2 | 24 |

Using the Kruskal Wallis test, age of FB patients is significant higher than both vmDIBH patient groups ($p < 0.001$). Type of surgery is, although a clear trend is seen, not significant differently (Chi square test: $p > 0.05$).

Analysis of set-up data

The systematic error in all patients varied between 1.2 and 2.0 mm; the random error between 2.2 and 3.3 mm. Mean of means varied between -0.6 and 0.9 mm (Table 2). The random error showed a significant difference in cranio-caudal direction for FB compared to vmDIBH_trained, whereas the mean of means showed a significant difference in anterior-posterior direction for FB compared with vmDIBH_trained and vmDIBH_untrained. Although the number of measurements for vmDIBH_untrained was larger than for FB patients ($p=0.03$), the number of set-up corrections was similar ($p>0.2$) (Table 3).

Table 2. Set-up data of all patients and subdivided by treatment group.

| | Mean of means μ (mm) | | | Systematic error Σ (mm) | | | Random error σ (mm) | | |
|-------------------------|--------------------------|-----|------|--------------------------------|-----|-----|----------------------------|------|-----|
| | LR | CC | AP | LR | CC | AP | LR | CC | AP |
| All patients (N=80) | 0.4 | 0.7 | -0.0 | 1.7 | 1.6 | 2.0 | 2.8 | 2.7 | 3.2 |
| Free Breathing (N=29) | 0.2 | 0.7 | 0.7 | 2.0 | 1.6 | 2.0 | 2.9 | 2.2 | 3.1 |
| vmDIBH_trained (N=25) | 0.4 | 0.6 | -0.3 | 1.2 | 1.9 | 2.0 | 2.8 | 3.0 | 3.1 |
| vmDIBH_untrained (N=26) | 0.6 | 0.9 | -0.6 | 1.7 | 1.4 | 1.9 | 2.6 | 3.0 | 3.3 |
| p value | 0.7 | 0.9 | 0.04 | | | | 0.7 | 0.01 | 0.8 |

Using a Kruskal-Wallis test, significant differences were found in mean of means (AP) and random error (CC) between the 3 groups. Detailed analysis using a Mann-Whitney test showed a significant difference in mean of means in AP direction between FB and vmDIBH_untrained ($p=0.02$) and in random error in CC direction between FB and vmDIBH_trained and vmDIBH_untrained ($p<0.02$).

Table 3. Median numbers of set up measurements and median number of corrections per patient for the three patient groups.

| | Median number of measurements (range) | Median number of corrections (range) |
|------------------|---------------------------------------|--------------------------------------|
| Free Breathing | 8 (4-12) | 1 (0-4) |
| vmDIBH_trained | 9 (4-18) | 1 (0-7) |
| vmDIBH_untrained | 10 (5-16) | 1 (0-4) |

Comparing the three groups with the Kruskal-Wallis test, vmDIBH_untrained patients had a larger number of set up measurements than FB patients ($p=0.03$). However, the number of corrections per patient was similar for the three patient groups ($p>0.2$).

Analysis of 2D EPID dosimetry data

The median percentage of pixels with an absolute gamma value >1 ($|\gamma|>1$) for all patients was 9.8% (2-25.8%) (Table 4). The percentage of deviating images was somewhat higher in vmDIBH patients (vmDIBH_trained (38%) and vmDIBH_untrained (43.6%)) than in FB (33.3%), but the difference was not statistically significant. Visual inspection of deviant gamma images showed no significant difference in the percentage of deviating gamma images attributed to translation in vmDIBH_trained, vmDIBH_untrained and FB patients (14.9% (0-66.6%), 12.7% (0-83.3%) and 0 (0-100%, respectively). No differences were seen between medio-lateral and latero-medial beams (all p-values >0.3).

Table 4. Analysis of 2D EPID dosimetry data.

| | Median % of pixels with $ \gamma >1$ per patient and range | Median % of deviating measurements and range | Median % of deviating measurements resulting from a translational set-up error and range |
|---------------------------------|--|--|--|
| All patients together N=1396 | 9.8 (2.0-25.8) | 38.0 (4.2-81.3) | 13.7 (0-100) |
| Free breathing N=241 | 8.2 (2.0-25.0) | 33.3 (7.1-64.3) | 0.0 (0.0-100) |
| vmDIBH_trained N=590 | 10.7 (2.3-1.6) | 38.0 (4.2-72.0) | 14.9 (0-66.6) |
| vmDIBH_untrained N=565 | 12.3 (4.8-25.8) | 43.6 (12.5-81.3) | 12.7 (0-83.3) |

No significant differences were found ($p>0.3$, Kruskal- Wallis test).

Discussion

To our knowledge, this is the first paper reporting on both set-up and *in vivo* dosimetric data obtained during vmDIBH. Although the dosimetry data showed remarkably large deviations, differences in set-up errors between free breathing and vmDIBH were extremely small (sub mm), and very comparable with data reported earlier in FB patients [19,20]. Therefore, we consider vmDIBH to be as reproducible as standard FB techniques. Furthermore, we showed that by careful step by step introduction of vmDIBH, an easy, widely applicable procedure can be obtained.

Comparison with literature

Set-up data specific for vmDIBH are sparsely available in literature. Lu *et al* performed repeated CT scans and showed that all patients (N=15) except one reasonably reproduced their position between different breath-holding cycles [7]. The Netherlands Cancer Institute described set-up data in patients during radiotherapy in DIBH. Set-up deviations (systematic error) in the order of 1.4-2.9 mm [13], ≤ 1.7 mm [11] and ≤ 1.4 mm were reported [12], which are comparable to the 1.2-2 mm found in the current study.

Theoretically, one would assume that Active Breathing Control (ABC), i.e. breath hold with an additive device, results in better set-up reproducibility. This assumption, however, is not confirmed in literature [9,21,22]. Recently, the results of the UK HeartSpare Study were published, which suggest that ABC and vmDIBH are comparable both in terms of positional reproducibility and normal tissue sparing. In addition, patients experience vmDIBH more comfortable than an ABC [23].

Although our set-up errors were reasonably small and comparable to literature, the measured dose showed remarkably large deviations, (i.e. $>10\%$ pixels with gamma values >1), both for vmDIBH and for FB patients. The clinical relevance of these deviations in 2D dosimetry is yet unclear. To our knowledge, only Fidanzio *et al* [24] published results of breast *in vivo* dosimetry using EPID, but they did not take into account the entire tangential field. They determined the ratio between reconstructed and planned dose at breast midpoint, thus verifying a point dose in the patient and not a 2D transit dose distribution at the EPID plane. When patient set-up variations were not taken into account, these ratios were within 5% in 72% of the checks. As no comparable dosimetry data are available in literature, and since our set-up data are comparable to those reported in literature, we assume that these dosimetry results would be found in other institutes as well, if measured. This assumption is supported by the paper of Topolnjak *et al* [25], who showed that set-up verification using EPIDs underestimates set-up errors in breast cancer patients by 20-50%, compared to cone beam CT data as a gold standard, thereby possibly explaining the poor dosimetry results. Another explanation for the high number of deviant dose measurements could be the high variation in tissue density, e.g. lung versus chest wall and breast versus air, such that small set-up or breath hold differences can easily lead to large differences between predicted and measured transit dose. The latter could also be an explanation for the, although not statistically significant,

differences in percentage of deviant dose measurements between FB and vmDIBH patients.

Limitations

There are some limitations of the study inherent to its design. First, the number of transit dose measurements was limited, since radiation fields used for breast cancer are often larger than the field of view of the EPID. The limited size of the flat panel portal imagers is an ongoing problem in case of large treatment volumes. Due to this limited amount of EPID dosimetry data, the study might have been underpowered to detect a significant difference among the three groups. Secondly, we verified 2D measured transit dose distributions and not the 3D delivered dose inside a patient. Gamma criteria were deliberately chosen larger (5% and 5 mm; in accordance with protocol at that time) than usually applied in other publications; in literature, gamma criteria of 3% and 3 mm are often used, but usually for pre-treatment verification purposes and for verification of the 3D dose inside the patient. Due to changes in breast anatomy and set-up errors, and their large impact on the measured transit dose, the broader gamma criteria were chosen to get a better threshold to decide whether the expected 3D dose differences would be clinically significant, and to balance clinical relevance and workload. Although for verification of the 3D delivered dose, 3D EPID dosimetry is preferred, Nijsten *et al* [26] showed earlier that 2D EPID dosimetry can be used to predict changes in DVH parameters, indicating that there is a relation between our 2D data and the 3D delivered dose, at least for this patient group using tangential breast fields for radiotherapy treatment.

To conclude, both set-up and 2D EPID dosimetry data indicate that reproducibility of radiotherapy for patients treated during FB and vmDIBH is similar. The observed differences in 2D EPID dosimetry were not statistically significant between vmDIBH and FB techniques. However, further investigation with 3D EPID dosimetry is recommended to investigate the clinical relevance of deviant gamma images. We are currently investigating how EPID dosimetry can be used to develop decision rules for adaptive radiotherapy.

References

1. Darby SC, Ewertz M, McGale P, Bennet AM, Blom-Goldman U, Brønnum D, *et al.* Risk of ischemic heart disease in women after radiotherapy for breast cancer. *The New England Journal of Medicine* 2013;368(11):987–998. doi:10.1056/NEJMoa1209825
2. Kubo A, Osaki K, Kawanaka T, Furutani S, Ikushima H, Nishitani H. Risk factors for radiation pneumonitis caused by whole breast irradiation following breast-conserving surgery. *The Journal of Medical Investigation : JMI* 2009;56(3–4):99–110
3. Ewertz M, Jensen AB. Late effects of breast cancer treatment and potentials for rehabilitation. *Acta Oncologica (Stockholm, Sweden)* 2011;50(2):187–193. doi:10.3109/0284186X.2010.533190
4. Offersen B, Højris I, Overgaard M. Radiation-induced heart morbidity after adjuvant radiotherapy of early breast cancer - Is it still an issue? *Radiotherapy and Oncology : Journal of the European Society for Therapeutic Radiology and Oncology* 2011;100(2):157–159. doi:10.1016/j.radonc.2011.08.007
5. Lu HM, Cash E, Chen MH, Chin L, Manning WJ, Harris J, *et al.* Reduction of cardiac volume in left-breast treatment fields by respiratory maneuvers: a CT study. *International Journal of Radiation Oncology, Biology, Physics* 2000;47(4):895–904
6. Ming Hui Chen, MD, MMSc; Michael L. Chuang, MS, Bruce A. Bornstein, MD, Rebecca Gelman, PhD, Jay R. Harris, MD; Warren J. Manning M. Impact of Respiratory Maneuvers on Cardiac Volume Within Left-Breast Radiation Portals. *Circulation* 1997;96:3269–3272
7. Korreman SS, Pedersen AN, Aarup LR, Nøttrup TJ, Specht L, Nyström H. Reduction of cardiac and pulmonary complication probabilities after breathing adapted radiotherapy for breast cancer. *International Journal of Radiation Oncology, Biology, Physics* 2006;65(5):1375–1380. doi:10.1016/j.ijrobp.2006.03.046
8. Pedersen AN, Korreman S, Nyström H, Specht L. Breathing adapted radiotherapy of breast cancer: reduction of cardiac and pulmonary doses using voluntary inspiration breath-hold. *Radiotherapy and Oncology : Journal of the European Society for Therapeutic Radiology and Oncology* 2004;72(1):53–60. doi:10.1016/j.radonc.2004.03.012
9. Remouchamps VM, Letts N, Vicini F a., Sharpe MB, Kestin LL, Chen PY, *et al.* Initial clinical experience with moderate deep-inspiration breath hold using an active breathing control device in the treatment of patients with left-sided breast cancer using external beam radiation therapy. *International Journal of Radiation Oncology*Biography*Physics* 2003;56(3):704–715. doi: 10.1016/S0360-3016(03)00010-5
10. Korreman SS, Pedersen AN, Nøttrup TJ, Specht L, Nyström H. Breathing adapted radiotherapy for breast cancer: comparison of free breathing gating with the breath-hold technique. *Radiotherapy and Oncology : Journal of the European Society for Therapeutic Radiology and Oncology* 2005;76(3):311–318. doi:10.1016/j.radonc.2005.07.009
11. Alderliesten T, Sonke JJ, Betgen A, Honnef J, van Vliet-Vroegindeweyj C, Remeijer P. Accuracy evaluation of a 3-dimensional surface imaging system for guidance in deep-inspiration breath-hold radiation therapy. *International Journal of Radiation Oncology, Biology, Physics* 2013;85(2):536–542. doi:10.1016/j.ijrobp.2012.04.004
12. Betgen A, Alderliesten T, Sonke JJ, van Vliet-Vroegindeweyj C, Bartelink H, Remeijer P. Assessment of set-up variability during deep inspiration breath hold radiotherapy for breast cancer patients by 3D-surface imaging. *Radiotherapy and Oncology : Journal of the European Society for Therapeutic Radiology and Oncology* 2013;106(2):225–230. doi:10.1016/j.radonc.2012.12.016

13. Borst GR, Sonke JJ, den Hollander S, Betgen A, Remeijer P, van Giersbergen A, *et al.* Clinical results of image-guided deep inspiration breath hold breast irradiation. *International Journal of Radiation Oncology, Biology, Physics* 2010;78(5):1345–1351. doi:10.1016/j.ijrobp.2009.10.006
14. Peulen H, Hanbeukers B, Boersma L, van Baardwijk A, van den Ende P, Houben R, *et al.* Forward intensity-modulated radiotherapy planning in breast cancer to improve dose homogeneity: feasibility of class solutions. *International Journal of Radiation Oncology, Biology, Physics* 2012;82(1):394–400. doi:10.1016/j.ijrobp.2010.09.005
15. Bel a, van Herk M, Bartelink H, Lebesque J V. A verification procedure to improve patient set-up accuracy using portal images. *Radiotherapy and Oncology : Journal of the European Society for Therapeutic Radiology and Oncology* 1993;29(2):253–260
16. Salm A Van Der, Strijbos J, Dijcks C, Murrer L, Borger J, Boersma L. Use of skin markers and electronic portal imaging to improve verification of tangential breast irradiation. *Radiotherapy and Oncology : Journal of the European Society for Therapeutic Radiology and Oncology* 2009;90(1): 106–109. doi:10.1016/j.radonc.2008.05.014
17. van Herk M, Remeijer P, Rasch C, Lebesque J V. The probability of correct target dosage: dose-population histograms for deriving treatment margins in radiotherapy. *International Journal of Radiation Oncology, Biology, Physics* 2000;47(4):1121–1135
18. Nijsten SMJJG, van Elmpt WJC, Jacobs M, Mijnheer BJ, Dekker a. L. a. J, Lambin P, *et al.* A global calibration model for a-Si EPIDs used for transit dosimetry. *Medical Physics* 2007;34(10):3872. doi:10.1118/1.2776244
19. Kukulowicz PF, Debrowski A, Gut P, Chmielewski L, Wieczorek A, Kedzierawski P. Evaluation of set-up deviations during the irradiation of patients suffering from breast cancer treated with two different techniques. *Radiotherapy and Oncology : Journal of the European Society for Therapeutic Radiology and Oncology* 2005;75(1):22–27. doi:10.1016/j.radonc.2005.02.004
20. Yang DS, Yoon WS, Chung SY, Lee J a, Lee S, Park YJ, *et al.* Set-up uncertainty during breast radiotherapy. Image-guided radiotherapy for patients with initial extensive variation. *Strahlentherapie Und Onkologie : Organ Der Deutschen Röntgengesellschaft . [et Al]* 2013;189(4):315–320. doi:10.1007/s00066-012-0271-4
21. Remouchamps VM, Letts N, Yan D, Vicini F a, Moreau M, Zielinski J a, *et al.* Three-dimensional evaluation of intra- and interfraction immobilization of lung and chest wall using active breathing control: A reproducibility study with breast cancer patients. *International Journal of Radiation Oncology*Biological*Physics* 2003;57(4):968–978. doi:10.1016/S0360-3016(03)00710-7
22. Moran JM, Balter JM, Ben-David M a, Marsh RB, Van Herk M, Pierce LJ. Short-term displacement and reproducibility of the breast and nodal targets under active breathing control. *International Journal of Radiation Oncology, Biology, Physics* 2007;68(2):541–546. doi:10.1016/j.ijrobp. 2007.01.028
23. Bartlett FR, Colgan RM, Carr K, Donovan EM, McNair H a, Locke I, *et al.* The UK HeartSpare Study: Randomised evaluation of voluntary deep-inspiratory breath-hold in women undergoing breast radiotherapy. *Radiotherapy and Oncology : Journal of the European Society for Therapeutic Radiology and Oncology* 2013;108(2):242–247. doi:10.1016/j.radonc.2013.04.021
24. Fidanzio A, Greco F, Mameli A, Azario L, Balducci M, Gambacorta MA, Frascino V, Cilla S, Sabatino D PA. Breast in vivo dosimetry by EPID. *Journal Appl Clin Med Phys* 2010;2(11):3275
25. Topolnjak R, Sonke JJ, Nijkamp J, Rasch C, Minkema D, Remeijer P, *et al.* Breast patient setup error assessment: comparison of electronic portal image devices and cone-beam computed tomography matching results. *International Journal of Radiation Oncology, Biology, Physics* 2010;78(4):1235–1243. doi:10.1016/j.ijrobp.2009.12.021

