System radiobiology modelling of radiation induced lung disease

Sangkyu Lee

Doctor of Philosophy

Department of Physics

McGill University

Montreal, Quebec

June 1, 2016

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy in Physics

ABSTRACT

Radiation induced lung disease (RILD) is a side effect of radiotherapy for treating thoracic cancers, limiting radiation dose to tumours and in turn the chance of treatment success. A current scheme for predicting and managing RILD risk is based on a population-based normal tissue complication probability (NTCP) model assuming the same response to given radiation dose in lung. However, recent research suggests that dose response can be modified by biological and clinical factors pertinent to pathogenesis of RILD. In this work, we explore systems radiobiology approaches to model RILD as a result of interactions between these factors. Clinical, dosimetric, and biological data on lung cancer patients were analyzed to identify markers associated with high RILD risk. Then, we applied machine learning methods to combine such markers into models that calculate patient-specific RILD risk. We investigated two RILD endpoints: radiation fibrosis (RF) and radiation pneumonitis (RP). RF is formation of scar tissues in lung and can be quantitatively measured from computed tomography (CT) images. We extended a classical NTCP model to explicitly model time-dependent dose response of RF risk. Our modelling results have shown significant change in dose-RF correlation after 3 months post-treatment as well as higher RF risk when tumour was in lower lung. We extended the dose modelling to intra-treatment CT images. However, we did not find association between early CT changes and biological states or clinical outcomes. Subsequent investigations on radiation pneumonitis (RP) also suggest that dose response is modified by factors not related to lung dose distribution, such as dose to heart or production of proteins

that are responsible for inflammatory reactions. We built an ensemble of Bayesian Networks (BN) to represent inter-relationships between such variables. The BN was trained by fusing the observed data with prior knowledge on causal relations. In order to account for a fractionation effect on RILD, we modelled the conventional (2 Gy per fraction) and stereotactic body radiotherapy groups separately. Utility of the BN ensemble models for both groups was demonstrated in two ways: 1) robust prediction that can handle uncertainties in data, and 2) hypothesis-generating potential of the network topology that were derived from data. In conclusion, we created mathematical models for both early and late RILD that could be used for patient-specific RILD risk adaptive planning. We propose that clinical and biological risk factors should be considered in addition to dosimetry information. Finally, we advocate a use of Bayesian network as a systems radiobiology approach to combine these factors.

ABRÉGÉ

La maladie pulmonaire radio-induite (MPRI) est un effet secondaire de la radiothérapie pour le traitement des cancers thoraciques, qui limite la dose de radiations aux tumeurs et affecte les chances d'un traitement réussi. Actuellement, la prediction et la gérance du risque de MPRI est basée sur des études de population utilisant un model de probabilité de complication des tissus sains (PCTS) qui suppose que la réponse pulmonaire est la méme pour une dose de radiation donnée. Cependant, des recherches récentes suggèrent que la réponse à la dose peut être modifiée par des facteurs biologiques et cliniques pertinents à la pathogenèse de la MPRI. Ici, nous explorons une approche de biologie des systèmes pour modéliser la MPRI résultant des interactions entres ces différents facteurs. Des données cliniques, dosimétriques et biologiques de patients souffrant d'un cancer du poumon ont êtê analysées pour identifier des marqueurs associés a un risque élevé de MPRI. Puis, nous avons appliqué des méthodes d'apprentissage statistique pour combiner ces marqueurs en models qui calculent le risque de MPRI de manière spécifique pour chaque patient. Nous avons investigué deux critères d'evaluation de MPRI: la fibrose-induite par rayonnement (FR) et la pneumonite-induite par rayonnement (PR). FR est la formation de tissus cicatriciel dans les poumons et peut être mesurée de manière quantitative avec des images de tomodensitométrie. Nous avons étendu le model classique de PCTS pour modéliser explicitement le risque de FR en réponse à la dose dépendamment du temps. Nos résultats de modélisation démontrent un changement significatif de la corrélation FR-dose trois mois après la fin du traitement ainsi qu'un

risque plus élevé de FR lorsque la tumeur est située dans la region inférieure des poumons. Des recherches ultérieures sur PR suggèrent aussi que la réponse à la dose est modifiée par des facteurs qui ne sont pas reliés à la distribution de la dose dans les poumons, comme la dose donnée au coeur ou la production de protéines responsables des réactions inflammatoires. Nous avons construit un ensemble de réseaux Bayesian (RB) pour représenter les relations entre ces variables. Le RB a êtê entrainé en fusionnant les données observées avec les connaissances préalables sur les relations de causalité. Afin d'incorporer les effets de la fractionation sur MPRI, nous avons modélisé la radiothérapie conventionnelle (fraction de 2 Gy) et la radiothérapie stéréotaxique en deux groupes séparés. L'utilité des ensembles de models de RB a êtê démontrée de deux faons: 1) une prédiction robuste capable de gérer les incertitudes dans les données, et 2) le potentiel de générer des nouvelles hypothèses avec la topologie du réseau obtenu avec les données. En conclusion, nous avons crée des models mathématiques des effets précoces et tardif de la MPRI qui peut être utilisé dans la planification adaptative des risques de MPRI de manière spécifique à chaque patient. Nous proposons que les facteurs cliniques et biologiques de risques soit considérés en plus des informations dosimétriques. Finalement, nous prnons l'utilisation des réseaux Bayesian comme une approche de biologie des systèmes pour combiner ces facteurs.

ACKNOWLEDGEMENTS

I would like to begin by acknowledging my supervisor Issam El Naqa for his guidance that he carried out with his intelligence and dedication. Meeting him the first time in May 2009 - the day he delivered an interview presentation about his vision on systems radiobiology - was a turning point of my academic career. That was the time he brought me into the world of machine learning which was obscure to most of us in this field. I see myself so lucky to having conducted this truly interesting research under his direction.

Second, I would also like to thank my co-supervisor Jan Seuntjens for supporting the biomarker protocol as well as his leadership in the department which brought the CREATE program. Many thanks to Norma Ybarra for all the hard work in the wet lab and educating me with her endless knowledge in biology. Asha Jeyaseelan, my "the other lung", worked diligently for recruiting patients and setting collaboration with CHUM, which she did exceptionally well with her charm. Thank you Seema Ambereen for continuing Asha's work to keep this project progressing. So many thanks to Neil Kopek and Nathalie Japkowicz for their expert advices and commitment in the thesis committee. The CT imaging projects would have been impossible without dedication from a radiologist Mahmood AlBahhar. Thanks to collaborating doctors, Sergio Faria, Jeff Bradley, Cliff Robinson, Michael Roach, Toni Vu, Edith Filion, Marie-Pierre Campeau, Louise Lambert, Pierre Del Vecchio, for recognizing the importance of this research, providing the clinical data, and helping me with

writing manuscripts. I also benefited from interesting discussion with experts outside my group, including Joe Deasy, Jung Hun Oh and Andrew Jackson.

I acknowledge the financial supports from: Systems Biology Training Program, CREATE Medical Physics Research Training Network grant, and Excellence in Radiation Research for the 21st Century.

Thank you Margery and Tatjana for your administrative helps on scholarships, travel grants, and of course, delivering paychecks! Thank you Jessica for translating the abstract. My next and the last appreciation goes to my fellow students, and my lovely family, who have been encouraging me rain or shine.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Contribution of authors

This thesis consists of four manuscripts, two published, one submitted and one in preparation. I served as the first author for all the manuscripts by performing major works including study design, data analysis, and writing manuscripts. However, each was made possible through collaborations with various co-authors with their specific contribution stated below.

The first manuscript (chapter 4), published under the title "Analytical modelling of regional radiotherapy dose response of lung", was authored by: Sangkyu Lee (SL), Gabriela Stroian (GS), Neil Kopek (NK), Mahmood AlBahhar (MA), Jan Seuntjens (JS) and Issam El Naqa (IEN). GS and JS helped with designing the study and wrote a part of software codes for data analysis. IEN also participated in the study design. MA verified image analysis results by reviewing follow-up CT images. NK provided patient outcome data.

The second manuscript (chapter 5), in preparation under a tentative title "Texture analysis on intra-treatment CT images and and its implication on RILD heterogeneity", was contributed by Norma Ybarra (NY), Krishinima Jeyaseelan (KJ), Mahmood AlBahaar (MA), Sergio Faria (SF), Neil Kopek (NK), Pascale Brisebois (PB), and Issam El Naqa (IEN). IEN participated in the study design. NY provided biological data by performing biological assays. KJ participated in biological assays

and patient recruitment. MA reviewed CT images and participated in the study design. SF, NK, and PB collected clinical data.

The third manuscript (chapter 7), published under the title "Bayesian Network ensemble as a multivariate strategy to predict radiation pneumonitis risk", was authored by: Sangkyu Lee, Norma Ybarra (NY), Krishinima Jeyaseelan (KJ), Sergio Faria (SF), Neil Kopek (NK), Pascale Brisebois (PB), Jeffery D. Bradley (JB), Clifford Robinson (CR), Jan Seuntjens (JS), and Issam El Naqa (IEN). IEN and JS participated in the study design. NY provided biological data by performing biological assays. KJ participated in biological assays and patient recruitment. SF, NK, PB, JB, and CR collected clinical data.

The fourth manuscript (chapter 8), submitted under the title"Modeling of radiation pneumonitis after lung stereotactic body radiotherapy: a Bayesian network approach", was authored by: Sangkyu Lee, Norma Ybarra (NY), Krishinima Jeyaseelan (KJ), Sergio Faria (SF), Neil Kopek (NK), Pascale Brisebois (PB), Toni Vu (TV), Edith Filion (EF), Marie-Pierre Campeau (MPC), Louise Lambert (LL), Pierre Del Vecchio (PDV), Diane Trudel (DT), Nidale El-Sokhn (NES), Michael Roach (MR), Clifford Robinson (CR), and Issam El Naqa (IEN). IEN and CR participated in the study design. NY provided biological data by performing biological assays. KJ participated in biological assays and patient recruitment. DT and NES also took part in patient recruitment and provided clinical data. SF, NK, PB, TV, EF, MPC, LL, PDV, MR, and CR collected clinical data.