26. Nijsten SMJJG, van Elmpt WJC, Mijnheer BJ, Minken a. WH, Persoon LCGG, Lambin P, *et al.* Prediction of DVH parameter changes due to setup errors for breast cancer treatment based on 2D portal dosimetry. *Medical Physics* 2009;36(1):83–94. doi:10.1118/1.3026660

Chapter 6

Implementation of a rapid learning platform: Predicting 2-year survival in laryngeal carcinoma patients in a clinical setting

Tim Lustberg, Michael Bailey, David Thwaites, Andrew Miller, Martin Carolan,
Lois Holloway, Emmanuel Rios Velazquez, Frank Hoebbers, Andre Dekker

Oncotarget. 2016 Jun 14; 7(24): 37288–37296.

Abstract

Background and Purpose:

To improve quality and personalization of oncology health care, decision aid tools are needed to advise physicians and patients. The aim of this work is to demonstrate the clinical relevance of a survival prediction model as a first step to multi institutional rapid learning and compare this to a clinical trial dataset.

Material and Methods:

Data extraction and mining tools were used to collect uncurated input parameters from Illawarra Cancer Care Centre's (clinical cohort) oncology information system. Prognosis categories previously established from the Maastricht Radiation Oncology (training cohort) dataset, were applied to the clinical cohort and the radiotherapy only arm of the RTOG-9111 (trial cohort).

Results:

Data mining identified 125 laryngeal carcinoma patients, ending up with 52 patients in the clinical cohort who were eligible to be evaluated by the model to predict 2-year survival and 177 for the trial cohort. The model was able to classify patients and predict survival in the clinical cohort, but for the trial cohort it failed to do so.

Conclusions:

The technical infrastructure and model is able to support the prognosis prediction of laryngeal carcinoma patients in a clinical cohort. The model does not perform well for the highly selective patient population in the trial cohort.

Keywords: larynx; survival prediction; rapid learning; model validation;

Introduction

Laryngeal carcinoma has a recorded incidence of 606 (2009) with 255 reported mortalities (2010) in Australia [1]. In Europe the recorded incidence (2012) was 39,900[2] and the USA reported an incidence of 12,630 (2014) with an estimated number of deaths of 3,610[3]. Treatment options for patients with early localized laryngeal carcinoma include surgery or radiotherapy, having equal outcome [4,5]. For advanced laryngeal cancer, surgery therapy using total laryngectomy has been the standard of care for decades [6], however nowadays also laryngeal preservation strategies using primary radiation or chemoradiation have been adopted [7,8].

New developments to further improve outcome in patients treated with radiotherapy, include the application of dose-escalation [9] and the use of more advanced radiotherapy technologies such as IMRT [10] and proton irradiation [11,12], to reduce side effects while maintaining local control.

Clinical Decision Support Systems (DSS) are a way to support the choice between the increasing number of radiotherapy techniques and technology options [13–15] both in terms of clinical benefit in the individual patient and in assigning resources to patient groups that benefit most from the new technology to address the concerns of keeping cancer care affordable [16]. To construct a DSS, predictive models need to be learned and validated.

Rapid learning health care is a way to learn predictive models. In rapid learning it is postulated that data routinely generated through patient care and clinical research feed into an ever-growing set of coordinated databases [17]. These coordinated datasets could then be used to learn and validate the model. In this study we present a first rapid learning approach that combines learning a predictive model from one clinical center (“training cohort”) and validating it in both another clinical center (“clinical cohort”) and a clinical trial dataset (“trial cohort”). A comprehensive technical infrastructure is proposed in which databases were coordinated spanning institutions and continents (Maastricht Radiation Oncology (MAASTRO) in Europe, the Radiation Therapy Oncology Group (RTOG) in North-America and Illawarra Cancer Care Centre (ICCC) in Australia).

The hypothesis of this study is that it is possible to implement an automated data extraction infrastructure for rapid learning that uses a model to predict survival in laryngeal carcinoma without any human evaluation of the data to show that routine clinical data is a valuable source of information that can be used to complement the current evidence base consisting mainly of clinical trial data. The

model was learned in one institution (MAASTRO) and applied in a patient care-driven regional cancer service (ICCC) and evaluated for a research-driven clinical trial collaborative group (RTOG).

Materials and methods

Clinical cohort

After internal review board approval, the data of laryngeal carcinoma patients was extracted from the Oncology Information System (OIS) of ICC (MOSAIQ, Elekta, Stockholm, Sweden) using a data integration tool (Kettle ,Pentaho Community Edition 5.0, Orlando, USA). To provide an automated infrastructure all model input parameters needed to be extracted from the OIS and stored in a data warehouse (MSSQL 2008). Patients were selected using the International Classification of Diseases (ICD) codes version 10. The ICD code for laryngeal carcinoma patients is C32 and all subcategories in this classification group. Patients treated with radiotherapy alone for a primary H&N disease were added to the clinical cohort data warehouse. These patients were diagnosed between April 1987 and February 2014. Over time more patients will be included because it is an automated system in which new patients can be included each time the software is executed. For patients with missing record and verify (R&V) data within the OIS (e.g. due to H&N diagnoses that were manually added to the treatment history for treatment they received elsewhere) an imputation algorithm was added to the data mining script: patients with a H&N diagnosis treated with radiotherapy before 2012 with no recorded delivered dose were assumed to be treated with the recorded prescribed dose. Each of the model parameters were extracted from the OIS with individual data integration programs. Quality assurance of the extracted data was undertaken via cross referencing with the OIS.

H&N predictive model

The data warehouse was queried and analyzed by a predictive model developed using Matlab 8.2.0 (The MathWorks Inc., Natick, MA, USA). The software applies the laryngeal carcinoma survival model [18] to the extracted data and reports the accuracy of the predictions created by the model. The model was fitted with Univariate Cox regression [18] which uses the following factors: age at the time of diagnosis, gender, T-stage, N-stage, hemoglobin level before treatment, tumor location and the biological equivalent dose in fractions of 2 Gray. These features

and other features were selected by a medical specialist to be analyzed by the Univariate Cox regression. In the original study [18] it was concluded that these features had a statistical relevance when predicting survival while others had not (e.g. the Tumor Volume computed from the PET scan). This resulted in a model with a baseline two-year survival of 0.1404. Table 1 shows the beta-coefficients of the model and the data input formatting that is used. The proportional hazards model resulting from this fit was implemented as a nomogram in the original study [18] to create an easy to use DSS for the physicians.

Table 1. Model coefficients together with the corresponding features and data format. The 0/1 input is the binary answer to the yes or no question of the model feature.

| Model Feature | Model Input | Model Beta |
|-------------------------------|-------------|------------|
| Age | number | 0.0454 |
| Gender is male? | 0/1 | 0.8715 |
| T2 classification | 0/1 | 0.1177 |
| T3 classification | 0/1 | 0.6795 |
| T4 classification | 0/1 | 1.2836 |
| N+ classification | 0/1 | 0.3623 |
| Tumor location is non-glottic | 0/1 | 0.2644 |
| Hemoglobin level | number* | -0.3190 |
| Total radiation dose | number | -0.0034 |

**in mmol/l*

For this study we implemented the original proportional hazard model as we are using a completely automated digital infrastructure. To evaluate the accuracy of the model the survival in months and the patient deceased status were also extracted from the OIS. To determine the survival of alive patients the last known registered contact within the OIS was used.

Model validation and statistics

A Receiver Operating Characteristic (ROC) curve computation module was used to compare the predicted survival with the actual survival of each patient population [18]. For the training cohort we applied an internal validation on the entire dataset. We compared internal validation to the external validations on the clinical and trial cohort. All validations result in an area-under-the-curve (AUC) that displays how well the model predicted the survival of the patients. An AUC of 0.5 indicates that the result is completely random meaning that the model is not able to predict outcome and an AUC of 1.0 indicates that the result is perfectly

matched meaning the model is a perfect outcome predictor. Bootstrapping was used to determine the uncertainty in the model's AUC reported by the program. Specifically, the AUC was determined a thousand times using the bootstrap function provided by the Matlab statistics toolbox. All cohorts were bootstrapped and in each bootstrap sample the model was applied to determine +/- 2 standard deviations of the AUC in all cohorts. Additionally, the predicted probability of survival was compared to the observed probability of survival for each prognosis group to assess the calibration of the model for each cohort.

Reference cohorts

To compare the effectiveness of the model in the clinical cohort we used the trial cohort (the RTOG-91-11 trial dataset [7]) and the training cohort (MAASTRO dataset [18]). With respect to the randomized RTOG-91-11 trial, only the patients treated with radiation only were selected (n = 177). The training and trial cohort were added to separate databases with the same data structure as the clinical cohort, this enabled the use of the same software for analyzing each cohort separately. An overview of the patient population of these datasets is given in Table 2. To perform a univariate survival analysis, the Kaplan Meier method was used. The prognosis groups were divided into 3 groups, classifying the 25% lowest survival predictions and 25% highest survival predictions as the poor and good prognosis group respectively. The middle 50% were classified as medium prognosis. The training cohort was used to create these thresholds for the poor, medium and good prognosis groups. These survival prediction thresholds dividing the training cohort were also applied to the clinical and trial cohort model outcomes. To compare the Kaplan Meier curves between cohorts and between prognosis groups the log-rank test was used. In all statistical tests p-values of less than 0.05 were assumed to indicate statistical significance.

Results

The data mining of the OIS resulted in an initial clinical cohort of 125 patients primarily diagnosed with laryngeal carcinoma. From this cohort, patients with missing data were then excluded; 13 patients because the diagnosis was not older than 2 years and thus it is impossible to assess 2-year survival for these patients; 3 patients due to a lack of treatment dose available in the OIS; 57 patients (the largest exclusion group) due to a lack of hemoglobin measurements before

treatment. This resulted in a clinical cohort containing 52 patients diagnosed between June 1993 and February 2012 with complete datasets suitable for analysis, at the time of modelling; in time the set will automatically grow.

The model predicted prognostic survival groups, resulted in the Kaplan Meier curves presented in Figure 1. The survival prediction thresholds to separate the poor, medium and good prognosis groups were 58% and 82% chance of 2-year survival. By definition these thresholds meant that the training cohort had 25%, 50% and 25% of patients in the poor, medium and good prognosis groups respectively. Applying the same thresholds to the clinical cohort gave a group distribution of 53%, 36% and 10% and for the trial cohort gave 55%, 41% and 4%. The Kaplan Meier curves of all groups in the clinical and trial cohort were compared with their equivalent in the training cohort. The survival prediction of each prognosis group in the clinical and trial cohort was not statistically different from the corresponding training cohort prognosis group ($p > 0.2$). The clinical cohort's poor and medium prognosis groups were statistically different ($p < 0.05$) but the medium and good prognosis groups were not ($p > 0.2$). The trial cohort comparison showed similar results.

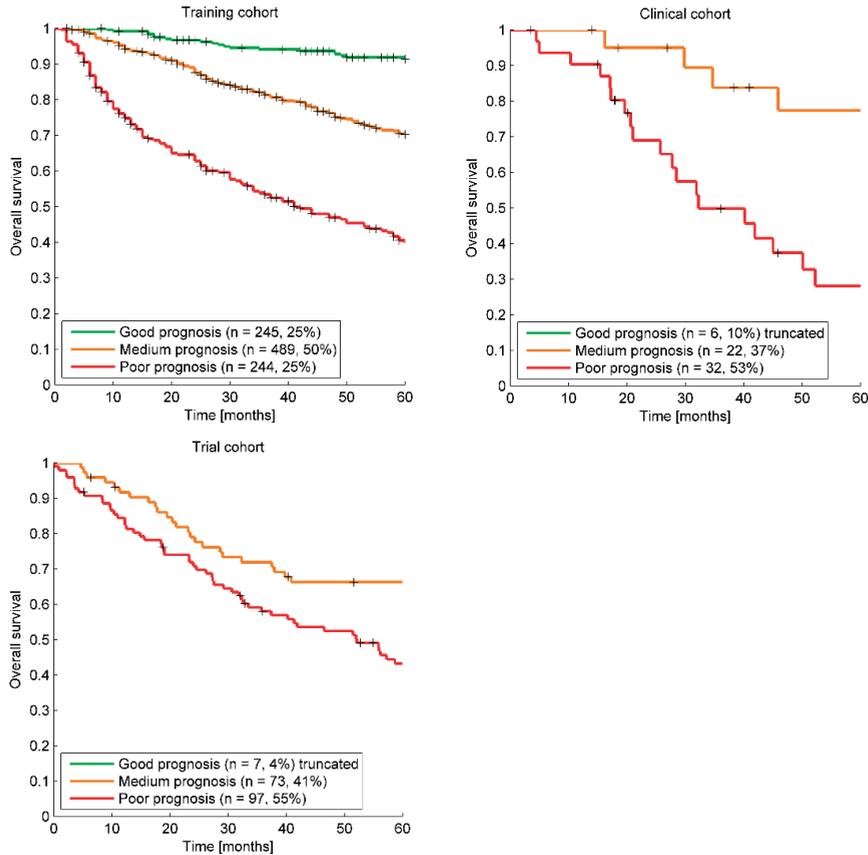


Figure 1. Kaplan Meier Curves for each cohort. The survival prediction thresholds to create the poor, medium and good prognosis groups were 57% and 81% chance of 2-year survival. This resulted in a group distribution of 53%, 36% and 10% and 53%, 42% and 5% for the poor, medium and good prognosis group for the clinical cohort and trial cohort respectively, while (by definition) the training cohort had 25%, 50% and 25% distribution.

The ROC computation resulted in AUC values of 0.77, 0.71 and 0.57 for training, clinical and trial cohort respectively. Bootstrapping (1000 samples) resulted in normally distributed AUC reliability intervals (+/-2SD) of 0.73 to 0.81, 0.55 to 0.88 and 0.47 to 0.67 for training, clinical and trial cohort respectively. The model calibration plots are presented in Figure 2. For the training cohort the observed 2-year survival is higher than predicted for the poor and medium prognosis group. The same can be concluded for the clinical cohort as the difference in survival for each prognosis group did not reach statistical significance ($p > 0.2$) between the training and clinical cohort. No statistical difference could be found between the prognosis

groups of the trial cohort ($p > 0.2$) and the poor and good prognosis group survival was different from the training cohort ($p < 0.05$).

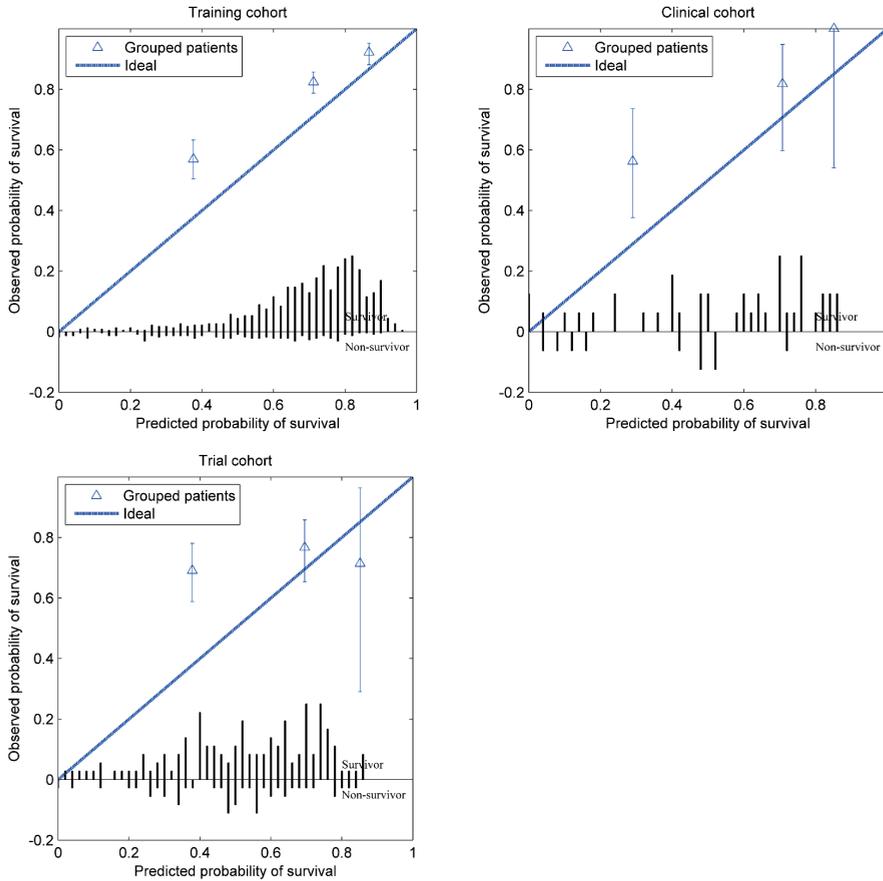


Figure 2. Calibration curve for each cohort showing the observed survival in relation to the predicted survival for the poor, medium and good prognosis groups. The bar graph shows the survivor and non-survivor distribution per predicted survival probability bin.

Discussion

We have implemented a survival prediction model for laryngeal carcinoma patients treated with primary radiation in a clinical cohort from a completely different geographical area (Australia vs The Netherlands) and evaluated the same model in a trial cohort from North America. In previous work this model was validated in other curated independent datasets from the Leuven Cancer

Institute (Belgium), VU University Medical Center (Netherlands), Netherlands Cancer Institute-Antoni van Leeuwenhoek Hospital (Netherlands) and Christie Hospital (UK)[18]. The AUCs reported were 0.68, 0.74, 0.71 and 0.76 for each mentioned group respectively. The uncurated clinical cohort had a comparable accuracy (AUC of 0.71). The model is able to predict 2-year survival for laryngeal cancer patients for the clinical cohort as the AUC is statistically different from 0.5. The AUC reliability interval was larger than observed for the training cohort which can be explained by the smaller size of the clinical cohort. Comparable results were shown for the Leuven dataset (AUC 0.50 – 0.82) which is similar in size ($n = 52$). Another important fact is that the applied model was learned from the training cohort so it will by definition perform better on this cohort. The model calibration plots (Figure 2) shows that the model is not perfectly calibrated for 2-year survival, underestimating survival especially in the poor prognosis group. The likely reason is that this proportional hazard model was trained in the original study for 5-year survival prediction and provides a baseline survival for all time points between 0 and 5 years. It is not uncommon for these types of models to be recalibrated after acquiring more data to improve the survival probability [19]. The observed survival in the clinical cohort is not statistically different from the observed survival in the training cohort as reported earlier in the Results section demonstrating that the model performance in the clinical cohort is comparable to the training cohort. The prognosis distribution of the clinical cohort is shifted towards the poor prognosis group; the main reason is the difference in patient population. As shown in Table 2 the clinical cohort patients have more T3 and T4 cancers, more often have N1 and N2 disease, more non-glottic cancers, receive a lower treatment dose and have lower hemoglobin levels. These are all unfavorable predictors for survival in the prediction model. The more advanced cancers and nodal metastasis might be explained by the socioeconomic difference between the Illawarra and Maastricht regions, as patients are referred to the ICCC at a later stage or wait longer to consult their physician. However, the observed survival is not statistically different from the training cohort for each prognosis group as reported in the results. This indicates that the training and clinical cohort are similar and that similar features seem to be predictive for survival. The inclusion of the clinical cohort patients in a future training cohort may find additional features specific to the poorer prognosis group, which is the subject of future work.

Table 2. Patient population model input parameter values. The clinical and trial cohort are compared to the training cohort to indicate a statistical difference.

| | | Training Cohort | | Clinical Cohort | | Trial Cohort | | Training VS | |
|----------------------|-------------|-----------------|----|-----------------|----|--------------|----|-------------|----------|
| | | # | % | # | % | # | % | Clinical | Trial |
| Total | | 978 | | 52 | | 177 | | | |
| Age | 47-60 years | 357 | 37 | 15 | 29 | 96 | 54 | p > 0.20 | p < 0.05 |
| | > 60 years | 621 | 63 | 37 | 71 | 81 | 46 | p > 0.20 | p < 0.05 |
| Gender | Male | 870 | 89 | 47 | 90 | 136 | 77 | p > 0.20 | p < 0.05 |
| | Female | 108 | 11 | 5 | 10 | 41 | 23 | p > 0.20 | p < 0.05 |
| T-classification | T1 | 524 | 54 | 18 | 35 | 0 | 0 | p < 0.05 | p < 0.05 |
| | T2 | 260 | 27 | 11 | 21 | 18 | 10 | p > 0.20 | p < 0.05 |
| | T3 | 128 | 13 | 14 | 27 | 144 | 81 | p < 0.05 | p < 0.05 |
| | T4 | 66 | 7 | 7 | 13 | 15 | 8 | p = 0.07 | p > 0.20 |
| | Missing | 0 | 0 | 2 | 4 | 0 | 0 | p < 0.05 | p > 0.20 |
| N-classification | N0 | 884 | 90 | 41 | 79 | 92 | 52 | p < 0.05 | p < 0.05 |
| | N+ | 98 | 10 | 11 | 21 | 85 | 48 | p < 0.05 | p < 0.05 |
| | Missing | 2 | 0 | 0 | 0 | 0 | 0 | p > 0.20 | p > 0.20 |
| Tumor location | Glottic | 723 | 74 | 27 | 52 | 49 | 28 | p < 0.05 | p < 0.05 |
| | Non-Glottic | 255 | 26 | 25 | 48 | 128 | 72 | p < 0.05 | p < 0.05 |
| Haemoglobin level | Low1 | 168 | 17 | 24 | 46 | 58 | 33 | p < 0.05 | p < 0.05 |
| | Normal-high | 667 | 68 | 28 | 54 | 116 | 66 | p < 0.05 | p > 0.20 |
| | Missing | 0 | 0 | 0 | 0 | 3 | 2 | p > 0.20 | p < 0.05 |
| Total radiation dose | <60Gy | 16 | 2 | 11 | 21 | 5 | 3 | p < 0.05 | p > 0.20 |
| | 60-66Gy | 437 | 45 | 22 | 42 | 1 | 1 | p > 0.20 | p < 0.05 |
| | >66Gy | 541 | 55 | 19 | 37 | 171 | 97 | p < 0.05 | p < 0.05 |

¹Male < 8.5 mmol/l, Female < 7.5 mmol/l)

Uncurated clinical data has been demonstrated to be sufficient to produce and validate useful models and DSS, however the work also indicates that prospective consistent data recording can improve opportunities to learn from clinical data. Increasing the numbers of patient records eligible to be entered into the modelling process can enable the addition of more model parameters and strengthen model performance. For this study the data quality was very high in comparison to similar studies[20] where less than 5 percent of the treated patient records were usable after data mining while in this study over 30 percent of the

treated patient records have been included. This can be explained by a previous retrospective study in this patient group. The original data for these patients was complimented with great detail, something that is not standard in a radiotherapy clinic.

The largest gap in data was caused by the poorly recorded hemoglobin level measurements before the start of radiotherapy. Because hemoglobin level is one of the input parameters with the strongest weight in the model a separate analysis was undertaken where the hemoglobin level was imputed with a low (7.0mmol/l), high (11.0mmol/l) and the training set median (9.1mmol/l) value. This imputation resulted in an enlarged clinical cohort of 109 patients. This resulted in an AUC interval increase of 0.60 to 0.83, 0.61 to 0.84, and 0.62 to 0.85 for the mentioned imputations respectively, intervals which are somewhat tighter than for the non-imputed clinical cohort which is likely caused by the increase in patient numbers. The different imputation methods for the hemoglobin level resulted in a very different distribution of the patients across survival groups as shown in the Kaplan Meier curves (Figure 3). This shows that the model is very sensitive to the hemoglobin level and that one has to take great care in choosing an appropriate imputation method but it also shows that the clinical cohort contains more information that could be utilized. In future work smarter ways of imputing missing values could be explored. An example of a smarter solution is a Bayesian Network model where for example the hemoglobin level could be derived by considering all other patient properties available instead of simple median calculation [21].

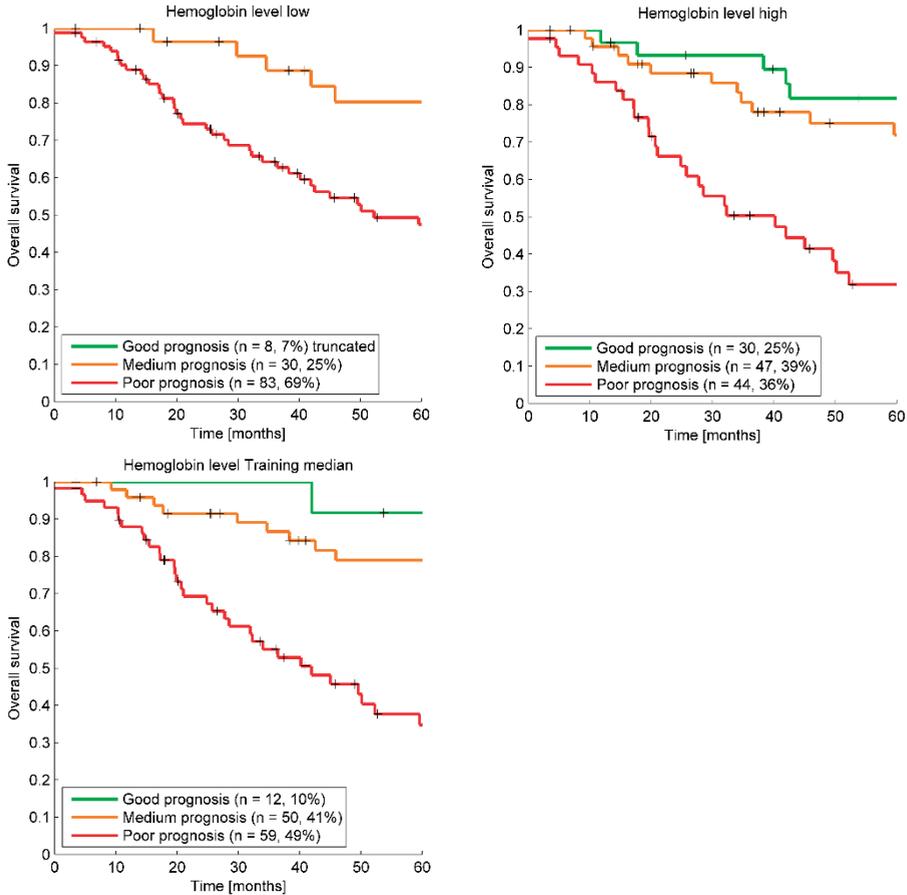


Figure 3. Kaplan Meier survival curves clinical cohort using a low, high, and training median imputation value demonstrating the effect of assuming hemoglobin values for the 57 patients in the clinical cohort that were missing a hemoglobin level measurement before treatment.

In both the previous study[18] and in this study the model is able to predict survival for the clinical cohorts. However, in this study we show that the model is not able to predict 2-year survival for laryngeal cancer patients for the trial cohort: The AUC shows that the prediction is not different from random and the calibration plot shows that the observed survival is statistically the same for each prognosis group in the trial cohort. The comparison between the trial cohort prognosis groups as reported in the results confirms this. The most likely cause of this poor performance and calibration is that the trial patient population is different from the training and clinical cohort. Table 2 shows cohort properties which are statistically different between training and trial cohort. The trial cohort

consists of younger patients, more females, almost exclusively T3 staged cancers, more nodal involvement, more non-glottic cancers and almost exclusively high radiation dosages. This patient-type is under represented in the training cohort. Although data quality is extremely high in a clinical trial, the patients in the trial cohort are highly selected being relatively young patients with advanced cancers that were treated perhaps with a higher quality level or with different treatments than standard practice or that were different by another unknown confounding factor. After training new model coefficients on the trial cohort and using the same data for validation the AUC was only 0.58. Even when using an optimistic overestimating validation, the model performs poorly, this means that the data does not contain the knowledge we need to predict survival for this specific cohort supporting our earlier statement that there might be another unknown factor that is of great influence on the survival for these patients. This difference between clinical routine and trial cohorts is one of the arguments against using solely evidence from clinical trials as the source of clinical guidelines in radiation oncology [22,23]. For rapid learning, if a training cohort is different from a validation cohort and/or the patients in the validation cohort are underrepresented in the training cohort, a poor performance of the model can be expected. To increase model performance more patients with different characteristics should be included in the training set during model learning. Including trial patients can have a negative effect on model performance; it has been reported that it can result in a biased model towards this population. As an example, the predicted survival (i.e. the calibration of the model) will be much higher than can be obtained in routine clinical practice as routine quality is expected to be lower than in clinical trials, which is known to affect survival [24]. The poor performance in the trial cohort also underlines the need of model commissioning. Before using any prediction model, it is important to verify if this model does indeed perform well in a specific population. Commissioning of hardware and software is a well-known process in radiotherapy and decision aid tools should undergo the same quality assurance procedures. A better modelling approach would be multi center rapid learning systems that can enable the model learning algorithm to learn from data present in multiple centers, using different machine learning approaches, such as Bayesian Networks, which can explicitly account for biased datasets. This could include the integration of large scale observational studies such as DAHANCA [25] and the integration of clinical routine and clinical trial cohorts as suggested by others[13,26]. In ongoing work on the rapid learning infrastructure, we use Semantic Web technology and ontologies such as the

Radiation Oncology Ontology [<https://bioportal.bioontology.org/ontologies/ROO>] to create well defined semantically interoperable data stores and secure messaging systems to facilitate multi institutional rapid learning.