CHAPTER 1
Introduction

Contents

1.1 Radiotherapy for lung cancer

Lung cancer is currently a leading cause of cancer-related death in Canada [1]. It is also the most challenging cancer site to cure, 5-year survival rate being the lowest (17%) and far below that (63%) of average cancer patients [1]. Radiotherapy (RT), the use of high-energy ionizing radiation to kill tumour cells, is one of the

main modalities in treating cancer and considered a standard care for a subset of lung cancer patients that are not eligible for surgery [2]. RT is intended to kill tumour cells by breaking down their deoxyribonucleic acid (DNA) strands, which leads to apoptosis (spontaneous death of cells) as well as by damaging tumour microenvironment (e.g. vasculature). Depending on the location of a radiation source, RT can be classified into external beam radiotherapy (radiation is given in beams that are generated externally and penetrate through a body and a tumour) and brachytherapy (the radiation source is seeded inside a tumour). EBRT has an advantage over surgery or brachytherapy for its non-invasiveness. However, irradiation of healthy tissues along beams' path is almost unavoidable and often leads to side effects that compromise patients' quality of life. As detailed in later sections in this chapter, lung is a complex radiosensitive organ. Thus, success of treating lung cancer with EBRT hinges on finding the right balance between the benefit of delivering tumoricidal dose of radiation and the cost of side effects that it might cause.

1.2 Basic EBRT workflow

After a patient is diagnosed with a malignant cancer and need for radiotherapy is determined, a CT image around the expected treatment site is taken (*simulation CT*). Patient positioning during the scan is set up to "simulate" the position during the treatment, which bears the name of this procedure. On the CT image, a targeted tumour volume and surrounding organs/tissues at potential risk of RT side effects (*organs at risk*) are contoured into *structures*. There are four types of target volumes with different extents as defined in International Commission on Radiation Units and Measurements (ICRU) report no. 62 [3]:

- Gross tumor volume (GTV): gross demonstrable extent and location of the tumor.
- Clinical target volume (CTV): GTV + subclinical malignant disease (e.g. microscopic tumor)
- Internal target volume (ITV): CTV + internal margin factoring into uncertainties in the position of the CTV (e.g. breathing-induced tumor motion)
- Planning target volume (PTV): ITV + set-up margin to account for uncertainties in patient positioning and the behaviour of a treatment machine. It is used in treatment planning as a target volume to which the prescribed dose is conformed as well as reporting of tumour dose.

The contoured CT image is sent to a *treatment planning system (TPS)* which determines the optimal treatment beam types, shapes and positions that would deliver the prescribed dose uniformly to the target while minimizing dose spills to organs at risks (OARs). Modern TPS is equipped with dose calculation engines that estimate the expected dose distribution under a given beam setup and patient anatomy. The dose distribution is summarized into *dose-volumetric histograms (DVHs)* for tumour volumes and organs of interests. A DVH of a given structure represents distribution of its subvolumes in dose levels (figure 1.1). A DVH can be expressed either in differential or cumulative distribution. For example, one point of a cumulative DVH at dose D denotes how much percentage volume of the structure receives more than dose D (written as V_D or VD). Several metrics like VD are extracted from DVHs for quality assurance of a treatment plan. For example, sharp falloff of the cumulative

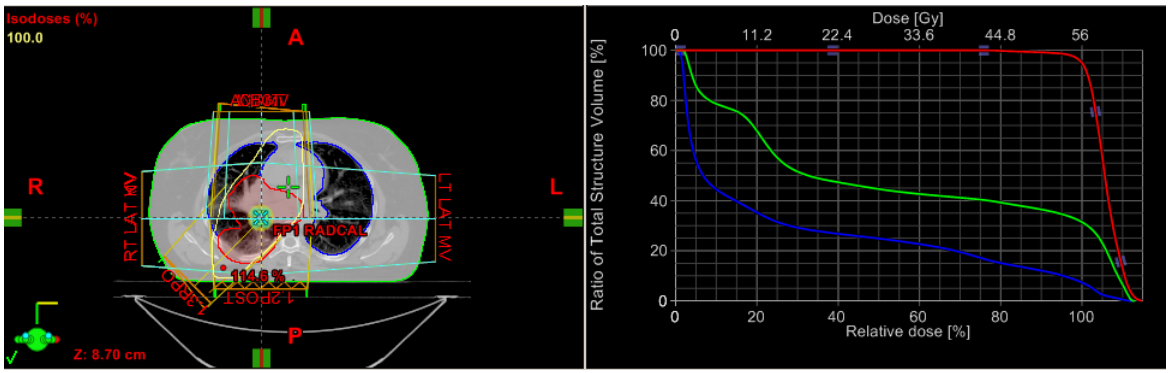

Figure 1.1: Example of a treatment plan for lung cancer. Left: an axial slice of simulation CT with superimposed PTV contour (red) and 4 treatment fields (light blue). Right: Cumulative dose volume histograms for this plan. Red: PTV, green: heart, and blue: lungs with PTV subtracted.

DVH of a PTV around a prescription dose denotes the degree of dose conformity to a tumour.

The finished treatment plan is then sent as a set of instructions to a linear accelerator (linac), a machine that generates high-energy X-ray treatment beams (fields). Success of RT depends crucially on how accurate the planned treatment is delivered to the patient. Efforts are made to match patient position and anatomy during the treatment to those at simulation as close as possible so that the planned dose distribution is reproduced during the treatment. Traditionally, this has been done by the use of marks or "tattoos" on patients' skin at the simulation: at the treatment session, these marks would be aligned with the optical light projections of treatment fields from a linac. Nowadays, dedicated on-board imaging modalities enabled image-guided radiotherapy (IGRT) where the patient images acquired at the treatment guide the fields to the desired positions. Cone-beam CT (CBCT), a widely used modality for modern IGRT, is acquired using an arm of X-ray source and a flat

panel detector that is mounted perpendicularly to a beam line. The CBCT image is then registered (spatially matched) to a planning CT, from which beam positions are calculated.

In case of lung cancer, breathing-induced target motion is taken into account when defining target volumes: simulation CT is taken at different breathing phases (4-dimensional CT or 4D-CT) from which the maximum extent of a target is delineated as an ITV. Prescription dose is then conformed to a PTV which contains a motion margin from the ITV. A gated delivery is an alternative to making a wide margin to handle motion: using real-time imaging, tumor position is constantly tracked during treatment and a beam is delivered when the target falls onto the designated location. For example, Cyberknife (Accuray Inc, Sunnyvale, CA) is a novel type of linac which is dedicated to treating small targets that are subject to motion. It is equipped with a robotic arm that can follow the movement of a target using real-time X-ray imaging.

1.3 Radiobiological models in radiotherapy

1.3.1 Basic principles

When designing radiotherapy, dose to a tumour or normal tissues is based on an empirical radiobiological model that relates planned dose distribution with expected outcomes of the therapy. There are two outcome metrics that constitute the model: *tumour control probability (TCP)* and *normal tissue complication probability (NTCP)*. TCP relates the probability of complete kill of tumour cells (tumour control) as a function of dose delivered to the tumour volume. Similarly, NTCP models dose-response of expected probability of a certain toxicity symptom (*endpoint*). The

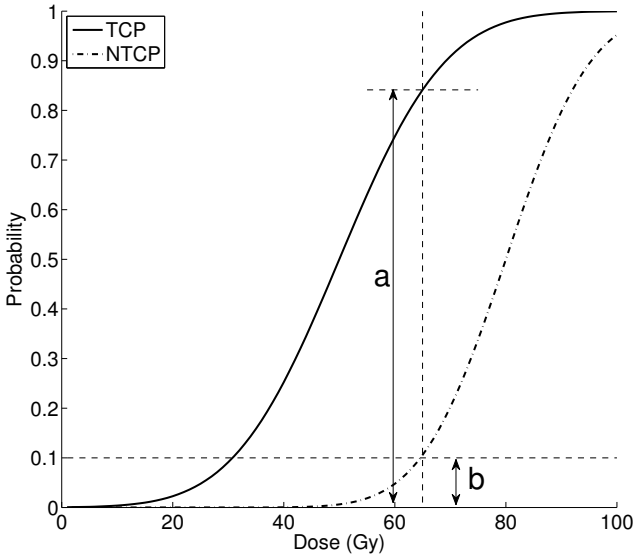

Figure 1.2: A typical sigmoid pattern shown in a hypothetical tumor control (TCP) and a normal tissue control probability (NTCP) function. The therapeutic ratio is defined as a ratio a/b.

shape of both curves is known to follow a sigmoidal function (figure 1.2). The most suitable radiotherapy maximizes the ratio of TCP to NTCP which is also known as the *therapeutic ratio*. TCP of 0.5 or higher and NTCP less than 0.05 is considered a good treatment [4].

1.3.2 Linear-quadratic model

Various patient specific and non-specific factors influence TCP, NTCP, and the therapeutic ratio. One of the most important factors is scheduling of dose delivery timing. Typically, radiotherapy is given with small fractions over a period of time (*fractionation*). The rationale behind this practice stems from an analytical model of a radiation effect on cell killing, so called the *linear-quadratic (LQ) model* [5]. This long-standing model in radiobiology was derived by fitting a linear-quadratic

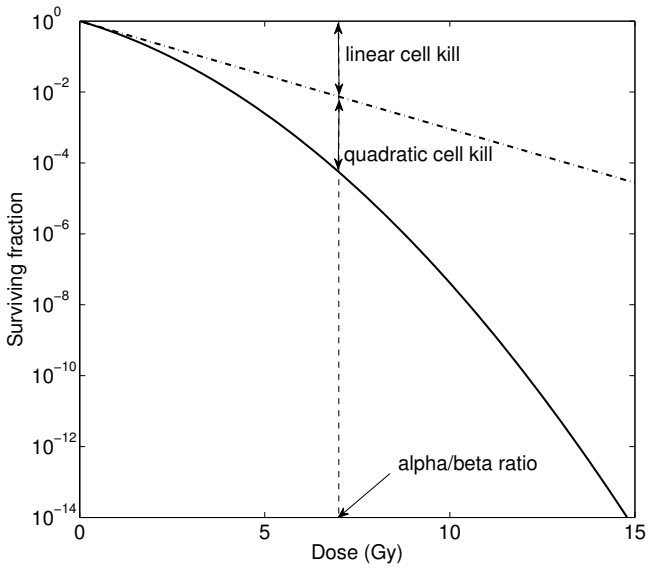

Figure 1.3: A hypothetical linear-quadratic (LQ) cell survival curve showing linear and quadratic components of cell kill.

equation to experimental data on survival of *in-vitro* cell lines to a single fraction of radiation. The model describes the log of a cell survival fraction (S) as sum of a linear and quadratic components with coefficients α and β respectively (figure 1.3, equation 1.3.2):

$$log(S) = -(\alpha D + \beta D^2)$$

$$S = e^{-\alpha D - \beta D^2}$$

If the dose D is fractionated into n sessions so that $D = nd$, the LQ-predicted survival (S_n) becomes:

$$
\begin{aligned}
log(S_n) &= -n(\alpha d + \beta d^2) \\
S_n &= e^{n(-\alpha d - \beta d^2)} \\
&= e^{(-\alpha n d - \beta n d^2)} \\
&= e^{(-\alpha D - \beta \frac{D^2}{n})} \\
&> e^{-\alpha D - \beta D^2}
\end{aligned}
$$

The non-linearity of the curve means that the cell survival as described by the LQ model would be higher when dose is more fractionated (figure 1.4). This is a prominent behaviour for normal tissues where cells are given more time to repopulate and repair sublethal damages between fractions. A ratio of the linear to quadratic parameters, α/β, determines the curvedness of the LQ curve and thus sensitivity of a tissue to fractionation (figure 1.4). It is also equal to the dose where linear and quadratic components of cell kill are equal (figure 1.3). Values of α/β vary among cell types. Normal tissues with lower α/β demonstrate a stronger fractionation effect and usually show slow radiation response later than a few months (e.g. lung, kidney). In contrast, other normal tissues with acute response (e.g. skin) and a majority of tumours tend to have higher α/β.

In contrast to normal tissues, fractionation could enhance damage to tumour cells by two mechanisms that are not accounted for by the classical LQ model: First, it allows more oxygen supplies to a tumour, which in turn creates more free radicals and thus enhances the biological effect in cell killing. Secondly, given sufficient

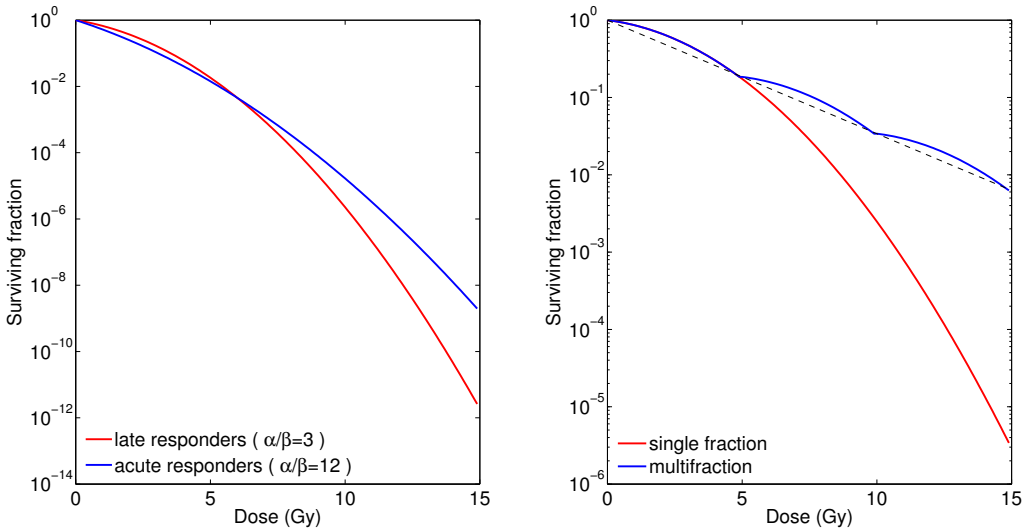

Figure 1.4: Factors that affect cell survival as explained by the linear quadratic (LQ) model. Left: Effect of an alpha-beta ratio on the shape of LQ cell survival curves. Right: Effect of fractionation on cell survival. A single dose of 15 Gy was compared to 3 fractions of 5 Gy.

time between fractionations, cancer cells are phased to the most radiosensitive cycle (reassortment). Therefore, it is generally a good strategy to increase fractionation to enhance the therapeutic ratio. Fractionation is more favourable when the value of α/β for a tumour is higher than normal tissues so that loss of TCP due to damage repair in tumour cells is outweighed by reduced toxicity risk. This is the case for lung cancer where tumour α/β is around 8-10 Gy [6] which is higher than that of radiation pneumonitis (4 Gy [7]) and fibrosis (2-3 Gy [7]). The Conventional fractionated radiotherapy (CFRT) for lung cancer prescribes 60-66 Gy in 2 Gy per fractions over 6 weeks.

Due to its simplicity, the LQ model has been widely used in preclinical research as well as clinical study design, especially for stereotactic radiotherapy (see section

1.4), for comparing dose between two different fractionation schemes. A quantity called *biologically effective dose (BED)* converts a nominal prescription dose to the equivalent that represents cell killing potential as predicted by the LQ model (equation 1.1):

$$BED = nd[1 + d/(\alpha/\beta)] \tag{1.1}$$

with n: number of fractions, d: nominal dose to a target per fraction.

Equivalent dose in 2 Gy fraction (EQD2), or normalized total dose (NTD) in old literatures, can be used instead of BED to estimate the equivalent biological effect if the fraction size were 2 Gy:

$$EQD_2 = nd(\frac{d + \alpha/\beta}{2 + \alpha/\beta}) \tag{1.2}$$

Despite its widespread use, the LQ model is under several criticisms for its over-simplistic assumptions and inaccuracy at high dose per fraction. The main limitation comes from the fact that it was derived from in-vitro cell survival data without modelling of other mechanisms that affect cell killing such as damage to microvasculature [8] or stroma (connective tissues providing mechanical support to a tumour) [9]. This could account for underestimation of survival by the LQ model in a high dose area (above 10-12 Gy) [10] where these mechanisms are triggered [11]. This inaccuracy might have ramifications in designing and evaluating SBRT regimen where a fraction size often exceeds 10 Gy. There has been alternative models that attempted to extend the LQ model to the high dose region. For example, Guerrero and Li [10] introduced into the LQ formula an additional term that bends the curve

linearly instead of quadratically at high dose in order to improve the fitting with empirical data.

1.4 Stereotactic body radiotherapy

Despite the therapeutic advantages of fractionation as explained in section 1.3, prescription of the smaller number of fractions (5 or less) with larger dose per fraction has been practiced to treat early stage (I and II) cancers. This regimen, called stereotactic body radiotherapy (SBRT), was inspired by clinical success of stereotactic radiosurgery (SRS) on brain tumours where a dose of 15-25 Gy is given in a single or 2 fractions. In recent years, SBRT has become a standard care for non-operable early stage NSCLC, demonstrating a local control rate close to 90% and limited toxicity rate [12]. Damage to tumour vasculature is an important mechanism of tumour killing by SBRT, which is not accounted by the LQ model. It has been shown that a single fraction of dose more than 10 Gy induces damage to blood vessels, leading to starvation and secondary killing of downstream tumour cells [13].

SBRT increases BED for both tumours and normal tissues due to the reduced number of fractions (equation 1.1). However, increased BED for lung does not raise a serious concern for RILD as long as treatment field sizes are reduced to match that of the tumours that are usually small in volume (typically less than 100 cc). Lung is thought to possess a physiological reserve capacity that compensates for a partial loss of its tissues [7]. As a result, lung displays a strong volume effect, meaning that RILD severity strongly depends on the irradiated volume (detailed in chapter 2). Rather, lung SBRT poses risks for potentially lethal side effects on other thoracic organs,

especially central structures such as great vessels (e.g. aorta) or central bronchus [14].

In order to restrict the volume effect on RILD risk, it is important to focus radiation to the small target while minimizing dose spills to normal tissues. Here are 3 most widely used delivery techniques for lung SBRT:

- 3-dimensional conformal radiotherapy (3D-CRT): Treatment consists of 7-10 coplanar fields aiming at a target from different angles. Each field is conformed to the cross-section of the PTV at the angle of the field using multi-leaf collimators (MLC) which are high atomic number materials on a linac head that blocks the radiation to shape the beam.
- Volumetric Arc Therapy (VMAT): A linac head rotates in an arc with a tumour at its focal point, and radiation is delivered during the rotation. It does away with the dead time that would be consumed by 3D-CRT while switching a linac head position.
- CyberKnife: A robotic non-coplanar delivery system that is dedicated to irradiating small targets.