In clinical practice TNM staging is used to estimate the prognosis for larynx patients and there is some evidence of additional single variable prognosis predictors[27,28]. To our knowledge there are no models to predict survival in laryngeal cancer patients. Some survival prediction models exist for other head and neck cancer patients [29,30]. The study [30] with a similar approach, large training and external validation cohorts found comparable results in model performance.

Conclusion

We were able to use a data mining system to automate the collection of model parameters in a totally different clinical cohort in a different country and healthcare system and predict 2-year survival for their patient population using uncurated clinical data. The study shows that routine clinical data contains valuable information that could be harvested to improve and personalize patient care and even more so if recorded in a detailed, structured manner. The results demonstrate that further investigations into the difference between clinical trial cohorts and clinical cohorts is necessary with the potential for rapid learning systems to provide evidence for patients who do not fit clinical trial criteria.

Acknowledgements

We would like to thank the NRG/RTOG for making the data of the RTOG-91-11 trial available; and the Illawarra Cancer Care Centre for funding Tim Lustberg's visit to Australia, within the NSW/University of Sydney datamining collaborative network project.

References

1. Australian Institute of Health and Welfare & Australasian Association of Cancer Registries 2012. Cancer in Australia an overview 2012. *Canberra: Australian Institute of Health and Welfare & Australasian Association of Cancer Registries*
2. UK CR. Laryngeal (larynx) cancer incidence statistics 2014. <http://www.cancerresearchuk.org/cancer-info/cancerstats/types/larynx/incidence/#In> [accessed April 16, 2015]
3. Howlader N, Noone A, Krapcho M, Grashell J, Miller D, Altekruse S, *et al.* SEER Cancer Statistics Review. *SEER Cancer Statistics Review* 2014. <http://seer.cancer.gov/statfacts/html/larynx.html> based on November 2013 SEER data submission, posted to the SEER web site, April 2014 [accessed April 16, 2015]
4. Hartl DM, Ferlito A, Brasnu DF, Langendijk JA, Rinaldo A, Silver CE, *et al.* Evidence-based review of treatment options for patients with glottic cancer. *Head Neck* 2011;33(11):1638–1648. doi:10.1002/hed.21528
5. Remmelts AJ, Hoebbers FJP, Klop WMC, Balm AJM, Hamming-Vrieze O, Brekel MWM van den. Evaluation of lasersurgery and radiotherapy as treatment modalities in early stage laryngeal carcinoma: tumour outcome and quality of voice. *Eur Arch Otorhinolaryngol* 2013;270(7): 2079–2087. doi:10.1007/s00405-013-2460-x
6. Wolf GT, Fisher SG, Hong WK, *et al.* Induction chemotherapy plus radiation compared with surgery plus radiation in patients with advanced laryngeal cancer. The Department of Veterans Affairs Laryngeal Cancer Study Group. *N Engl J Med* 1991;324(24):1685–1690. doi:10.1056/NEJM199106133242402
7. Forastiere AA, Goepfert H, Maor M, Pajak TF, Weber R, Morrison W, *et al.* Concurrent chemotherapy and radiotherapy for organ preservation in advanced laryngeal cancer. *N Engl J Med* 2003;349(22):2091–2098. doi:10.1056/NEJMoa031317
8. Forastiere AA, Zhang Q, Weber RS, Maor MH, Goepfert H, Pajak TF, *et al.* Long-term results of RTOG 91-11: a comparison of three nonsurgical treatment strategies to preserve the larynx in patients with locally advanced larynx cancer. *J Clin Oncol* 2013;31(7):845–852. doi:10.1200/JCO.2012.43.6097
9. Leclerc M, Maingon P, Hamoir M, Dalban C, Calais G, Nuyts S, *et al.* A dose escalation study with intensity modulated radiation therapy (IMRT) in T2N0, T2N1, T3N0 squamous cell carcinomas (SCC) of the oropharynx, larynx and hypopharynx using a simultaneous integrated boost (SIB) approach. *Radiother Oncol* 2013;106(3):333–340. doi:10.1016/j.radonc.2013.03.002
10. Nutting CM, Morden JP, Harrington KJ, Urbano TG, Bhide SA, Clark C, *et al.* Parotid-sparing intensity modulated versus conventional radiotherapy in head and neck cancer (PARSPORT): a phase 3 multicentre randomised controlled trial. *Lancet Oncol* 2011;12(2):127–136. doi:10.1016/S1470-2045(10)70290-4
11. Ramaekers BLT, Grutters JPC, Pijls-Johannesma M, Lambin P, Joore MA, Langendijk JA. Protons in head-and-neck cancer: bridging the gap of evidence. *Int J Radiat Oncol Biol Phys* 2013;85(5):1282–1288. doi:10.1016/j.ijrobp.2012.11.006
12. van der Laan HP, van de Water TA, van Herpt HE, Christianen MEMC, Bijl HP, Korevaar EW, *et al.* The potential of intensity-modulated proton radiotherapy to reduce swallowing dysfunction in the treatment of head and neck cancer: A planning comparative study. *Acta Oncol* 2013;52(3):561–569. doi:10.3109/0284186X.2012.692885

13. Lambin P, Roelofs E, Reymen B, Velazquez ER, Buijsen J, Zegers CML, *et al.* "Rapid Learning health care in oncology" - an approach towards decision support systems enabling customised radiotherapy'. *Radiother Oncol* 2013;109(1):159–164. doi:10.1016/j.radonc.2013.07.007
14. Lambin P, van Stiphout RGPM, Starmans MHW, Rios-Velazquez E, Nalbantov G, Aerts HJWL, *et al.* Predicting outcomes in radiation oncology--multifactorial decision support systems. *Nat Rev Clin Oncol* 2013;10(1):27–40. doi:10.1038/nrclinonc.2012.196
15. van der Laan HP, van de Water TA, van Herpt HE, Christianen MEMC, Bijl HP, Korevaar EW, *et al.* The potential of intensity-modulated proton radiotherapy to reduce swallowing dysfunction in the treatment of head and neck cancer: A planning comparative study. *Acta Oncol* 2013;52(3):561–569. doi:10.3109/0284186X.2012.692885
16. Sullivan R, Peppercorn J, Sikora K, Zalcborg J, Meropol NJ, Amir E, *et al.* Delivering affordable cancer care in high-income countries. *Lancet Oncol* 2011;12(10):933–980. doi:10.1016/S1470-2045(11)70141-3
17. Abernethy AP, Etheredge LM, Ganz PA, Wallace P, German RR, Neti C, *et al.* Rapid-learning system for cancer care. *J Clin Oncol* 2010;28(27):4268–4274. doi:10.1200/JCO.2010.28.5478
18. Egelmeier AGTM, Velazquez ER, de Jong JMA, Oberije C, Geussens Y, Nuyts S, *et al.* Development and validation of a nomogram for prediction of survival and local control in laryngeal carcinoma patients treated with radiotherapy alone: a cohort study based on 994 patients. *Radiother Oncol* 2011;100(1):108–115. doi:10.1016/j.radonc.2011.06.023
19. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. *BMC Medical Research Methodology* 2013;13(1):33. doi:10.1186/1471-2288-13-33
20. Dekker A, Vinod S, Holloway L, George A, Goozee G, Delaney G, *et al.* Rapid learning in practice: A lung cancer survival decision support system in routine patient care data. *Radiation Oncology* 2014;111(Suppl1):35
21. Jayasurya K, Fung G, Yu S, Dehing-Oberije C, De Ruyscher D, Hope A, *et al.* Comparison of Bayesian network and support vector machine models for two-year survival prediction in lung cancer patients treated with radiotherapy. *Med Phys* 2010;37(4):1401–1407
22. Grand MM, O'Brien PC. Obstacles to participation in randomised cancer clinical trials: a systematic review of the literature. *J Med Imaging Radiat Oncol* 2012;56(1):31–39. doi:10.1111/j.1754-9485.2011.02337.x
23. Dekker ALAJ, Gulliford SL, Ebert MA, Orton CG. Point/Counterpoint. Future radiotherapy practice will be based on evidence from retrospective interrogation of linked clinical data sources rather than prospective randomized controlled clinical trials. *Med Phys* 2014;41(3):030601. doi:10.1118/1.4832139
24. Fairchild A, Straube W, Laurie F, Followill D. Does Quality of Radiation Therapy Predict Outcomes of Multicenter Cooperative Group Trials? A Literature Review. *International Journal of Radiation Oncology • Biology • Physics* 2013;87(2):246–260. doi:10.1016/j.ijrobp.2013.03.036
25. Primdahl H, Nielsen AL, Larsen S, Andersen E, Ipsen M, Lajer C, *et al.* Changes from 1992 to 2002 in the pretreatment delay for patients with squamous cell carcinoma of larynx or pharynx: a Danish nationwide survey from DAHANCA. *Acta Oncol* 2006;45(2):156–161. doi:10.1080/02841860500423948
26. Murthy VH, Krumholz HM, Gross CP. Participation in cancer clinical trials: race-, sex-, and age-based disparities. *JAMA* 2004;291(22):2720–2726. doi:10.1001/jama.291.22.2720
27. Sachse F, Stoll W, Rudack C. Evaluation of treatment results with regard to initial anterior commissure involvement in early glottic carcinoma treated by external partial surgery or transoral laser microresection. *Head Neck* 2009;31(4):531–537. doi:10.1002/hed.20997

Chapter 6

28. Hermans R, Van den Bogaert W, Rijnders A, Baert AL. Value of computed tomography as outcome predictor of supraglottic squamous cell carcinoma treated by definitive radiation therapy. *International Journal of Radiation Oncology*Biophysics* 1999;44(4):755–765. doi:10.1016/S0360-3016(99)00039-5
29. Cho JK, Lee GJ, Yi KI, Cho KS, Choi N, Kim JS, *et al.* Development and external validation of nomograms predictive of response to radiation therapy and overall survival in nasopharyngeal cancer patients. *European Journal of Cancer* 2015;51(10):1303–1311. doi:10.1016/j.ejca.2015.04.003
30. Tang LQ, Li CF, Li J, Chen WH, Chen QY, Yuan LX, *et al.* Establishment and Validation of Prognostic Nomograms for Endemic Nasopharyngeal Carcinoma. *J Natl Cancer Inst* 2016;108(1). doi:10.1093/jnci/djv291

Chapter 7

Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer

Tim Lustberg, Johan van Soest, Mark Gooding, Devis Peressutti, Paul Aljabar,
Judith van der Stoep, Wouter van Elmpt, Andre Dekker

Radiother Oncol. 2018 Feb;126(2):312-317

Abstract

Background and Purpose:

Contouring of organs at risk (OARs) is an important but time consuming part of radiotherapy treatment planning. The aim of this study was to investigate whether using institutional created software-generated contouring will save time if used as a starting point for manual OAR contouring for lung cancer patients.

Material and Methods:

Twenty CT scans of stage I-III NSCLC patients were used to compare user adjusted contours after an atlas-based and deep learning contour, against manual delineation. The lungs, esophagus, spinal cord, heart and mediastinum were contoured for this study. The time to perform the manual tasks was recorded.

Results:

With a median time of 20 minutes for manual contouring, the total median time saved was 7.8 minutes when using atlas-based contouring and 10 minutes for deep learning contouring. Both atlas based and deep learning adjustment times were significantly lower than manual contouring time for all OARs except for the left lung and esophagus of the atlas based contouring.

Conclusions:

User adjustment of software generated contours is a viable strategy to reduce contouring time of OARs for lung radiotherapy while conforming to local clinical standards. In addition, deep learning contouring shows promising results compared to existing solutions.

Key words: lung cancer; organs at risk; radiotherapy; atlas contouring; deep learning contouring;

Introduction

The contouring of organs at risk (OAR) and target volumes is an important aspect of treatment planning in radiation oncology. This process is time consuming and the quality of contours depends on the skill level of the observer [1]. Automatic contouring software could potentially speed up the process and improve consistency between observers. There are a number of commercially available products but these are not frequently used in clinical practice [2].

In the last two decades a lot of effort was put into learning new ways to recognize structures in a range of different imaging modalities (CT, PET, and MRI). Approaches range from knowledge-based algorithms such as atlas-based contouring, machine learning and statistical shape and appearance models; region-based methods such as adaptive thresholding, graph cuts and watershed contouring; or a combination of the knowledge- and region-based methods [2]. In products commercially available in 2014, all vendors use a form of atlas-based contouring and approximately half complement this with a model-based method but these are generally limited to certain OARs [2]. Recently, machine learning techniques, and deep learning methods in particular, have become popular for a wider range of tasks. These approaches, based on artificial neural networks, have shown outstanding capabilities, outperforming most classification and regression methods to date. The main advantage of deep learning methods is the ability to automatically learn the most suitable data representation for the task at hand.

The clinical applicability of automatic contouring software is well-reported for regions such as head and neck, breast, and abdomen [2–6]. For the lung, there are a number of studies reporting automatic contouring [7–14]. Some focus solely on a single method and OAR or the Gross Tumor Volume (GTV) and only evaluate the accuracy of the method without any clinical implications such as time gain. Two studies address the usability of atlas-based contouring for the thorax OARs [9,14]. Dolz et al provides a framework for a region-based contouring technique in clinical practice [13], they advise further investigation into more accurate atlas selection methods to improve the clinical usability. All these methods should be followed by manual correction of the imperfections of the software contouring with the techniques currently available to assess the clinical usability of these methods. In this study, we hypothesize that using a software-generated contour created with an institution specific model as a starting point for OAR contouring will reduce contouring time for lung cancer patients in a clinical

setting. For this we evaluate two contouring methods, atlas-based contouring and one novel method based on deep learning in a clinical representative scenario.

Materials and methods

Atlas-based contours

A commercial atlas-based contouring software (Mirada RTx 1.6 and Workflow Box 1.4, Mirada Medical Ltd., Oxford, United Kingdom) was used to automatically generate contours of the OARs. The atlas-based contouring employed 20 stage I NSCLC patients collected from clinical practice at our institute with minimal geometric distortions and small lesion volumes. These atlas patients were contoured by a senior radiotherapy technician specialized in the thorax region using institutional guidelines and carefully inspected by radiation oncologist for correctness.

Deep Learning contours

A prototype of deep learning contouring software (“Mirada DLC Expert” prototype, Mirada Medical Ltd., Oxford, United Kingdom) was used to create contours of the OARs. Both automated contouring methods were performed on a standard desktop computer, additionally the DLC used a graphics card to perform the calculations. The prototype uses a deep learning model based on convolutional neural networks [15], a sub-class of deep learning techniques tailored to process imaging data. Convolutional neural networks employ models with a large number of degrees of freedom (in the order of millions), and are therefore able to learn complex non-linear relationships within the imaging data. Training such models requires a large amount of imaging data and uses a backpropagation algorithm based on a stochastic gradient descent to optimize the free parameters [15]. Contours of 450 lung patients were collected from clinical practice and used to train the model.

Patient selection

Twenty consecutive stage I-III NSCLC patients treated in the period January to February 2016 were selected from routine clinical practice and the mid-ventilation phase of a 4D CT scan (Siemens Biograph PET/CT or Sensation Open CT

scanner) was used to contour the OARs. The OARs were defined by institutional guidelines and comprised the left lung, right lung, heart, spinal cord, esophagus, and mediastinum.

Contour methods

For this study we created 5 contour sets: A manual contour (MC); an atlas-based contour (AC); a user adjustment of the atlas-based contour (UAC), meaning the atlas-based contour was used as a starting point and adjustments by a radiotherapy technician were allowed to meet institutional guidelines; a deep learning contour (DLC); and a user adjustment of the deep learning contour (UDLC). A single radiotherapy technician performed contouring tasks to prevent any inter-observer variability. Manual contouring tasks were performed using the software used in clinical practice (Eclipse, version 11.0, Varian, Palo Alto, United States of America). For each of the contouring tasks, the time required for completion was recorded per patient and OAR.