1.5 Basic lung anatomy and histology

Human lung has two parts, left and right lung, connected by a main airway (figure 1.5). Each lung is divided into lobes: right lung into three (superior or upper, middle, inferior or lower) and left lung into two (upper, lower). An airway or a respiratory tract is a passage conducting air from the nose or mouth to lungs. An airway first branches at the end of trachea, also known as the carina, into two primary bronchi which enter left and right lung. The bronchi further branch into

finer tubes called bronchioles in several stages. The endings of the bronchioles join lung parenchyma which are structural units in lung tissues where air exchange occurs. Lung parenchyma consist of alveoli (singular: alveolus), alveolar ducts, and capillary beds. *Alveoli* are air sacs that correspond to the terminal ends of the respiratory tree. A thin membrane of the alveoli ($\sim$0.2 μm) enables gas diffusion from alveoli atrium into the bloodstream in the capillary bed surrounding the alveoli. An empty space between parenchyma, intra-alveolar septum, is filled with stromal cells such as extracellular matrix (ECM) that provide a mechanical support to the alveoli.

Alveoli are composed of two types of cells: type I and II pneumocytes (figure 1.6). Type I pneumocytes are epithelial cells that cover a vast majority (90-95%) of the surface of an alveolar sac, which is a main site for gas exchange. Type II cells take up a small fraction of alveolar surface ($\leq$5%) but have several important functions: they produce pulmonary surfactants which decrease surface tension within alveoli. They can also differentiate into type I pneumocytes in case of structural damage to alveoli. Alveoli also home some macrophages (dust cells) which help neutralize pathogenic particles entering alveoli.

1.6 Radiation induced lung disease

External radiotherapy inevitably irradiates healthy tissues surrounding a tumour. In cases where treatment sites are located in thorax such as lung, esophagus, breast, or thyroid, lung is subject to the non-negligible risk of radiation induced side effects. Since lung functions heavily affect patients' survival and quality of life, minimizing dose to lung is a one of the major objectives for thoracic RT planning and delivery.

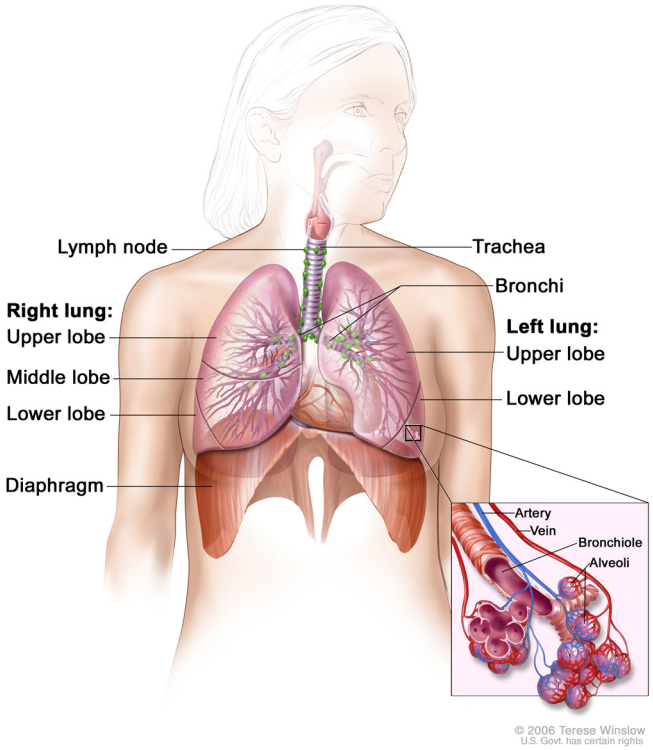

Figure 1.5: Basic anatomy of the respiratory system (with permission from Terese Winslow).

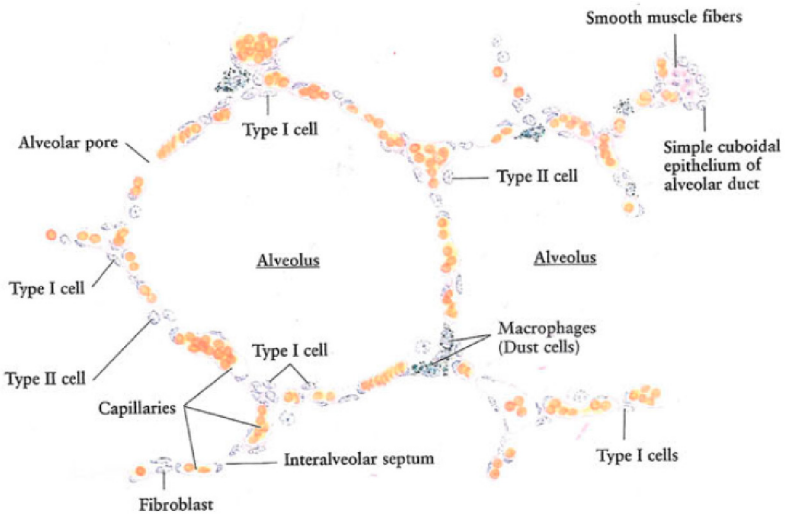

Figure 1.6: Histology of lung parenchyma (with permission from Zhang [15]).

Radiation induced lung disease (RILD) is a broad range of symptoms and manifestations that are caused by microscopic changes in lung (figure 1.8). Radiotherapy-related lung diseases develop in two phases - acute (radiation pneumonitis) and late/chronic (radiation fibrosis) - with different clinical presentations:

- Radiation pneumonitis (RP) is an acute inflammatory lung disease that accompanies fever, cough and dyspnea (difficulty in breathing). It typically occurs 1-3 months or before 10 months after RT [16] with a few weeks of latency period following irradiation [17]. RP is potentially life threatening and often requires administration of steroids to appease the symptoms. Non-symptomatic RP can be observed radiographically as a diffuse change in the irradiated region, also referred to as ground glass opacity [18] (figure 1.7).
- Radiation fibrosis (RF) refers to formation of fibrotic tissues in the area of high radiation dose regions. Fibrosis tends to develop later than RP (months

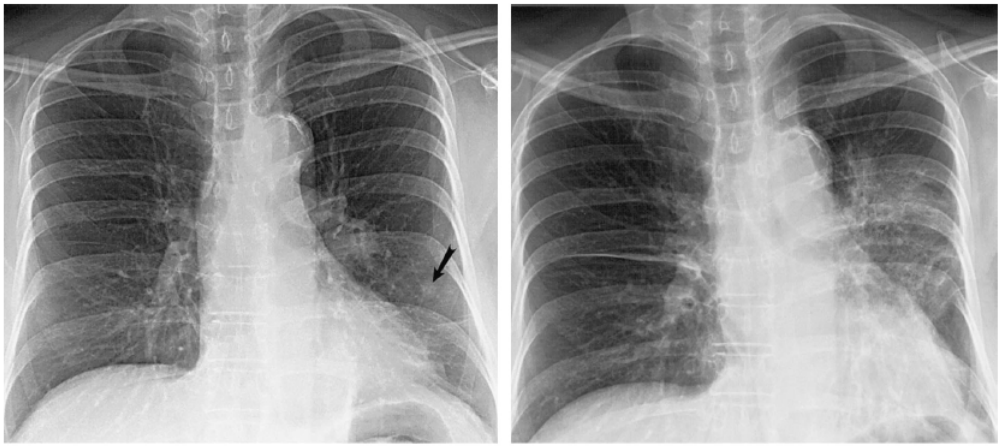

Figure 1.7: Ground-glass opacity from radiation pneumonitis. Left: pre-RT chest radiograph with a tumour mass (arrow), B: Chest radiograph obtained 2 months after completion of RT shows ground-glass opacity adjacent to the tumour site. With permission from [23]

to years after the treatment) and leaves permanent scars in lung, which is seen in CT or radiographs as high-density areas. Although occurrence of RF is known to correlate with RP [19], a majority of RT-treated lung cancer patients show some degree of radiological changes without symptoms [20]. Still, severe loss of air exchange space by fibrotic tissues impairs gas exchange and can be clinically observable as a low oxygen level in blood (hypoxemia). Moreover, fibrosis tends to develop in a high dose region near a tumour [18] [21] which can be confused with tumour relapse and thus poses a challenge to accurate diagnosis of local control [22]. Examples for different patterns of RF can be seen in figure 1.9.

1.6.1 Clinical grading of RILD

There are a number of standards for grading treatment-induced side effects. This study follows a widely-used Common Toxicity Criteria for Adverse Events (CTCAE)

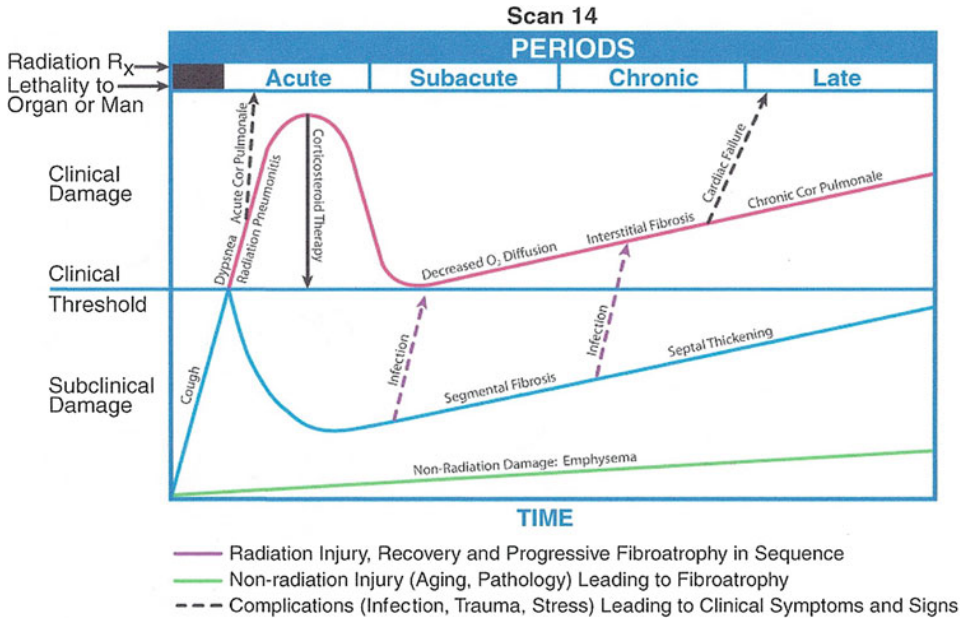

Figure 1.8: Continuum of clinical and subclinical manifestations of RILD. With permission from Rubin et al. [24].