Subjective scoring of contours

The software generated contours were subjectively scored by the technician from one to four. (1) None of the results would form a useful basis for further editing, no time is expected to be saved contouring is expected compared to manual contouring; (2) Some of the results form a useful basis for further editing, little time would be saved contouring is expected compared to manual contouring; (3) Many of the results form a useful basis for further editing, a moderate time saving is expected compared to manual contouring; (4) Most of the results form a useful basis for further editing, a significant time saving is expected compared to manual contouring.

Contour consistency measurement

All contours were exported and analyzed in Matlab 8.6 (The MathWorks Inc., Natick, MA, USA). To assess similarity, the manual contour was compared to the software generated and user adjusted contours using several metrics. To quantify the similarity between the different contour sets, the contours were projected onto a three-dimensional grid matching the dimensions of the corresponding CT scan. The Dice index, which is the volume of the union normalized by the mean of the two volumes, of each OAR in comparison with the manual contour was calculated. The distance between the surfaces of each contour was measured

using a nearest neighbor Euclidean distance calculation, the maximum nearest neighbor Euclidean distance, i.e. the Hausdorff distance, were compared. To evaluate differences in the results of the Dice index, Hausdorff distances and contouring time, a ranked Wilcoxon test was performed; p-values smaller than 0.05 were assumed to be statistically significant.

Results

Contouring time and consistency

The total median time saved was 7.8 minutes [range -2.2 to 13 minutes] and 10 minutes [range 5.2 - 15 minutes] for the UAC and UDLC respectively with respect to the MC. This is a large reduction compared to the median time required to contour all OARs for a lung case, which was 20 minutes. An overview of the recorded times per OAR is given in Figure 1. Both lungs and the spinal cord show significant time reductions for the UAC and UDLC ($p < 0.05$), except for the left lung UAC. This was the result of one large outlier in the dataset, where the time needed to contour was 0.9 minutes and the time to adjust the AC was 3.6 minutes. The AC was incorrect due to the fact the lung had collapsed (Figure 2A). Repeating the ranked Wilcoxon test without this sample resulted in a significant time gain for the left lung ($p = 0.014$). The DLC remained robust in the case where the lung had collapsed (Figure 2A). The UDLC performs best with a median time to evaluate and correct under one minute for the lungs and spinal cord. Comparing the Dice and Hausdorff distance for the DLC and UDLC to MC results in no significant differences ($p > 0.05$), this supports the assumption that the time needed to adjust these OARs represents the time the technician needs to decide the contour meets clinical guidelines and can be used without major adjustments. An overview of the Dice scores and Hausdorff distances can be found in Figure 3 and 4.

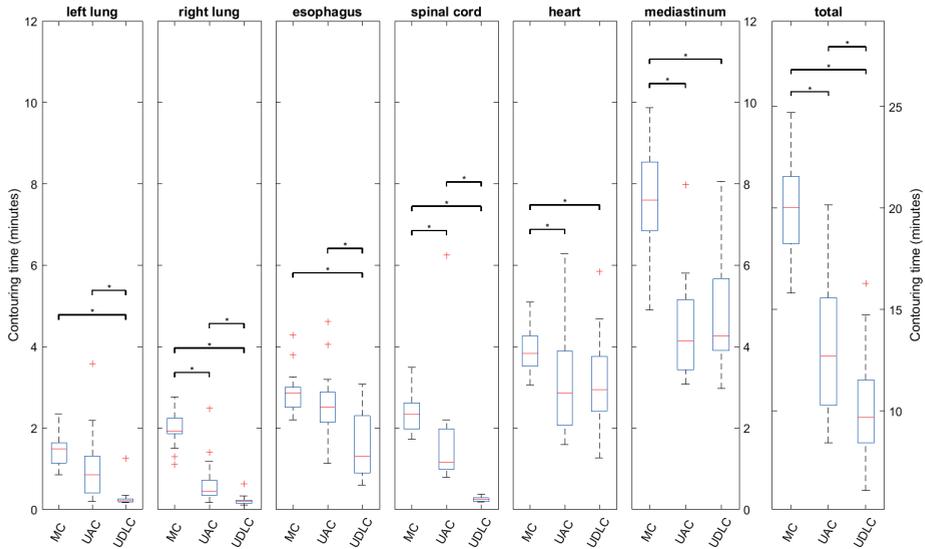


Figure 1. Contouring time of the manual contour (MC) user adjusted atlas-based contour (UAC) and user adjusted deep learning contour (UDLC) displayed for each OAR and the total time of all OARs. * indicates significant difference between manual and adjustment timing ($p < 0.05$, ranked Wilcoxon test).

The UAC of the esophagus did not result in a significant time reduction with a median time saving of 0.3 minutes [range -1.9 to 1.9]. By contrast, the UDLC resulted in a significant time reduction of 1.5 minutes [range 0 - 3.2]. Figure 2B shows an example slice of the esophagus contour where the DLC matches the MC but the AC is different. The heart and mediastinum showed a significant time reductions for both UAC and UDLC when compared to the MC, however, no significant difference was observed using UDLC over UAC ($p = 0.65$ and $p = 0.32$ for the OARs respectively). Figure 2C shows an example of the heart, where the AC matches the MC but the DLC is different. The mediastinum had a median time reduction 3.4 minutes [range 0.6 - 5.7], which is the largest contribution to the overall time reduction. Comparing the AC and DLC to the UAC and UDLC for the esophagus, heart and mediastinum resulted in a significant different time required when compared to the MC ($p < 0.05$).

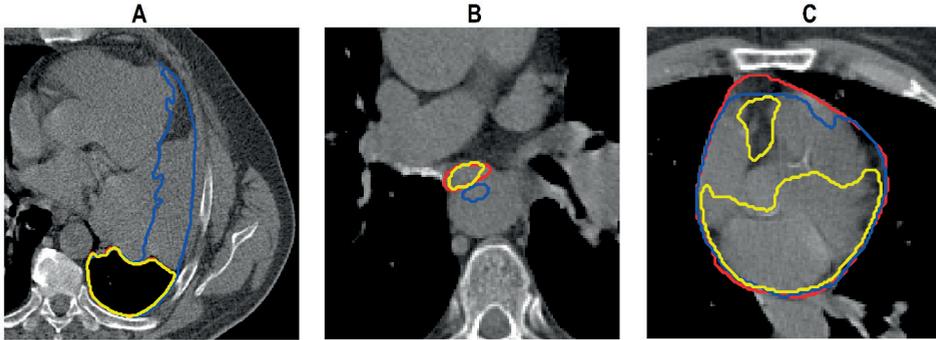


Figure 2. Example cases showing manual contour (MC, red), atlas-based contour (AC, blue) and the deep learning contour (DLC, yellow) for left lung (A) the esophagus (B) and heart (C).

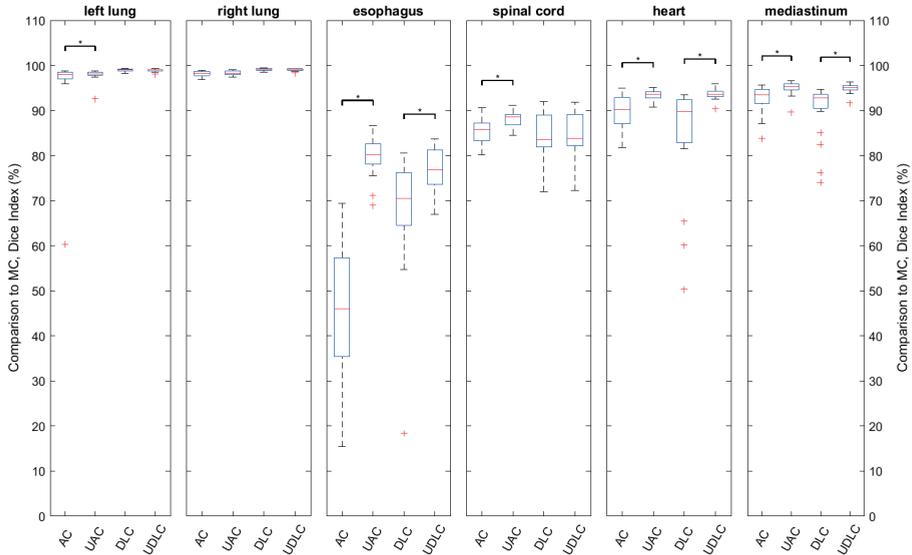


Figure 3. Dice scores comparing the atlas contour (AC), user adjusted atlas contour (UAC), deep learning contour (DLC) and user adjusted deep learning contour (UDLC) to the manual delineation MC displayed for each OAR. *indicates significant difference between a software generated contour and user adjusted contour ($p < 0.05$, ranked Wilcoxon test)

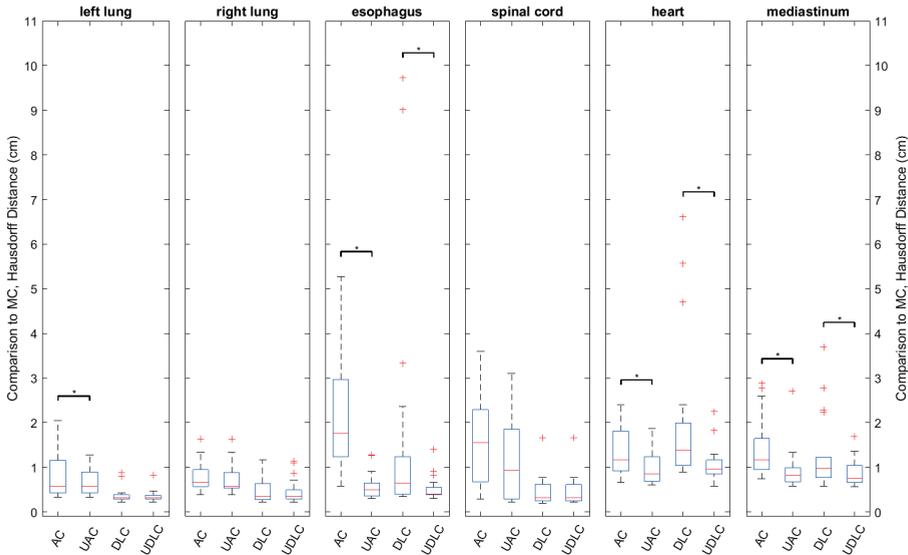


Figure 4. Hausdorff distance comparing the atlas contour (AC), user adjusted atlas contour (UAC), deep learning contour (DLC) and user adjusted deep learning contour (UDLC) to the manual contour MC displayed for each OAR. *indicates significant difference between a software generated and user adjusted contour ($p < 0.05$, ranked Wilcoxon test).

Subjective scoring of contours

For the lungs and spinal cord, AC performed well according to the subjective score of the technician, with a median score of 4 [range 2 - 4]. DLC performed even better for these OARs, where every contour was scored as 4. The esophagus AC performed poorly according to the technician with all contours having a score of 1. The DLC was perceived to be an improvement in most cases with a median score of 3 [range 1 - 4]. The AC of the heart and mediastinum were perceived to be slightly better than the DLC with respective median scores of 3 [range 2 - 3] and 3 [range 1 - 3]. The OARs with a median subjective score of 4 were consistent (median Dice score >90%, median Hausdorff distance <1.5 cm) when compared against the MC, however the spinal cord performed slightly worse (median Dice score 83%, median Hausdorff distance 1.6 cm).

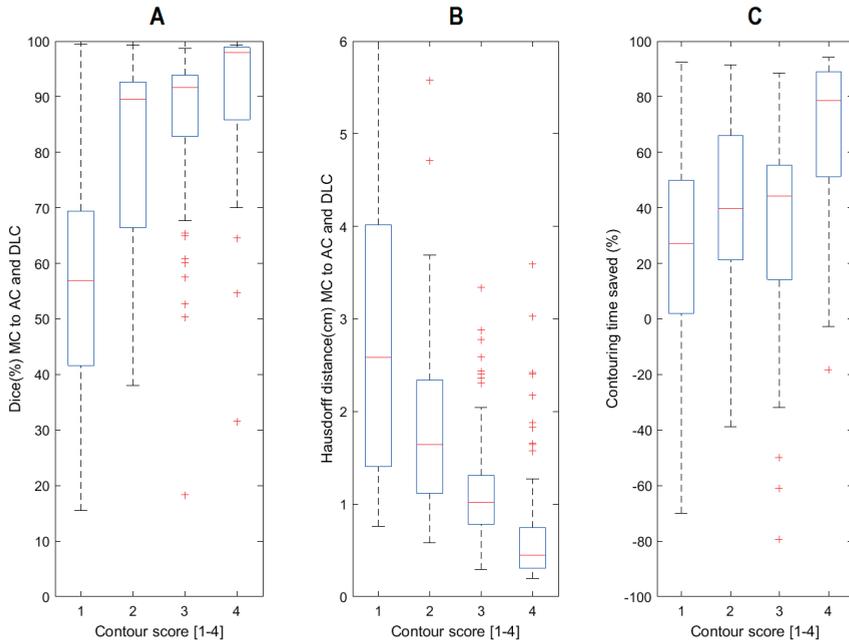


Figure 5. The comparison results of the manual contour (MC) and the atlas based contour (AC) and deep learning contour (DLC) showing the Dice (A) and the Hausdorff distance (B) each subjective score group. Part C shows the time saved adjusting the software generated contours as a percentage of the manual contour time, displayed for each score group.

Grouping the contours per subjective score instead of the corresponding OAR or contouring method resulted in the Dice scores, Hausdorff distances and time saved as shown in Figure 5. Because the median time to contour an OAR varies greatly, e.g. 1.5 minutes for the left lung compared with 7.6 minutes for the mediastinum, the time saved was expressed in percentage of the time to manually contour the OAR. Comparing the manual contour to the software generated contours for OARs with a subjective score of 1 resulted in a median Dice score of 57% [range 16 - 99]; a median Hausdorff distance of 2.6 cm [range 0.4 - 4.5]; and a median time saved of 27% [range -70% to 92%]. Performing the same analysis for the OARs with a subjective score of 4 resulted in a median Dice score of 98% [range 70% - 99%]; a median Hausdorff distance of 0.4 cm [range 0.2 - 1.3]; and a median time saved of 79% [range -3% to 94%].

Discussion

The methods presented in this study can be used to evaluate any commercially available auto-contouring software. Sharp et al provide a comprehensive overview of the software that was available a few years ago [2]. Evaluating contour consistency and recording time saved can be used to setup a commissioning protocol to use auto-contouring software in clinical practice and can be used to evaluate the cost-effectiveness of such products. Systematically creating a model for a treatment site, creating a manual, software generated and user adjusted contour and evaluating them will provide the necessary information to safely introduce contouring software into clinical practice. Having a low-tech alternative, i.e. user adjustment strategies as opposed to a software strategy [14], could support the clinical use of auto-contouring software in the near future.

In similar studies, it is shown that auto-contouring techniques can save time [16,17], while other studies show promising new techniques that could potentially save time if investigated [10,18]. Young et al showed a potential time save for contouring lymph node for pelvic cancer patients [19]. Grouping the Dice scores, Hausdorff distances and time saved for each subjective score group comparing the manual to the user adjusted contours shows a relation between the consistency measurements and the subjective score. This demonstrates the ability of a radiation technician, while accounting for clinical guidelines, to judge whether contour quality is sufficient to lead to time savings compared with routine clinical practice. The technician can delete poor software generated contours and create a manual contour instead. This could form a basis of a viable and principled strategy for introducing imperfect automated contouring methods into clinical practice in order to save time while maintaining local clinical guidelines.

The results observed in the esophagus, heart and mediastinum highlight several issues with knowledge-based contouring software. First, the manual and user adjusted contours were significantly different while both were accepted as contours in compliance with local clinical guidelines by the same observer (intra-observer variability). For instance, the Dice score of the esophagus shows that even after adjusting the software generated contour to meet the clinical guidelines, the median Dice score is 78%, when comparing to the manual contour. Inter-observer variability in contours is reported to be an uncertainty that could greatly improve the quality of radiotherapy plans if reduced [1,3,20]. Van Baardwijk et al investigated a model-based contouring method for the Gross Tumor Volume (GTV) in CT-PET scans and showed a decrease in Inter-observer

variability [7]. Further investigation is needed to determine whether the automated contouring methods as described in this work could potentially reduce the intra- and inter-observer variability for OARs.

Using knowledge-based auto-contouring software, such as atlas-based and deep learning contouring, might improve the consistency of contours created by different observers because the software creates a contour which is the consensus of multiple observers when learning the model. In a future study, multiple observers should be included to determine the areas where they agree and where they do not. Comparing if the differences of the software generated contours happen in these specific parts of the contour will be a better quantification if the atlas is ready for clinical practice. This information, combined with the earlier suggested user adjustment method, could help with an important aspect of bringing automated contour techniques in the clinic, acceptance of these methods by the clinicians [2].

Deep learning is a state of the art machine learning technique that is utilized for many applications [15]. In health care specifically, there are several studies which investigate the utilization of deep learning for image contouring [21–23]. In our study, there are some promising results utilizing deep learning for the automatic contours of OARs for lung cancer patients. The deep learning contouring outperformed the atlas-based contouring for lungs and spinal cord. The deep learning performed better for the esophagus but further improvements remain necessary. In some cases, the user adjustment of the DLC did not save much time because slices near the stomach were missing and the technician judged these are clinically relevant to include. A similar case can be made for the heart example given in Figure 2C. The clinical training data included substantial differences of opinion in the training data where some observers include or exclude the vessels at the top of the heart differently, therefore the deep learning contouring method is trying to combine these differences of opinion, leading to inaccuracy. For both cases, reaching consensus between observers prior to training DLC could provide improved performance.

As pointed out by Nelms et al. variation in contouring can have a huge impact on the dosimetric properties of a radiotherapy treatment plan for head and neck cancers [24]. In a contouring peer-review of lung cancer, the effect of contouring deviation on the resulting radiation treatment was shown [25]. The latter study discusses the dosimetric differences compared to the clinical guidelines and concludes that further investigation on the actual impact on tumor control and normal tissue toxicity is needed. In our study, we used the objective assessment of

the technician to determine if contours meet clinical guidelines. Further investigation is needed to show the impact of these possible differences in contouring methods on dosimetry and finally treatment outcome.

Comparing the results of the automated contouring methods should be done with caution. The atlases are created from a highly curated set including an expert technician performing the contouring tasks that is validated by the radiation oncologist [26] while the DLC is learned on clinical data, which includes multiple observer preferences and possible imperfections. Comparing these two methods might be considered an unfair comparison, however, in this study we compared the current state of the art of both auto-contouring methods [22,26].

Time saving depends roughly on two main factors, the visualization of the boundary of the organ (e.g. lungs vs. esophagus) and the volume of the OAR to contour (e.g. mediastinal structures). High contrast edges, for instance the lungs, are easier to detect for both software and a human observer while low contrast edges, for instance the esophagus, are much harder. Automatic delineation methods typically are less accurate for small visible soft-tissue boundaries, which increases the time needed for adjustments, on top of this it is also more difficult for the human to distinguish where the contour should be, again increasing the time needed to adjust. Even if we assume that the auto-contouring techniques will reach human level contouring performance in the future, a human observer will probably still need time to evaluate difficult to contour OARs such as the esophagus.

Conclusion

Automatic contouring software as a starting point for clinical contours of OARs in lung radiation therapy allows for a significant time gain when contouring lungs, spinal cord, heart and mediastinum. DLC shows promising results with regard to the creation of institution-based models and to automatically generate high quality contours, providing a greater time saving compared to existing solutions. In addition, clinicians are able to assess if a software generated contour will potentially save time or not.

Acknowledgement

This work was funded by the Eurostars Cloud-Atlas project. (Project ID 9 297). MAASTRO clinic has research agreements with Mirada Medical Ltd.

References

1. Vinod SK, Jameson MG, Min M, Holloway LC. Uncertainties in volume delineation in radiation oncology: A systematic review and recommendations for future studies. *Radiother Oncol J Eur Soc Ther Radiol Oncol* 2016;121(2):169–179. doi:10.1016/j.radonc.2016.09.009
2. Sharp G, Fritscher KD, Pekar V, Peroni M, Shusharina N, Veeraraghavan H, *et al.* Vision 20/20: Perspectives on automated image segmentation for radiotherapy. *Med Phys* 2014;41(5). doi: 10.1118/1.4871620
3. Reed VK, Woodward WA, Zhang L, Strom EA, Perkins GH, Tereffe W, *et al.* Automatic segmentation of whole breast using an atlas approach and deformable image registration. *Int J Radiat Oncol Biol Phys* 2009;73(5):1493–1500. doi:10.1016/j.ijrobp.2008.07.001
4. Comparison and Evaluation of Methods for Liver Segmentation From CT Datasets - IEEE Xplore Document. <http://ieeexplore.ieee.org/document/4781564/> [accessed January 23, 2017]
5. Stapleford LJ, Lawson JD, Perkins C, Edelman S, Davis L, McDonald MW, *et al.* Evaluation of Automatic Atlas-Based Lymph Node Segmentation for Head-and-Neck Cancer. *Int J Radiat Oncol* 2010;77(3):959–966. doi:10.1016/j.ijrobp.2009.09.023
6. Hoang Duc AK, Eminowicz G, Mendes R, Wong SL, McClelland J, Modat M, *et al.* Validation of clinical acceptability of an atlas-based segmentation algorithm for the delineation of organs at risk in head and neck cancer. *Med Phys* 2015;42(9):5027–5034. doi:10.1118/1.4927567
7. van Baardwijk A, Bosmans G, Boersma L, Buijssen J, Wanders S, Hochstenbag M, *et al.* PET-CT-Based Auto-Contouring in Non-Small-Cell Lung Cancer Correlates With Pathology and Reduces Interobserver Variability in the Delineation of the Primary Tumor and Involved Nodal Volumes. *Int J Radiat Oncol* 2007;68(3):771–778. doi:10.1016/j.ijrobp.2006.12.067
8. Luo Y, Liao Z, Jiang W, Gomez D, Williamson R, Court L, *et al.* TU-H-CAMPUS-JeP2-05: Can Automatic Delineation of Cardiac Substructures On Noncontrast CT Be Used for Cardiac Toxicity Analysis? *Med Phys* 2016;43(6):3783. doi:10.1118/1.4957688
9. Kim J, Han J, Ailawadi S, Baker J, Hsia A, Xu Z, *et al.* SU-F-J-113: Multi-Atlas Based Automatic Organ Segmentation for Lung Radiotherapy Planning. *Med Phys* 2016;43(6):3433. doi:10.1118/1.4956021
10. Meng Q, Kitasaka T, Nimura Y, Oda M, Ueno J, Mori K. Automatic segmentation of airway tree based on local intensity filter and machine learning technique in 3D chest CT volume. *Int J Comput Assist Radiol Surg* 2017;12(2):245–261. doi:10.1007/s11548-016-1492-2
11. Kopriva I, Ju W, Zhang B, Shi F, Xiang D, Yu K, *et al.* Single-channel Sparse Nonnegative Blind Source Separation Method for Automatic 3D Delineation of Lung Tumor in PET Images. *IEEE J Biomed Health Inform* 2016. doi:10.1109/JBHI.2016.2624798
12. Reboças Filho PP, Cortez PC, da Silva Barros AC, C Albuquerque VH, R S Tavares JM. Novel and powerful 3D adaptive crisp active contour method applied in the segmentation of CT lung images. *Med Image Anal* 2017;35:503–516. doi:10.1016/j.media.2016.09.002
13. Dolz J, Kirişli HA, Fechter T, Karnitzki S, Oehlke O, Nestle U, *et al.* Interactive contour delineation of organs at risk in radiotherapy: Clinical evaluation on NSCLC patients. *Med Phys* 2016;43(5):2569. doi:10.1118/1.4947484
14. Schreibmann E, Marcus DM, Fox T. Multiatlas segmentation of thoracic and abdominal anatomy with level set-based local search. *J Appl Clin Med Phys* 2014;15(4):22–38. doi:10.1120/jacmp.v15i4.4468
15. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521(7553):436–444. doi:10.1038/nature14539

16. Haas B, Coradi T, Scholz M, Kunz P, Huber M, Oppitz U, *et al.* Automatic segmentation of thoracic and pelvic CT images for radiotherapy planning using implicit anatomic knowledge and organ-specific segmentation strategies. *Phys Med Biol* 2008;53(6):1751. doi:10.1088/0031-9155/53/6/017
17. Lim JY, Leech M. Use of auto-segmentation in the delineation of target volumes and organs at risk in head and neck. *Acta Oncol* 2016;55(7):799–806. doi:10.3109/0284186X.2016.1173723
18. Wang J, Guo H. Automatic Approach for Lung Segmentation with Juxta-Pleural Nodules from Thoracic CT Based on Contour Tracing and Correction. *Comput Math Methods Med* 2016;2016:2962047. doi:10.1155/2016/2962047
19. Young AV, Wortham A, Wernick I, Evans A, Ennis RD. Atlas-Based Segmentation Improves Consistency and Decreases Time Required for Contouring Postoperative Endometrial Cancer Nodal Volumes. *Int J Radiat Oncol* 2011;79(3):943–947. doi:10.1016/j.ijrobp.2010.04.063
20. Louie AV, Rodrigues G, Olsthoorn J, Palma D, Yu E, Yaremko B, *et al.* Inter-observer and intra-observer reliability for lung cancer target volume delineation in the 4D-CT era. *Radiother Oncol J Eur Soc Ther Radiol Oncol* 2010;95(2):166–171. doi:10.1016/j.radonc.2009.12.028
21. Hu P, Wu F, Peng J, Liang P, Kong D. Automatic 3D liver segmentation based on deep learning and globally optimized surface evolution. *Phys Med Biol* 2016;61(24):8676–8698. doi:10.1088/1361-6560/61/24/8676
22. Mansoor A, Cerrolaza JJ, Perez G, Biggs E, Nino G, Linguraru MG. Marginal Shape Deep Learning: Applications to Pediatric Lung Field Segmentation. *Proc SPIE-- Int Soc Opt Eng* 2017;10133. doi:10.1117/12.2254412
23. Suzuki K. Overview of deep learning in medical imaging. *Radiol Phys Technol* 2017. doi:10.1007/s12194-017-0406-5
24. Nelms BE, Tomé WA, Robinson G, Wheeler J. Variations in the Contouring of Organs at Risk: Test Case From a Patient With Oropharyngeal Cancer. *Int J Radiat Oncol* 2012;82(1):368–378. doi:10.1016/j.ijrobp.2010.10.019
25. Lo AC, Liu M, Chan E, Lund C, Truong PT, Loewen S, *et al.* The Impact of Peer Review of Volume Delineation in Stereotactic Body Radiation Therapy Planning for Primary Lung Cancer: A Multicenter Quality Assurance Study. *J Thorac Oncol* 2014;9(4):527–533. doi:10.1097/JTO.0000000000000119
26. Peressutti D, Schipaanboord B, van Soest J, Lustberg T, van Elmpt W, Kadir T, *et al.* TU-AB-202-10: How Effective Are Current Atlas Selection Methods for Atlas-Based Auto-Contouring in Radiotherapy Planning? *Med Phys* 2016;43(6Part33):3738–3739. doi:10.1118/1.4957432

Chapter 8

Discussion and future prospective

Tim Lustberg

In this chapter, lessons learned from the work of thesis and suggestions for future directions are given. One of the main conclusions is that using Big Data technology to innovate health care is still a very challenging field with many factors to consider. The current data collection methods: clinical trials, clinical registries and routine clinical data, all have their use in medical research as described in Chapter 2. All of these sources are a valuable source of information to innovate health care, and instead of viewing them as competing – as is sometimes postulated [1], they should be valued as complementary. We need all the data we can get to innovate health care.

Data representation and technology

Big Data technology is dependent on the available data; low quality data will result in low quality models [2–4]. Data quality problems originate from missing data, unstructured data and ambiguous data. Trying to gain new insights using the available data from one cohort and trying to validate this insight at a different cohort highlights these problems. This approach is considered the best practice according to the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD)[5]. As an example, looking at dyspnea for lung cancer patients treated with radiotherapy, the care provider responsible for the validation cohort might simply not record dyspnea (missing data), or record only the severe cases of dyspnea (biased missing data). The dyspnea might be noted as free text, often in native language and with abbreviations (unstructured, ambiguous data). Translating this free text into a coding system, e.g. Systematized Nomenclature of Medicine - Clinical Terms (SNOMED-CT), can again result in errors and uncertainties. For instance, deciding what grade A (the patient sometimes experiences severe shortness of breath) dyspnea finding is according to SNOMED-CT can result in different classifications depending on the expert creating the mapping (ambiguous data). As pointed out by K. Fessele “To promote high-quality, high-value cancer care, big data platforms must be constructed from standardized data sources to support extraction of meaningful, comparable insights.”[6]. Applying technology as described in Chapter 4 takes a step towards creating these standardized data sources for secondary use. It describes a pragmatic approach to apply this to medical imaging data, but the approach can be extended to other health care data sources and even beyond health care.

Challenges of collecting and exchanging data

The first challenge of any Big Data project is acquisition of the data. In our field, this data exists in various clinical systems and for good reasons they are not easily accessible. A common approach is creating a copy of the clinical systems' databases at a moment that is suitable to the health care provider. To interpret the data of these systems, the data is processed by highly customized, expensive software (professor-ware). For instance, the rapid learning platform described in Chapter 6 took 2 months to set up, for an already existing model with very little data elements required. To investigate and demonstrate the use of the technology these approaches are well suited but implementing Big Data technology into production-level software will require more scalable data acquisition solutions. Initiatives like the Clinical Building Blocks (CBB) [7] aim to provide a standard to collect and exchange a bare minimum set of information about a patient on a national level for the Netherlands. Each patient record would represent the information in a similar manner if the vendors implement the CBB as prescribed. This enables easy exchange of this basic information in a multi-vendor environment, for instance by a HL7 FHIR [8] interface on top of their database. It is a good thing that such standardized data capture and exchange is now becoming mandatory in the Netherlands e.g. through the so-called Basisgegevensset Zorg [9].

Natural Language Processing (NLP) is the process of machine interpreting natural language to extract structured data, e.g. SNOMED-CT codes. NLP faces the same challenges as discussed in the paragraph above, with the added complexity of the many small variations in the use of language between clinicians recording the information. Free text in health care is plagued with ambiguous acronyms, shorthand, spelling errors and native language, which make it harder to standardize the use of this kind of tools [3,10]. There are promising advances in this field, such as Hassan et al. achieving human level of language translation between Chinese and English [11]. Research has shown the potential of text mining the electronic medical records and referral letters [12–14], but there is still need for improvement. Increasing effectiveness of these tools will help to speed up the data acquisition process without forcing clinicians to provide only structured information.

Syntactic interoperability, being able to read a data field, is the first step to understanding any data source. Being able to read a data element is not enough. The meaning of what the data element represents and who added this

information is needed to make the data truly semantically interoperable. Having standardized imports such as HL7 FHIR will help automate this process, but as demonstrated in Chapter 4 there is a need for easy domain expert interaction with the data, as they understand what the data represents – in other words HL7 FHIR does not by itself achieve semantic interoperability. Following the example of Chapter 4, the clinicians who created the delineations are the ones who know what they meant when creating the delineations. Recording who created this mapping can help users of the data. For instance, the delineation map could have been created by a researcher who was creating his own cohort when the domain experts are unavailable. Having this information in a machine-readable format will help other users of the data source decide if the source is suitable for their intended analysis.