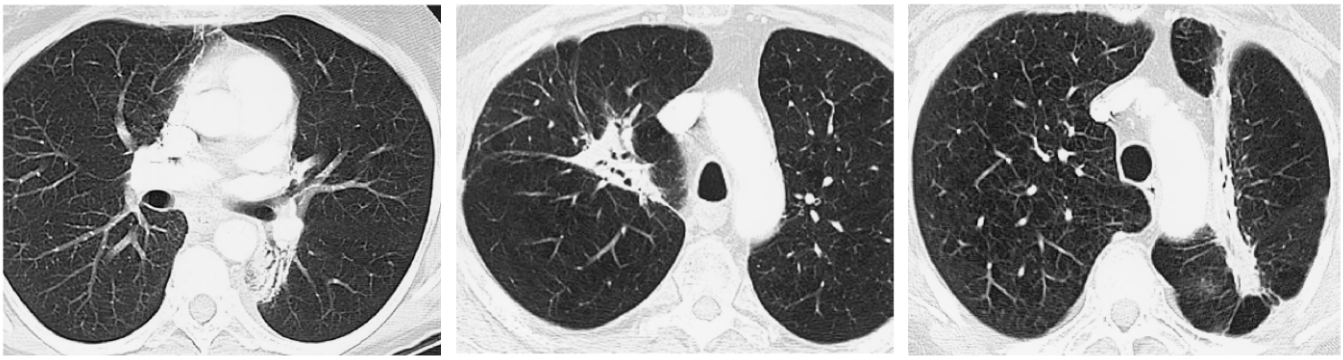

Figure 1.9: Three distinct patterns of radiation fibrosis in computed tomography images. Left: modified conventional pattern, 5 months post-RT. Also shown are volume loss, consolidation, and air bronchograms (air-filled bronchi made visible by increased density in adjacent alveoli. Middle: masslike pattern in right lung, 14 months post RT. Right: scarlike pattern, 5 years post-RT. Band-like consolidation can be seen across left lung. Reproduced with permission from [23].

Table 1.1: CTC scoring criteria for RP and fibrosis (version 4)

grade	Pneumonitis	Fibrosis
1	Asymptomatic; clinical or diagnostic observation only; intervention not indicated	Mild hypoxemia; radiologic pulmonary fibrosis <25% of lung volume
2	Symptomatic; medical intervention indicated; limiting instrumental activities of daily living	Moderate hypoxemia; evidence of pulmonary hypertension; radiographic pulmonary fibrosis 25 - 50%
3	Severe symptoms; limiting self care daily life; oxygen indicated	Severe hypoxemia; evidence of right-sided heart failure; radiographic pulmonary fibrosis > 50-75%
4	Life-threatening respiratory compromise; urgent intervention indicated (e.g. tracheotomy or intubation)	Life threatening consequence; intubation with ventilation support indicated; radiographic fibrosis > 75%
5	Death	Death

version 4 which stratifies the severity of RP and RF (table 1.1). Clinical definition of toxicity events and reported event rates vary amongst clinical studies depending on the usage of different toxicity criteria or cutoff grade. Rates of grade 2 or higher RP requiring therapeutic intervention from conventional fractionation RT range from 5 to 50% [16]. SBRT is known to decrease the risk of symptomatic RP (see section 1.4): reported rates are between 9% and 28% . Compared to symptomatic pneumonitis, episodes of radiographic change are common: incidence rate of grade 1 or higher RF is approximately 50% from conventional [25] and 60-80% from SBRT [26].

1.7 Paradigm of personalized radiotherapy and RILD

RILD may not always threaten patients' survival by its symptoms, but it does so indirectly by limiting prescribed dose to tumour and thereby limiting the chance

of the success of RT. A trial from the University of Michigan on dose escalation, an increase in a tumour dose over a conventional level, showed a promising result of improved survival [27]. In this trial, dose was escalated in different amounts for each patient according to their predicted NTCP [28]; The patients with lower expected toxicity received more tumour dose in order to enhance the chance of survival. Thus, it is very important to have accurate and quantitative NTCP prediction to design a trial that might establish a new treatment without additional morbidity.

However, the NTCP model that they used was derived from a population-based study, which assumes the same expected toxicity if dose were the same across all the patients. So, when it comes to designing a radiotherapy plan, we have not realized "personalization" in its authentic sense. As seen in later chapters, a population-based RILD risk model has its limit in accuracy due to other patient-specific factors that influence the vulnerability of patients. Challenges remain in discovering those factors, testing their robustness and incorporating them into a mathematical model that could be served as one element for RT plan optimization.

1.8 Thesis objectives and organization

Our main hypothesis is that prediction of RILD risk can be improved if we take into account and model patient-specific factors that modify dose-response relationships. The efforts to test this hypothesis can be summarized into the two specific objectives:

1. Exploration of a large amount of patient data to identify relevant variables that can capture biological and physical characteristics of RILD.

2. Combination of such variables into a mathematical model that calculates patient-specific RILD risk.

The thesis is organized into the following chapters: the next chapter will summarize radiobiological NTCP models in the literature and their applications to RILD. The chapter will also discuss patient specific factors that can extend the classical dosimetric NTCP models. Chapter 3 will elaborate on biological RILD risk factors and also biological mechanism of RILD which provides a rationale for the biomarkers. In chapter 4, applicability of a classical NTCP approach will be tested in the case of the dose response of radiation fibrosis measured by CT imaging. Chapter 5 will then attempt to explain patient heterogeneity in the dose response of the CT change. Then we will switch attention from describing the response to predicting the response, with a specific goal of predicting symptomatic radiation pneumonitis from patient specific factors. Chapter 6 will go over computational approaches used in the later chapters to build RILD prediction models from multitudes of factors. The following 2 chapters will illustrate the clinical application of such multivariate modelling approach to two different fractionation groups: CFRT (chapter 7) and SBRT (chapter 8). Chapter 9 will summarize and discuss the limitations of this work as well as potential future directions.

REFERENCES

[1] Canadian Cancer Society, Canadian cancer statistics 2015, Tech. rep. (2015).

[2] Treatment of non-small cell lung cancer, http://www.cancer.ca/en/cancer-information/cancer-type/lung/treatment/?region=on, accessed: 2016-01-02.

[3] International Commission on Radiation Units and Measurements (ICRU), ICRU Report 60: Prescribing, Recording, and Reporting Photon Beam Therapy (Supplement to ICRU Report 50), ICRU, Bethesda, MD, 1999.

[4] E. B. Podgorsak, Radiation Oncology Physics: A Handbook for Teachers and Students, International Atomic Energy Agency, Vienna, Austria, 2005.

[5] B. G. Douglas, J. F. Fowler, The effect of multiple small doses of x rays on skin reactions in the mouse and a basic interpretation, Radiation research 66 (2) (1976) 401–426.

[6] E. Hall, A. Giaccia, Radiobiology for the Radiologist, Lippincott Williams & Wilkins, 2006.

[7] J. B. S. M. Bentzen, J. Z. Skoczylas, Quantitative clinical radiobiology of early and late lung reactions, International Journal of Radiation Biology 76 (4) (2000) 453–462.

[8] Z. Fuks, R. Kolesnick, Engaging the vascular component of the tumor response, Cancer Cell 8 (2) (2005) 89–91.

[9] B. F. Schneider, D. A. Eberhard, L. E. Steiner, Histopathology of arteriovenous malformations after gamma knife radiosurgery, J Neurosurg 87 (3) (1997) 352–7.

[10] M. Guerrero, X. A. Li, Extending the linear-quadratic model for large fraction doses pertinent to stereotactic radiotherapy, Physics in Medicine and Biology 49 (20) (2004) 4825.

[11] J. P. Kirkpatrick, J. J. Meyer, L. B. Marks, The linear-quadratic model is inappropriate to model high dose per fraction effects in radiosurgery, Seminars in Radiation Oncology 18 (4) (2008) 240–3.

[12] T. Xia, H. Li, Q. Sun, et al., Promising clinical outcome of stereotactic body radiation therapy for patients with inoperable stage I/II non-small-cell lung cancer, International Journal of Radiation Oncology*Biology*Physics 66 (1) (2006) 117 – 125.

[13] C. W. Song, L. C. Cho, J. Yuan, K. E. Dusenbery, R. J. Griffin, S. H. Levitt, Radiobiology of stereotactic body radiation therapy/stereotactic radiosurgery and the linear-quadratic model, International Journal of Radiation Oncology*Biology*Physics 87 (1) (2013) 18–19.

[14] R. D. Timmerman, An overview of hypofractionation and introduction to this issue of seminars in radiation oncology, Seminars in Radiation Oncology 18 (4) (2008) 215 – 222.

[15] S. Zhang, An Atlas of Histology, Springer, 1999.

[16] L. B. Marks, S. M. Bentzen, J. O. Deasy, et al., Radiation dose-volume effects in the lung, International Journal of Radiation Oncology*Biology*Physics 76 (3, Supplement 1) (2010) S70 – S76.

[17] J. P. Williams, C. J. Johnston, J. N. Finkelstein, Treatment for radiation-induced pulmonary late effects: spoiled for choice or looking in the wrong direction?, Curr Drug Targets 11 (11) (2010) 1386–94.

[18] T. R. Koenig, R. F. Munden, J. J. Erasmus, et al., Radiation injury of the lung after three-dimensional conformal radiation therapy, Am. J. Roentgenol. 178 (6) (2002) 1383–1388.

[19] V. Moiseenko, T. Craig, A. Bezjak, J. V. Dyk, Dose-volume analysis of lung complications in the radiation treatment of malignant thymoma: a retrospective review, Radiotherapy and Oncology 67 (3) (2003) 265 – 274.

[20] S. Faria, M. Aslani, F. Tafazoli, L. Souhami, C. Freeman, The challenge of scoring radiation-induced lung toxicity, Clinical Oncology 21 (5) (2009) 371 – 375.

[21] T. Aoki, Y. Nagata, Y. Negoro, et al., Evaluation of lung injury after three-dimensional conformal stereotactic radiation therapy for solitary lung tumors: Ct appearance, Radiology 230 (1) (2004) 101–108.

[22] K. Huang, M. Dahele, S. Senan, et al., Radiographic changes after lung stereotactic ablative radiotherapy (SABR) - can we distinguish recurrence from fibrosis? a systematic review of the literature, Radiotherapy and Oncology 102 (3) (2012) 335 – 342.

[23] Y. W. Choi, R. F. Munden, J. J. Erasmus, et al., Effects of radiation therapy on the lung: radiologic appearances and differential diagnosis, Radiographics 24 (2004) 985–997.

[24] P. Rubin, L. Constine, L. Marks, ALERT: Adverse Late Effects of Cancer Treatment: Volume 2: Normal Tissue Specific Sites and Systems, Medical Radiology, Springer, 2013.

[25] F.-M. Kong, J. A. Hayman, K. A. Griffith, et al., Final toxicity results of a radiation-dose escalation study in patients with non-small-cell lung cancer (NSCLC): Predictors for radiation pneumonitis and fibrosis, International Journal of Radiation Oncology*Biology*Physics 65 (4) (2006) 1075 – 1086.

[26] E. X. Huang, A. J. Hope, P. E. Lindsay, et al., Heart irradiation as a risk factor for radiation pneumonitis, Acta Oncologica 50 (1) (2011) 51–60.

[27] F.-M. Kong, R. K. Ten Haken, M. J. Schipper, et al., High-dose radiation improved local tumor control and overall survival in patients with inoperable/unresectable non-small-cell lung cancer: Long-term results of a radiation dose escalation study, International Journal of Radiation Oncology*Biology*Physics 63 (2) (2005) 324 – 333.

[28] R. K. Ten Haken, M. K. Martel, M. L. Kessler, et al., Use of veff and iso-ntcp in the implementation of dose escalation protocols, International Journal of Radiation Oncology*Biology*Physics 27 (3) (1993) 689 – 695.

CHAPTER 2
Current status of normal tissue complication probability modelling

Contents

2.1 Introduction

This chapter will review a brief history and current status of modelling normal tissue probability (NTCP) and application of the NTCP principle to RILD.

NTCP models, like other disease models in general, can be divided into two categories which have a major philosophical difference in where models should originate from:

Phenomenological models. A model is designed and fitted in a way that it best describes empirical data, independent of theories or assumptions. Phenomenological models are generally simple and easy to implement. On the other hand, it is risky to extrapolate the model outside the empirical data from which the model parameters were derived from. Inaccuracy of the linear-quadratic model (section 1.3.2) in large fraction dose serves as a good example.

Mechanistic models. A model is built by fitting data to pre-specified underlying processes. Parameters of the model have biological/physical representations. Compared to phenomenological models, it takes more effort to design and implement. Especially, it is the case for any model that has to do with biological phenomena with high complexity and incomplete knowledge. Nevertheless, the model tends to be more general and robust to extrapolation due to its solid theoretical underpinnings.

Otherwise, NTCP models can be classified in terms of the scope of risk factors:

Dosimetric models. Historically, NTCP has been perceived as a population-averaged relationship between dose to an organ of interest and risk of complication to the organ. As dose calculation is part of treatment planning processes, NTCP

can readily be computed using the calculated 3D dose distribution or DVHs. Modern radiotherapy generates highly heterogeneous dose distribution thanks to inverse plan optimization and conformal delivery techniques such as multi-leaf collimators. Dosimetric NTCP models differ on how to convert this heterogeneity into a surrogate index that correlates best with outcomes. Most clinical NTCP models, such as the ones for biological based treatment planning system (BBTPS), are designed phenomenologically, while there are efforts to adopt a more mechanistic approach [1]. Dosimetric NTCP models have been incorporated into treatment planning and protocol design, often in the form of *dose constraints* that serve as an upper limit on planned dose to the organ.

Non-dosimetric models. Recently, researchers are beginning to extend conventional NTCP models with non-dosimetric parameters to identify more radiosensitive subpopulation. However, most of these extended models remain in research and have yet to make clinical application. The biggest obstacle would be availability of data for such non-dosimetric markers, as they may not be collected as routinely as dosimetric data. Hopefully, advances in functional imaging and biological assays are expected to facilitate external validation of these models, which is a necessary (but not sufficient) step for research translation.

This chapter is organized as follows: first, perspectives on general dosimetric NTCP modelling approaches, either phenomenological or mechanistic, will be discussed. Then, past results on modelling radiation pneumonitis (RP) and fibrosis (RF) using the dosimetric NTCP approaches will be presented. In addition, non-dosimetric RILD factors and recently developed models will be mentioned.

2.2 NTCP modelling approaches

This section will review two most commonly used dosimetric NTCP models.

2.2.1 Lyman-Kutcher-Burman (LKB) model

The LKB model [2] [3] [4] is a phenomenological model that relates 3D dose distribution to NTCP via a multi-parameter sigmoid function. It was initiated by a collective effort by Emami et al. [5] to create a compilation of toxicity reports in the literature that could serve as a guideline for treatment planning. Normal tissue toxicity risk was expressed in terms of $TD_{50/5}$ (tolerance dose at 50% chances of complication within 5 years) and $TD_{5/5}$ (tolerance dose at 5% chances of complication within 5 years). The toxicity data compilation included 28 critical organs in adult irradiated to conventionally fractionated (180-200 cGy per day/5 days a week) radiotherapy. The two tolerance dose values were assigned to each organ and each one of the three partially irradiated volumes (one-third, two-thirds, and the whole organ). Burman et al. [4] fitted the three-parameter model that was initially conceived by Lyman [2] to the toxicity data by Emami et al, where complication probability (NTCP) is expressed as a function of a single dose (D) that is given uniformly to a partial volume (v) (equation 2.1, 2.2, 2.3):

$$NTCP = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{(-x^2/2)} dx \tag{2.1}$$

$$t = \frac{D - TD_{50}(v)}{m \cdot TD_{50}(v)} \tag{2.2}$$

$$TD_{50}(v) = TD_{50}(1) v^{-n} \tag{2.3}$$

The model has 3 free parameters to be determined from observed data using maximum likelihood fitting: $TD_{50}(1)$ represents a tolerance dose at which complication rate reaches 50% when it is given uniformly to the entire organ/tissue. m indicates how steep the NTCP rises with increasing dose. n governs sensitivity of the NTCP to irradiated partial volume. Later, it was adapted to heterogeneous dose distribution via a histogram reduction technique by Kutcher et al. [3]. It converts a non-uniform dose volume histogram into the uniform equivalent of a single dose D to a volume v which then can be used as input parameters for the LKB model (equation 2.2 and 2.3). There are two different reduction schemes: first, a DVH can be expressed as a maximum dose in the distribution D to a effective volume v (equation 2.4):

$$v = \sum_i (\frac{D_i}{D})^{\frac{1}{n}} v_i \tag{2.4}$$

where v_i is a partial volume in a differential DVH dose bin D_i.

Another choice is to find an equivalent dose D_{eff} to the whole organ ($v = 1$) (equation 2.5):

$$D_{eff} = [\sum_i v_i (D_i)^{\frac{1}{n}}]^n \tag{2.5}$$

The generalized equivalent uniform dose (gEUD) [6] is a single-parameter model that has the same functional form as the equation 2.5, except that the volume parameter n is replaced by its reciprocal $a = 1/n$ (equation 2.6). The gEUD was proposed as a metric that summarizes non-uniform dose distribution into a uniform equivalent that could be used for optimizing treatment plans.

$$gEUD = (\sum_{i=1}^{N} v_i D_i^a)^{\frac{1}{a}} \tag{2.6}$$

The value of a governs which region of a DVH is weighted more to the gEUD: for negative a, a low dose part of the DVH counts more; for large positive value of a, the gEUD is correlated more to a high dose region in the DVH; when a is close to 1, every DVH bin is weighted the same and the gEUD approaches an arithmetic average dose. Figure 2.1 demonstrates the influence of a on sensitivity to toxicity (tolerance dose) to irradiated volume. The value of a found for lung in literatures is close to unity (1.03) [7], indicating strong volume dependence of RILD. This also means that MLD and gEUD are fairly close and exchangeable.

The gEUD is considered a parameter that is more clinically relevant than traditional DVH parameters. Some commercial TP softwares incorporated the gEUD into dose constraints for inverse planning. It was suggested that gEUD-optimized treatment plans improved normal tissue sparing in the optimized plan [9]. However, the International Commission on Radiation Units and Measurements (ICRU) report guarded against using gEUD for treatment optimization parameter without accurate knowledge on the parameter a [10].

2.2.2 Critical element and volume model

The critical element and volume model [11] [12] [13] is a mechanistic model that explains an onset of complication based on two main assumptions:

- An organ is composed of *functional subunits* (FSUs) which are the smallest unit of tissue elements that are structurally or functionally discrete.