FAIR-data platform using Semantic Web

The FAIR-data principles provide guidelines for creating data stores for research in a sustainable matter [15]. These principles state that data has to be stored in a Findable, Accessible, Interoperable, and Reusable manner. They provide clear instructions on the required meta-data, data governance and data management but do not dictate the use of any particular technology. Relational databases do not store meta-data accessible for machines about the database itself. This knowledge is stored in the database documentation and the source code of the application using the database. For instance, having a column called patient weight does not contain all necessary information (e.g. what is unit of the weight recorded or was the weight recorded with or without clothes). When combining data from multiple sources, which is needed for Big Data analysis, this kind of information is needed to prevent errors or bias in the results. Semantic Web technology [16] should be used to add the needed meta-data in machine-readable format to comply with the FAIR-data guidelines.

Figure 1 provides a schematic overview of the FAIR-data platform design, applying all lessons learned in this thesis. The first step is to import data and add the context of this data to the system. A widely supported open source initiative to perform this import and mapping is the Resource Mapping Language (RML) [17,18]. RML provides tools to map source data to an ontology without the need for a specification of the technological implementation of the source data. This is the responsibility of the RML-processor. An example of an Open-Source project

which can parse RML is Ontop [19]. The example in Figure 1 shows the data source “niet kleincellig longcarcinoom” which is Dutch for non-small cell lung carcinoma. For this example, we assume that the patient related information is already mapped and available in the FAIR-data platform. Using the RML language, the value can be linked to the Radiation Oncology Ontology (ROO) [20] Unique Resource Identifier (URI) for patient has disease (roo:100020). The mapping has to be created by an employee with technical skills and internal knowledge of the source database. After mapping, the standardized sources can be imported automatically without any human interaction, like CBBs exported with HL7 FHIR. However, for the foreseeable future there will a need for tools to add the same context to data sources who do not provide it itself, such as spreadsheets with data manually collected by a researcher.

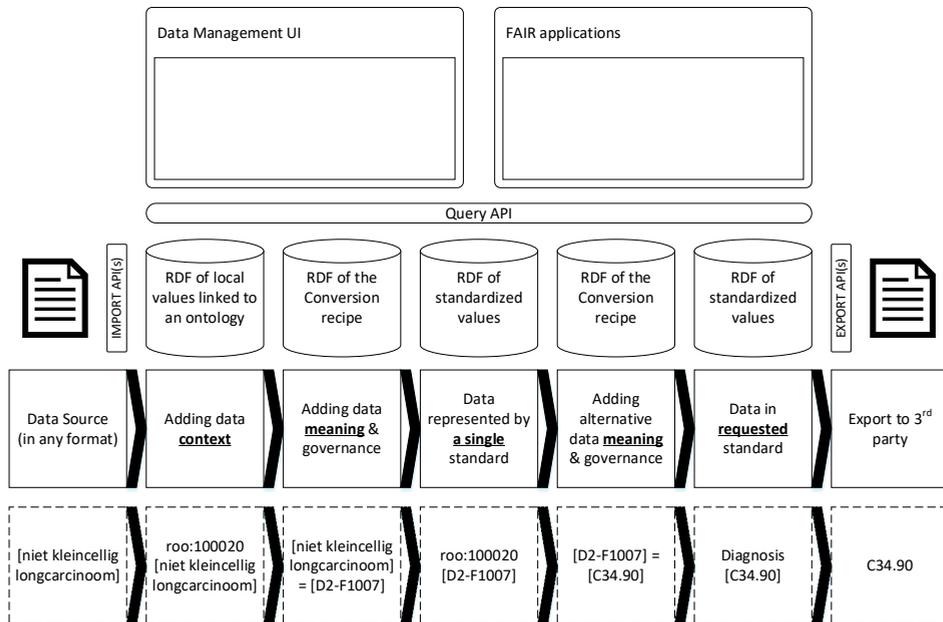


Figure 1: Schematic FAIR-platform workflow (including example)

After adding the context, meaning needs to be added in a machine-readable format. The data management UI needs to provide domain experts with the option to add mappings. In the example, the Dutch NSSLIC value needs to be mapped to the corresponding SNOMED-CT code (D2-F1007). This information is stored in the Web Ontology Language (OWL) format, and can be used by the Semantic Web reasoning engine to create the data represented by a single standard. OWL

is the ontology representation in Semantic Web format that can be queried by a machine. When adding these mappings, the API should also add the user providing the mapping, like the terminology linker of Chapter 4.

To satisfy the findable and accessible requirements of FAIR, the query API needs to provide summary information about the underlying data sources, e.g. how many lung cancer patients are available in the database. This enables a machine to generate catalogues about the available data, which the researcher can use to investigate if the data source can be used to answer their specific hypothesis. The EMIF-Platform is a commercially available product that has implemented to catalogue and governance very clearly [21]. All data sources have a custodian who can be contacted through the application to request access to the data. This is the first step in setting up a legal research collaboration, which has to be done outside of the platform. The custodian is also responsible for adding the meta-data for the data source so that the catalogue can be filled correctly.

Health care professionals do not tend to agree on a single standard. This underlines the need to have domain experts involved in the conversion between standards if possible. Sometimes this can result in a loss of detail of the data, but it can still serve a purpose. This is not the case for the example, SNOMED-CT provides mappings to other coding systems, and the NSSLC finding is semantically identical to a corresponding International Classification of Diseases version 10 code (ICD-10). Because the mapping is already provided by SNOMED-CT, no further human-interaction is required when a user of the platform asks a question in ICD-10 instead of SNOMED-CT. There are many cases where the ontology to convert between systems does not exist, the same mapping process described above can be used to have domain experts provide their insights on the mapping.

Having access to the source data, how it was converted, by whom and having it placed in context will greatly increase the effectiveness and transparency when applying Big Data technologies.

Big Data quality management challenges

A limitation of this work is the applicability to clinical practice. The TRIPOD statement provides insight on how to judge quality of Big Data models [5]. Performing model validations retrospectively is very achievable with the right amount of effort. However, with current regulation and legislation it is hard to create a self-learning system and use it for patient care. Productizing of software takes several months, with good reason as it is meant to ensure the safety of patients. A governing entity does not have the right tools or knowledge to evaluate

software that will learn something new every day based on the available data. Zhang et al. have acknowledged these issues and are providing a clear overview of the techniques currently available [11]. More research, together with commercial partners, is needed to understand how to safely use Big Data technology for patient care.

Clinical applicability of Big Data technology

Technologically Big Data has the potential to move health care forward to be more personalized, affordable and effective. However, many studies are focused too much on what the technology can offer. Acceptance of the technology by clinical users is an important factor for the applicability. Physicians need to be able to trust the tools they are given and it has to solve a problem for them. Big Data technology like Deep Learning can be very powerful, but they are harder to understand because they are a black box. Classical machine learning approaches, such as linear regression, may result in sub-optimal models but still be favored over the Deep Learning models because they are understandable. Chapter 7 describes a good example how to implement Big Data technology while keeping the clinical professional engaged. It solved a clear problem: delineating organs at risk takes a lot of time, speeding up this process can help clinical personnel find time to focus on other tasks. The Deep Learning algorithm was very well suited to learn the delineation tasks, but providing these delineations as suggestions to the clinical personnel helped in the acceptance of the results. The clinicians are and feel responsible for the delineation so being in control of the end result while still solving their problem (saving time) demonstrated a useful approach when trying to implement Big Data technology into clinical care. In time, when the algorithm gets better and the trust in the technology increases, the human check might become obsolete – as we have seen in for instance radiotherapy treatment planning. In future, more studies are needed to demonstrate the usefulness of applying Big Data technology in this manner. Big Data studies should focus more on a problem that the clinician wants solved in a manner that is acceptable to them as the end users of that solution.

Quality data collection is, and always will be, hard work. Using modern day technology as described in this work will definitely help streamline the process and prevent rework when processing the data for secondary use. A better solution would be to make the source data recorded by clinical staff FAIR. For this,

clinical staff needs to be educated in the importance of quality data collections and companies need to be convinced that they should invest in this kind of technology which allow easier data capture and better reuse. New technological solutions, like a dictionary suggesting the right terms/spelling when entering data, can help improve data quality when recording information into the Electronic Medical Record. It is our tasks, as data scientists, to help clinical personnel and companies developing medical applications to make this technological leap. Having immediate benefits for the care professionals will help involve them in better data collection. If a certain insight is currently not available to them and they can see this improving as they improve their data recording practices, they will be motivated to put in the work.

Technology alone cannot drive innovation in health care. Chapter 5-7 all demonstrate that the success lies in the combination of technology and human interaction.

References

1. Dekker ALAJ, Gulliford SL, Ebert MA, Orton CG. Point/Counterpoint. Future radiotherapy practice will be based on evidence from retrospective interrogation of linked clinical data sources rather than prospective randomized controlled clinical trials. *Med Phys* 2014;41(3):030601. doi:10.1118/1.4832139
2. Arora VM. Harnessing the Power of Big Data to Improve Graduate Medical Education: Big Idea or Bust? *Acad Med J Assoc Am Med Coll* 2018. doi:10.1097/ACM.0000000000002209
3. Mehta N, Pandit A. Concurrence of big data analytics and healthcare: A systematic review. *Int J Med Inf* 2018;114:57–65. doi:10.1016/j.ijmedinf.2018.03.013
4. Helbing D. The hidden danger of big data. <https://www.aljazeera.com/indepth/opinion/2016/08/hidden-danger-big-data-160816140935829.html> [accessed March 16, 2018]
5. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Med* 2015;13(1):1. doi:10.1186/s12916-014-0241-z
6. Fessele KL. The Rise of Big Data in Oncology. *Semin Oncol Nurs* 2018. doi:10.1016/j.soncn.2018.03.008
7. HCIM Mainpage - Zorginformatiebouwstenen. https://zibs.nl/wiki/HCIM_Mainpage [accessed May 7, 2018]
8. Bender D, Sartipi K. HL7 FHIR: An Agile and RESTful approach to healthcare information exchange. *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems, Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*. 2013. doi:10.1109/CBMS.2013.6627810
9. Basisgegevensset Zorg vastgesteld als standaard voor uitwisseling patiëntgegevens - ZonMw. <https://www.zonmw.nl/nl/actueel/nieuws/detail/item/basisgegevensset-zorg-vastgesteld-als-standaard-voor-uitwisseling-patientgegevens/> [accessed May 14, 2018]
10. Khalifa A, Meystre S. Adapting existing natural language processing resources for cardiovascular risk factors identification in clinical notes. *J Biomed Inform* 2015;58:S128–S132. doi:10.1016/j.jbi.2015.08.002
11. Hassan H, Aue A, Chen C, Chowdhary V, Clark J, Federmann C, et al. Achieving Human Parity on Automatic Chinese to English News Translation. *ArXiv180305567 Cs* 2018
12. Hegde H, Shimpi N, Glurich I, Acharya A. Tobacco use status from clinical notes using Natural Language Processing and rule based algorithm. *Technol Health Care Off J Eur Soc Eng Med* 2018. doi:10.3233/THC-171127
13. Kreuzthaler M, Pfeifer B, Vera Ramos JA, Kramer D, Grogger V, Bredenfeldt S, et al. EHR Text Categorization for Enhanced Patient-Based Document Navigation. *Stud Health Technol Inform* 2018;248:100–107
14. Todd J, Richards B, Vanstone BJ, Gepp A. Text Mining and Automation for Processing of Patient Referrals. *Appl Clin Inform* 2018;9(1):232–237. doi:10.1055/s-0038-1639482
15. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 2016;3:160018. doi:10.1038/sdata.2016.18
16. Semantic Web - W3C. <https://www.w3.org/standards/semanticweb/> [accessed May 9, 2018]
17. Dimou A, Kontokostas D, Freudenberg M, Verborgh R, Lehmann J, Mannens E, et al. Assessing and Refining Mappings to RDF to Improve Dataset Quality. vol. 9367, *14th International Semantic Web Conference*. 2015

Chapter 8

18. Dimou A, De Nies T, Verborgh R, Mannens E, Van de Walle R. Automated Metadata Generation for Linked Data Generation and Publishing Workflows. vol. 1593, *9th Workshop on Linked Data on the Web*. 2016
19. Ontop. <https://ontop.inf.unibz.it/> [accessed May 8, 2018]
20. Traverso A, van Soest J, Wee L, Dekker A. The Radiation Oncology Ontology (ROO): publishing linked data in radiation 1 oncology using Semantic Web and Ontology techniques. *Submiss*
21. EMIF Platform - EMIF. <http://www.emif.eu/about/emif-platform> [accessed May 8, 2018]

Summary

To improve our health care systems, we need to learn from all the data available to us. Collecting, combining and learning from large amounts of data is often referred as Big Data. Dan Ariely stated: *“Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it...”*. This is represented by the many definitions available for Big Data. To provide context on the subject **Chapter 2** explains the meaning and current state of Big Data in radiation oncology. To utilize ‘Big Data’ in healthcare the data has to be FAIR (Findable, Accessible, Interoperable and Reusable). But the FAIR-data principles are guidelines, they do not enforce the use of certain technology. Chapter 3 and 4 demonstrate what (partially) applying these guidelines will achieve.

Chapter 3 explains the technology used in the Semantic DICOM (SeDI) project. One of the biggest sources of data in radiation oncology are the digital representations of anatomy and treatment intent, saved as DICOM images. These images consist of two main components, the information about the image (metadata) and the actual image. Traditionally the metadata is partly stored in a separate database next to the DICOM-files. When using a traditional PACS system, trying to gather all CT-scans for lung cancer patients with a tumor of a certain volume, is almost impossible. However, using the Semantic Web technology explained in Chapter 3, we were able to answer this question.

Chapter 4 builds upon the technology presented in Chapter 3, further expanding the use of FAIR data guidelines to improve data utilization. One of the remaining challenges that wasn’t solved by SeDI is the free-text delineation names in the DICOM metadata fields. These names are added by the radiation oncology staff when delineating the organs and tumor(s) of a patient. Inconsistencies in these names, due to language, different guidelines, typos, etc., can be a huge problem in large scale image feature extraction. Chapter 4 provides a scalable solution to overcome these challenges using Semantic Web technology to create linked data out of the free-text source data.

Chapter 5 describes a clinical study which used an early version of the software described in Chapter 4. The goal of the study was to compare set-up and 2D EPID dosimetry data of breast cancer patients treated during voluntary moderately Deep Inspiration Breath Hold and free breathing. Identifying the correct DICOM images and processing them automatically resulted in the insight that both image sources indicate that reproducibility of radiotherapy for patients treated during free breathing and voluntary moderately Deep Inspiration Breath Hold is similar.

Chapter 6 describes the validation of a survival prediction model of larynx cancer patients using a rapid learning platform. For this study, a datamining infrastructure was set up to collect the data required for the prediction model to create a comparison to the original clinical dataset. A third dataset was included from a clinical trial performed by the RTOG. Analyzing the data available through routine clinical practice and collected a prospective clinical trial provided insights into the performance of the survival prediction model. The study showed the model performed as expected for the clinical validation data but performed poorly on the clinical trial dataset.

Chapter 7 evaluates the usefulness of automatic contouring of organs at risk (OARs) for lung cancer patients. To perform this evaluation two new auto-contouring methods were compared to the current clinical standard of contouring for lung cancer patients. The time required to perform the delineations following the standard procedures, adjusting an atlas based contour and adjusting a contour generated by a Deep Learning algorithm were recorded. The study showed that for these patients, pre-generating the contours and correcting the inconsistencies saved time over the current clinical practice while adhering to the local guidelines.