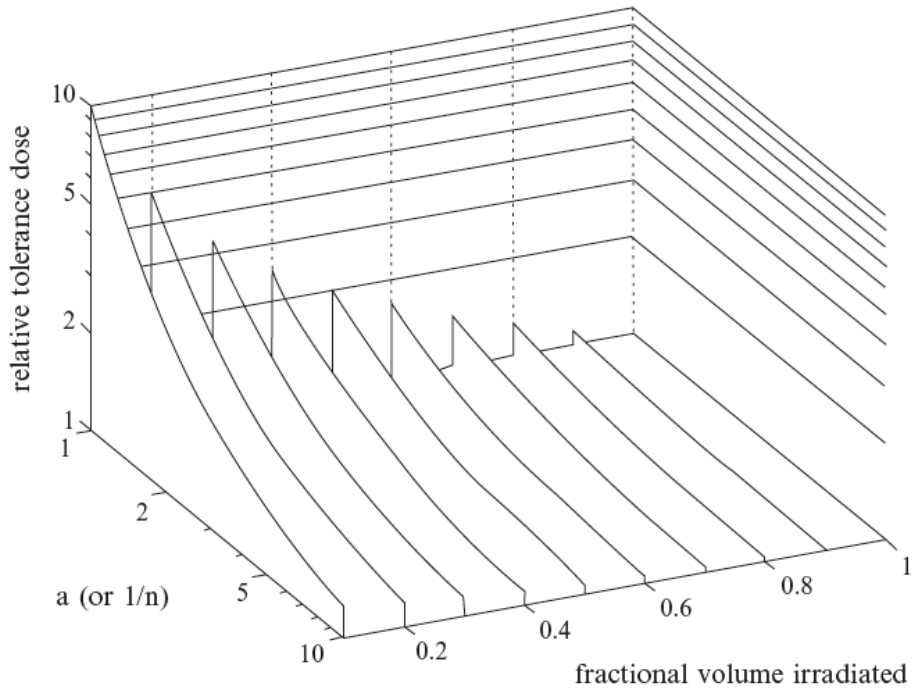

Figure 2.1: Dependence of relative tolerance dose on fractional volume irradiated and the parameter a of the gEUD model. It can be seen that the toxicity risk increases rapidly with volume when a close to 1, whereas the risk is rather insensitive to volume at larger a. Permission from Moiseenko et al. [8]

- The organ develops complication after it loses more than a critical number of FSUs.

The concept of FSUs was developed by Withers et al. [11] who created two different definitions depending on tissue types:

- Structurally defined FSUs: An FSU can be defined as a structural element that makes up an organ such as nephrons for a kidney and acini for lung.
- Structurally undefined FSUs: An FSU is defined as the maximum area or volume that can be repopulated by one clonogen rather than visible structures. This definition of FSUs is used for describing the organization of tissues such as skins, mucosa, or glial tissues.

Modelling NTCP using the critical volume approach is a two-step process: First, probability of a single FSU damage (P_{FSU}) is related to a local dose (d) using a phenomenological sigmoidal model (equation 2.7)

$$P_{FSU}(d) \quad = \quad 1/[1 + (d_{1/2}/d)^k] \tag{2.7}$$

where $d_{1/2}$ is the dose at which 50% of the FSUs are damaged and k is a parameter for the steepness of the sigmoidal curve. It is assumed that one FSU is small enough and dose distribution to a unit is nearly uniform.

The next step relates NTCP to the number of damaged FSUs. This relationship depends on arrangement of FSUs in an organ or whether they are connected in series (e.g. a spinal cord, optic nerves) or parallel (e.g. kidney, lung). The NTCP model for serial FSU arrangement is specifically called the *critical element model* [12]. It

assumes that complication arises when at least one FSU is damaged. When an organ receives a heterogeneous dose distribution with dose values $\{d_i\}$ to fractional volumes $\{v_i\}$, the NTCP for that organ is then expressed as follows (equation 2.8):

$$NTCP = 1 - \prod_i [1 - P_{FSU}(d_i)]^{v_i} \tag{2.8}$$

For the organs where FSUs are connected in parallel (the *critical volume model*), complication is assumed to occur when more than a certain fraction of FSUs, or *the functional reserve*, are damaged [14]. This model first calculates the damaged fraction of FSUs, f_{dam}, by integrating P_{FSU} over an organ:

$$f_{dam} = \sum_i v_i P_{FSU}(d_i) \tag{2.9}$$

When deriving NTCP as a function of f_{dam}, it is considered that every patient has different functional reserves, which is modelled as a gaussian distribution with a mean v_{50} and standard deviation σ_v^2. Then, the NTCP is expressed as a cumulative probability distribution of a functional reserve: in other words, NTCP is the probability that a patient's functional reserve is smaller than a damaged fraction f_{dam} (2.10):

$$NTCP(f_{dam}) = \frac{1}{\sqrt{2\pi\sigma_v^2}} \int_0^{f_{dam}} exp[-\frac{(v - v_{50})^2}{2\sigma_v^2}]dv \tag{2.10}$$

The biggest difference between critical element and critical volume NTCP lies in a response to volume at fixed dose. The critical element NTCP shows linear increase with volume. This can be explained in a pure probabilistic view: a chain-like

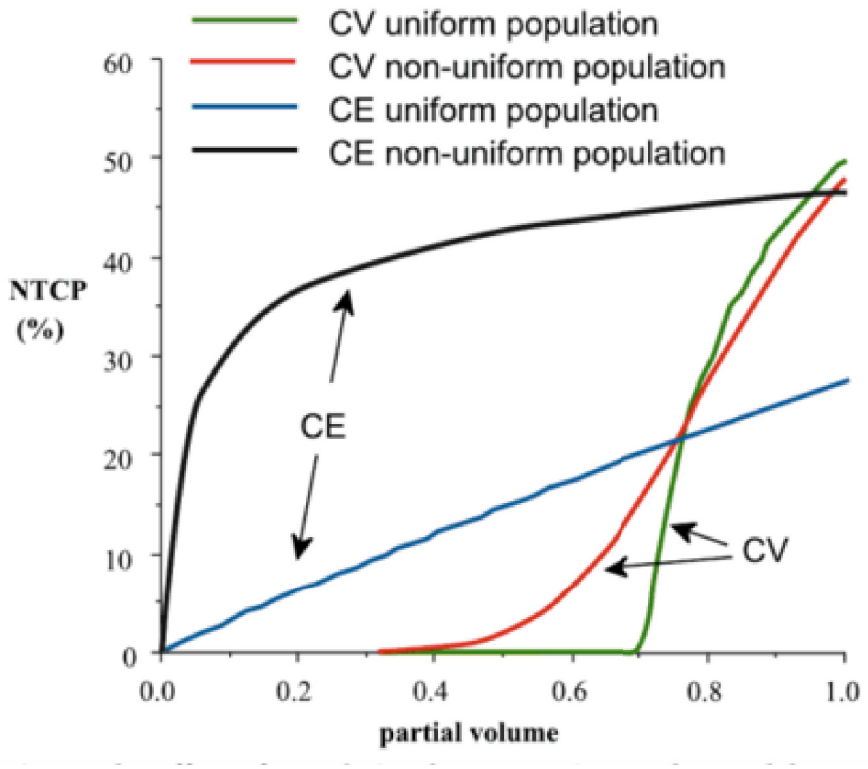

Figure 2.2: Comparison in volume response between critical element and critical volume NTCP models. An effect of population spread in model parameters were tested. Permission from [15].

structure is more likely to break if larger number of links are covered by radiation. In contrast, the critical volume response exhibits a threshold behaviour, which originates from a concept of a function reserve (figure 2.2).

The other category of tissue, such as skin or mucosa, is organized by a clump of independent FSUs aligned in a 3-dimensional array. This category called a graded response displays a continuous scale of response rather than binary-type events. Unlike the critical volume or element models, the graded response NTCP is known

Type	Description	Examples	Model
Critical Element (CE)	A complication is caused by irreparable damage to any of FSUs	spinal cord, nerves, peritoneum	Chain made up of many friable links
Critical Volume (CV)	Damage to the fraction of FSUs above functional reserve leads to a complication	kidney, liver, lung	Rope made of of many strands which will hold until many are broken
Graded Response (GR)	Response occurs on a continuous scale	skin, mucosa	Granula clump of "dosimeters"

Table 2.1: Three main types of tissue architectures described in the critical volume model. Reproduced with permission from [15]

to be independent of irradiated volume [15]. Table 2.1 summarizes the modelling of tissue organization and radiation response for each tissue type.

2.3 DVH-based RILD models

2.3.1 RP models

When it comes to RP for an endpoint, it is consensus that lung is a parallel organ: the earliest estimate by Burman et al. [4] on the volume parameter n for the LKB model was 0.87. A more recent pooled analysis by Marks et al. in 2010 [7] points to a value that is closer to 1 (mean: 1.03, 95% confidence interval: 0.67-1.39). This has two implications: first, an arithmetic average dose to lung or mean lung dose (MLD) is close to gEUD and thus serves as a fair average for non-uniform dose distributions. Second, as explained by the critical volume model (section 2.2.2), a parallel organ is expected to exhibit a threshold behaviour, where NTCP sharply increase beyond a certain fractional volume. Such volume is defined using a threshold

dose, in other words, as a percentage of volume of lung receiving more than a threshold dose (VDth). These two types of parameters, MLD along with VDth at various threshold dose, were investigated for their predictive power by numerous studies.

In 2010, a collective effort was made by the American Society for Radiation Oncology (ASTRO) to compile the published dose-volume studies on various organs. This work, called the Quantitative Analysis of Normal Tissue Effects in the Clinic (QUANTEC) [17], was aimed at updating the toxicity guidelines by Emami et al. [5] in wake of evolving dose delivery techniques that create more complex dose distributions. The QUANTEC report for lung reviewed more than 70 published clinical studies on "standard" radiotherapy - conventionally fractionated 3D-conformal photon treatments - excluding special techniques such as SBRT or IMRT (figure 2.3). The MLD models turned out to be fairly consistent across the studies, showing strong correlation with RP risk. This agreed with a preceding multi-institutional study by Kwa et al. [18]. However, it can be seen from figure 2.3 that RP risk increases slowly with MLD. As a result, it would be difficult to define a safe threshold for MLD, although the metric is still useful for relative toxicity assessment. Moreover, with regards to the VDth studies, the QUANTEC found inconsistency in which threshold dose correlates best with RP. These VDth parameters are highly correlated to each other [19], and the best predictive parameter depends on delivery techniques or beam arrangement [7]. The main cause of this inconsistency would be that any VDth is a single point in a DVH and thus cannot reflect the entire dose distribution, unlike the MLD which uses all the dose bins for its calculation. A study by Seppenwoolde

et al. [20] also supports the claim that dose response of RP is best fit by the MLD-based linear model rather than the threshold model such as VDth. Nevertheless, the QUANTEC report proposed the recommendation that V20 and MLD are limited to 30-35% and 20-23 Gy respectively.