Finally, **Chapter 8** summarizes the lessons learned. A FAIR-data platform design is explained to provide future direction for data research.

Samenvatting

Om innovatie in de zorg te verbeteren is het belangrijk dat we leren om alle beschikbare gegevens nuttig in te zetten. Het verzamelen, combineren en verwerken van gegevens op grote schaal wordt vaker onder de noemer “Big Data” geschaard. Een bekend citaat van Dan Ariely is: *“Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it...”*. Het komische citaat geeft aandacht aan het probleem dat er vele definities van Big Data te vinden zijn. In **hoofdstuk 2** wordt deze definitie en de huidige status van Big Data verder uitgewerkt voor radiotherapeutische oncologie. Om de zorg daadwerkelijk in het Big Data domein te krijgen is het belangrijk dat gegevens FAIR (Findable, Accessible, Interoperable, Reusable) opgeslagen worden. De Nederlandse vertaling van de term FAIR is dat de gegevens vindbaar, toegankelijk, interoperabel en herbruikbaar moeten zijn. De FAIR-richtlijnen stellen geen eisen aan de technologie die ingezet wordt om het doel te bereiken. Hoofdstuk 3 en 4 beschrijven software implementaties die (deels) conformeren aan de FAIR-richtlijnen, en wat het voordeel hiervan is.

In **hoofdstuk 3** wordt de software die geschreven is voor het Semantic DICOM (SeDI) project beschreven. Een van de grootste bronnen aan gegevens binnen de radiotherapeutische oncologie zijn de vele medische beelden die gemaakt worden als onderdeel van het diagnose- en behandelproces, deze worden opgeslagen als DICOM-beelden. Dit type beeldrepresentatie is opgedeeld in twee onderdelen: het daadwerkelijke beeld (de pixels) en de gegevens over het beeld (meta-data). De huidige klinische praktijk is dat de beelden in een PACS worden opgeslagen, waarbij de meta-data voor een deel opgeslagen wordt in een relationele database. Om praktische redenen zijn niet alle meta-data beschikbaar via de relationele database en moeten de bestanden uitgelezen worden om de informatie te achterhalen. Daarnaast is er meestal geen optie om additionele gegevens over het beeld op te slaan in deze database. Hierdoor zijn sommige onderzoeksvragen moeilijk te beantwoorden. Het verzamelen van alle beelden van longkankerpatiënten met een tumor van een bepaald volume is met een PACS niet mogelijk zonder alle beeldbestanden uit te lezen. Het is met de Semantic Web technologie uit hoofdstuk 3 wel mogelijk om dit snel en efficiënt te verzamelen.

Hoofdstuk 4 beschrijft de oplossing voor een van overgebleven uitdagingen van hoofdstuk 3, de naamgeving van de intekeningen. De naamgeving van de tumoren orgaanontrek intekeningen worden opgeslagen als vrije tekst. SeDI slaat deze meta-data op inclusief typefouten, taalverschillen en definitie verschillen. Voor

automatisering kunnen deze verschillen een groot probleem zijn. Het hindert de verwerking van deze gegevens op grote schaal. De softwareoplossing uit hoofdstuk 4 beschrijft een schaalbare oplossing voor dit probleem met minimale invoer van de gebruikers. Met behulp van Semantic Web technologie worden de vrije tekstvelden gekoppeld aan een definitie die leesbaar is voor mens en machine.

Hoofdstuk 5 beschrijft een klinische studie waar gebruik wordt gemaakt van een eerdere versie van de software uit hoofdstuk 4. Voor de studie zijn de set-up beelden en 2D EPID dosimetrie beelden van borstkankerpatiënten verzameld. De beelden van twee patiëntengroepen werden met elkaar vergeleken: patiënten die tijdens de behandeling vrij mochten ademhalen en patiënten waarbij gebruik werd gemaakt van een techniek waarbij de patiënt vrijwillig de adem inhoudt (voluntary moderately Deep Inspiration Breath Hold). De medische beelden zijn met de software geïdentificeerd en automatisch verwerkt om de reproduceerbaarheid van de behandelingen voor beide technieken te evalueren. Uit deze analyse is te concluderen dat de reproduceerbaarheid vergelijkbaar is.

Hoofdstuk 6 beschrijft de validatie van een voorspellingsmodel voor larynxkankerpatiënten gebruikmakend van een “rapid learning” platform. Voor deze studie werd een datamining infrastructuur opgezet om structureel klinische gegevens te verzamelen die nodig zijn voor het voorspellingsmodel. De resultaten van het voorspellingsmodel werden vergeleken met de klinische dataset waar het model op geleerd was. Daarnaast werd een derde dataset geanalyseerd van een klinisch trial uitgevoerd door de RTOG. De validatie toonde aan dat het voorspellingsmodel vergelijkbaar presteerde voor de klinische validatie dataset maar het voorspellingsmodel presteerde slecht voor de klinische trial dataset.

Hoofdstuk 7 evalueert de klinische inzetbaarheid van automatische intekensoftware voor risico-organen bij longkankerpatiënten. Hierbij werden twee automatische inteken methodes vergeleken met de huidige klinische handmatige standaardmethode. De automatische intekeningen werden gemaakt met behulp van atlas gebaseerde intekeningen en “Deep Learning” algoritme. Om de klinische inzetbaarheid te meten werd de benodigde tijd om de organen in te tekenen gemeten voor de standaardmethode en de benodigde tijd om de automatische intekeningen aan te passen tot ze aan de klinische standaard voldeden. De studie toonde aan dat het inzetten van automatische intekensoftware met een menselijke validatie tijd bespaart in de kliniek terwijl de lokale richtlijnen blijven gehandhaafd.

Tot slot, vat **hoofdstuk 8** de inzichten verkregen door de verschillende studies samen. Daarnaast wordt een FAIR-data platform beschreven om richting te geven aan vervolgonderzoek.

Valorisation Addendum

Providing the best quality care while keeping it affordable is one of the biggest challenges of the 21st century. Finding solutions to start overcoming this challenge will require data. The work presented in this thesis contributes to the implementation of data driven medicine to help care providers use this technology.

The projects funded by Eurostars have a strong focus on commercialization of the research. Project groups consist of academic and one or more commercial parties. During the Semantic DICOM project (Chapter 3) a new kind of picture archive application was developed. The main focus of this software is to easily find medical images and link derived data from these images directly to the source. SOHARD software GmbH (Fürth, Germany) was the commercial party involved in the project and is selling the software created in this project, mainly for research purposes. The image analysis application was also shared with the community through an Open Source project on bitbucket (<https://bitbucket.org/account/user/maastro/projects/MIA>). The application can process radiotherapy DICOM objects to calculate dose volume histograms. With these software products medical images and their derived features can be made FAIR to improve the data management of research projects. The second project, CloudAtlas, involved both Mirada Medical Ltd (Oxford, UK) and SOHARD software. The original project plan was to provide a cloud platform to perform atlas contouring. The idea was too ambitious to realize in the two years of the project, so the team focused on the work necessary to eventually provide this solution. The project resulted into two commercial products. Mirada Medical developed a new automatic contouring method using Deep Learning technology which can be used to improve delineation efficiency in clinical practice - they have received FDA clearance for this product in July 2018. SOHARD software created a radiotherapy DICOM web-viewer, which was currently unavailable as a commercial product but is expected to be a valuable tool for researchers.

Medical Data Works B.V.

Medical Data Works is a start-up which was started in 2017 as a result of the various research projects of the MAASTRO Clinic knowledge engineering team. Medical Data Works provides consultancy to help start or improve data collection, integration or analysis projects. Sharing insights learned and supporting the software created by the knowledge engineering team with care providers is our main focus. Our current domain is radiation oncology, but the technology could be applied to medical oncology and even other medical specialties. Medical Data Works closed its first financial year with a profit.



The FAIR platform described in **Chapter 8** is scheduled to be developed by Medical Data Works in 2019. To support the consulting activities of the company the platform will be developed with production grade software development requirements to ensure it can be adopted by hospital IT-departments. The software which is installed on network infrastructure of the hospital, even if it is for research purposes, needs to be supported. The IT-department needs to have Service Level Agreements (SLAs) with a third-party. This is often a roadblock for software written for research purposes, it is not in the interest of the researcher to support their software once the work is published. Medical Data Works wants to adapt Open Source software from successful research projects and provide support through SLAs to help researchers overcome this roadblock and improve utilization of research software.

The biggest concern is scalability. The software in its current state is highly dependent on specialists (data scientists) for configuration and utilization. Medical Data Works can provide this service by creating a larger team of data scientists. However, the technical people required for the job are in short supply in the Netherlands, even worldwide. When the company is involved in enough projects, scaling the available personnel will be challenging. The other challenge is the cost of the consultancy and support. Setting up a FAIR-data platform requires a data scientist to analyze the data and create the transformations. This is time consuming and heavily dependent on the source data systems, any change in the source will have to be incorporated in the data capture and transformation part of the FAIR-data platform. The goal is to simplify the process with smart software solutions to enable a team of service engineers to perform these tasks.

Curriculum Vitae

Tim Lustberg was born on the 5th of October 1987 in Brunssum in the Netherlands. After graduating high-school in 2006, he studied Electrical Engineering at Zuyd University in Heerlen. While his specialization was electronics and telecommunication, Tim always had an interest in software development. As part of the graduation internship for the Bachelor degree in 2010 he was part of the physics department of MAASTRO clinic, writing software to capture images during radiation treatment. The goal of the experiment was to improve the current Dose Guided Radiation Therapy quality assurance program. After graduating he started as a medical physics engineer / software developer with the IT and physics department of MAASTRO clinic. After two years of writing software, maintaining medical devices and learning about the clinic process, he applied for a PhD studentship. The goal of this research was to learn on the job and provide practical insights on Big Data technology and applications in health care. To apply all lessons learned, he co-founded Medical Data Works B.V. in 2017 to keep contributing to innovation in health care at a practical level. And finally, in 2018, under supervision of Prof. dr. ir. Frank Verhaegen and Prof. dr. ir. Andre Dekker, he completed his PhD thesis on Big Data technology in health care.

List of publications

1. **Lustberg T**, Bailey M, Thwaites DI, Miller A, Carolan M, Holloway L, Velazquez ER, Hoebbers F, Dekker A. Implementation of a rapid learning platform: Predicting 2-year survival in laryngeal carcinoma patients in a clinical setting. *Oncotarget* 2016;7(24):37288–37296. doi:10.18632/oncotarget.8755
2. **Lustberg T**, van Soest J, Jochems A, Deist T, van Wijk Y, Walsh S, Lambin P, Dekker A. Big Data in radiation therapy: challenges and opportunities. *Br J Radiol* 2017;90(1069):20160689. doi:10.1259/bjr.20160689
3. **Lustberg T**, van Soest J, Gooding M, Peressutti D, Aljabar P, van der Stoep J, van Elmpt W, Dekker A. Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer. *Radiother Oncol J Eur Soc Ther Radiol Oncol* 2018;126(2):312–317. doi:10.1016/j.radonc.2017.11.012
4. **Lustberg T**, van Soest J, Fick P, Fijten R, Hendriks T, Puts S, Dekker A. Radiation Oncology Terminology Linker: A Step Towards a Linked Data Knowledge Base. *Stud Health Technol Inform* 2018;247:855–859
5. Brouwers PJAM, **Lustberg T**, Borger JH, van Baardwijk AAW, Jager JJ, Murrer LHP, Nijsten SMJJG, Reymen BH, van Loon JGM, Boersma LJ. Set-up verification and 2-dimensional electronic portal imaging device dosimetry during breath hold compared with free breathing in breast cancer radiation therapy. *Pract Radiat Oncol* 2015;5(3):e135-141. doi:10.1016/j.prro.2014.10.005
6. Van Soest J, **Lustberg T**, Grittner D, Marshall MS, Persoon L, Nijsten B, Feltens P, Dekker A. Towards a semantic PACS: Using Semantic Web technology to represent imaging data. *Stud Health Technol Inform* 2014;205:166–170
7. Cheng Q, Roelofs E, Ramaekers BLT, Eekers D, van Soest J, **Lustberg T**, Hendriks T, Hoebbers F, van der Laan HP, Korevaar EW, Dekker A, Langendijk JA, Lambin P. 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. *Radiother Oncol J Eur Soc Ther Radiol Oncol* 2016;118(2):281–285. doi:10.1016/j.radonc.2015.12.029
8. Deist TM, Dankers FJWM, Valdes G, Wijsman R, Hsu IC, Oberije C, **Lustberg T**, van Soest J, Hoebbers F, Jochems A, El Naqa I, Wee L, Morin O, Raleigh DR, Bots W, Kaanders JH, Belderbos J, Kwint M, Solberg T, Monshouwer R,

- Bussink J, Dekker A, Lambin P. Machine learning algorithms for outcome prediction in (chemo)radiotherapy: An empirical comparison of classifiers. *Med Phys* 2018;45(7):3449–3459. doi:10.1002/mp.12967
9. Lambin P, Zindler J, Vanneste BGL, De Voorde LV, Eekers D, Compter I, Panth KM, Peerlings J, Larue RTHM, Deist TM, Jochems A, **Lustberg T**, van Soest J, de Jong EEC, Even AJG, Reymen B, Rekers N, van Gisbergen M, Roelofs E, Carvalho S, Leijenaar RTH, Zegers CML, Jacobs M, van Timmeren J, Brouwers P, Lal JA, Dubois L, Yaromina A, Van Limbergen EJ, Berbee M, van Elmpt W, Oberije C, Ramaekers B, Dekker A, Boersma LJ, Hoebbers F, Smits KM, Berlanga AJ, Walsh S. Decision support systems for personalized and participative radiation oncology. *Adv Drug Deliv Rev* 2016. doi:10.1016/j.addr.2016.01.006
 10. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, Sanduleanu S, Larue RTHM, Even AJG, Jochems A, van Wijk Y, Woodruff H, van Soest J, **Lustberg T**, Roelofs E, van Elmpt W, Dekker A, Mottaghy FM, Wildberger JE, Walsh S. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 2017;14(12): 749–762. doi:10.1038/nrclinonc.2017.141
 11. Peressutti D, Schipaanboord B, van Soest J, **Lustberg T**, van Elmpt W, Kadir T, Dekker A, Gooding M. TU-AB-202-10: How Effective Are Current Atlas Selection Methods for Atlas-Based Auto-Contouring in Radiotherapy Planning? *Med Phys* 2016;43(6Part33):3738–3739. doi:10.1118/1.4957432
 12. Persoon LCGG, Nijsten SMJJG, Wilbrink FJ, Podesta M, Snaith J a. D, **Lustberg T**, Elmpt WJC van, Gils F van, Verhaegen F. Interfractional trend analysis of dose differences based on 2D transit portal dosimetry. *Phys Med Biol* 2012;57(20):6445. doi:10.1088/0031-9155/57/20/6445
 13. Podesta M, Nijsten SMJJG, Snaith J, Orlandini M, **Lustberg T**, Emans D, Aland T, Verhaegen F. Measured vs simulated portal images for low MU fields on three accelerator types: possible consequences for 2D portal dosimetry. *Med Phys* 2012;39(12):7470–7479. doi:10.1118/1.4761950
 14. Schipaanboord B, Boukerroui D, Peressutti D, van Soest J, **Lustberg T**, Kadir T, Dekker A, van Elmpt W, Gooding M. Can atlas-based auto-segmentation ever be perfect? Insights from Extreme Value Theory. *IEEE Trans Med Imaging* 2018. doi:10.1109/TMI.2018.2856464