The dose-RILD relationships for SBRT are not as well established. Dose tolerance for SBRT is expected to be considerably different from CFRT due to different dose distribution and fractionation [21]. Previous RTOG trials for lung SBRT (0236, 0813, 0915) used lung dose constraints of V20 <10% and $D_{1000cc} < 2.7Gy$ per fraction. These limits were derived from toxicity results from preclinical and clinical data for CFRT using linear-quadratic model based conversion with an additional conservative margin [22]. The most predictive dose-volumetric metrics vary across studies. According to a literature survey by Yamashita et al. [23], MLD was most frequently reported as a significant factor for grade 2 or higher RP, while V5, V20, V25 were found significant by a single study. Different use of fractionation schedules convolutes dosimetric modeling from SBRT patients. Some studies use EQD2 to bring dose distributions with different fractionation into a common scale [20] (equation 1.2), but it is not always enforced.

2.3.2 Fibrosis models

It has been suggested that radiation fibrosis (RF) exhibits more serial response than RP: Moiseenko et al. [24] fitted the LKB model to thymoma patient data with an endpoint defined as absence/presence of radiographic changes. They found the best-fit volume parameter $n = 0.5$, which was lower than that ($\sim$1) for RP. This was supported by Vågane et al. [25] who demonstrated that $n = 0.5$ led to less fitting

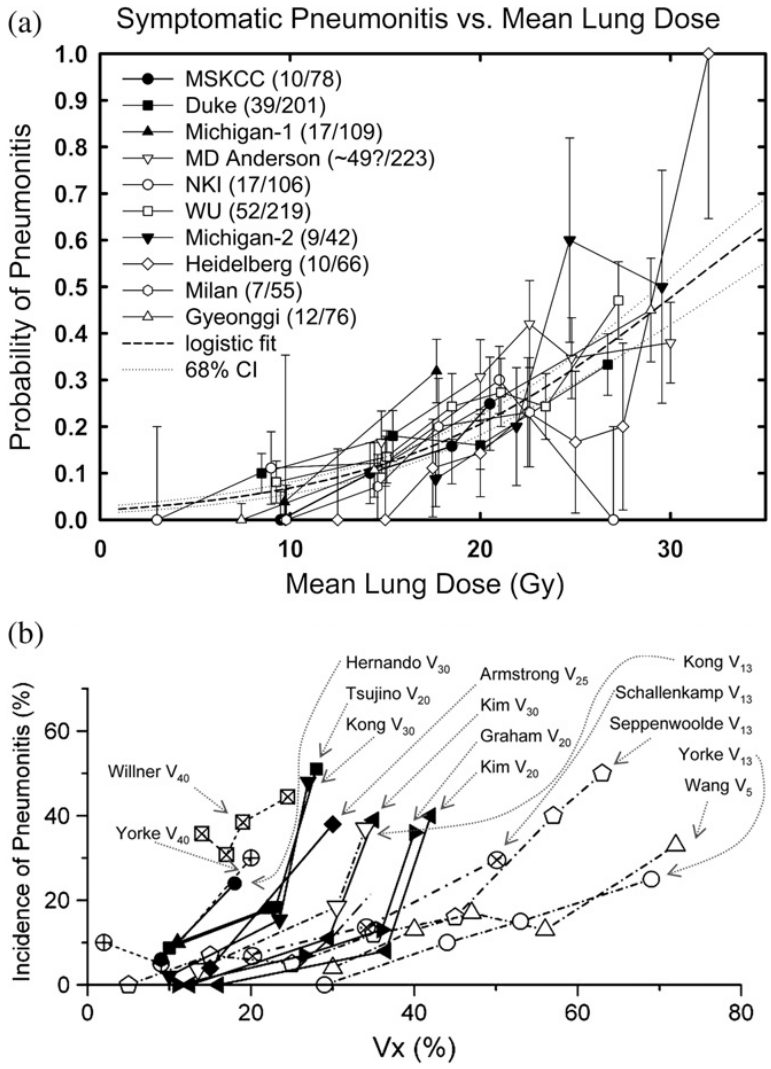

Figure 2.3: The QUANTEC compilation of mean lung dose (left) or VDth (rights) models for radiation pneumonitis. The dashed line on the top graph is logistic fit to the compiled data. Permission from Marks et al. [7]

error than Burman et al.'s value $n = 0.87$. Lower n implies that dose response depends more on maximum dose and a volume effect is less apparent when fibrosis is defined in a binary fashion.

Whereas RP is a systemic response, RF tends to display non-uniformity in irradiated regions. CT densitometry is a widely used non-invasive technique to investigate the extent of the disease. Collagen deposition in affected lung tissues increases physical tissue density, which is indirectly measured by CT imaging in terms of X-ray attenuating property. In general, there is linear relationship between the physical density of biological tissues and an attenuation coefficient, with the exception of bone which has higher atomic numbers and thus drastically changes attenuation properties. CT calibration is performed by fitting a linear function to CT attenuation measurement of materials with known physical density.

A Hounsfield unit (HU) is a unit for X-ray attenuation at every image voxel. It is scaled in a way that air (lowest density material) is assigned to -1000 and water is given 0. Any other tissues are assigned to a HU value depending on CT measurement of a linear attenuation coefficient (μ) which is energy loss of X-ray photons per length (equation 2.11):

$$HU = 1000\frac{\mu - \mu_{water}}{\mu_{water} - \mu_{air}} \tag{2.11}$$

Fibrotic response from radiotherapy was characterized using CT densitometry in different ways. Increase in average lung HU or density after RT was reported [26] [27] [28] albeit with low sensitivity [26]. An alternative method is to measure correlation between regional CT density and local dose deposition. Strong correlation

was demonstrated by studies on both CFRT [29], [30] and SBRT [31] [32] cohorts. These studies evaluated the correlation via a dose response curve (DRC) which is obtained by dividing contoured lung into several isodose subvolumes and evaluate average CT change within the volumes. Rosen et al. [30] measured HU changes from baseline in the regions that were delineated by radiologists. From the range of density changes for the fibrotic regions, they established threshold density values for each grade of fibrosis. Probability of fibrosis was defined as a fraction of lung voxels that were classified as fibrosis by applying these threshold values. This DRC approach was also adopted by Stroian et al. [33]. Ghobadi et al. [34] found that local spread of CT numbers, as well as an average, is correlated to other independent endpoints such as histology and breathing rate.

The DRC could be a useful dose sensitivity indicator since it decouples a volume from a dose effect. Furthermore, unlike DVH parameters, spatial information can be preserved via the subregion analysis.

2.4 Non-DVH RILD factors and models

2.4.1 Irradiation to lower lung

One major limitation in any DVH-derived NTCP models, including MLD, VDth, and the LKB model, is that spatial dose information (e.g. location of a high dose gradient region) is lost while constructing a DVH. Several studies report that lower lung is more sensitive to radiation and thus contributes more to RILD. One of the earliest findings was a series of animal experiments by Liao [35], Travis [36], and colleagues. They irradiated different subregions of mice lung and found that a breathing rate (a surrogate for RILD) and a mortality rate was higher when the subregions were

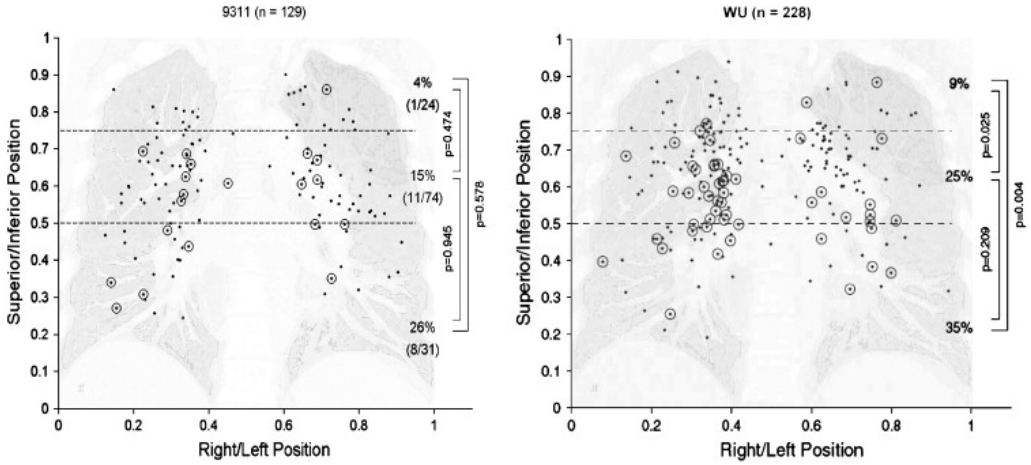

Figure 2.4: Distribution of tumour centroid locations (points) projected to a coronal plane from two datasets: RTOG 9311 [41] (left), Washington University [40] (WU, right). Circled points represent RP events. From the WU dataset, significant difference in RP frequency was found between lower 50% and upper 25% of lung. With permission from [40].

located in lower lung. Similar findings followed from clinical studies [37] [38] [39] [40]. Hope et al. [39] and Bradley et al. [40] incorporated this effect into their multivariate models by introducing a factor for a centre-of-mass of tumour location in superior-inferior direction (COMSI) which was associated with grade ≥ 2 RP (figure 2.4).

As a possible explanation of the location effect, Travis et al. [36] suggested that a lower lobe could have higher density of alveoli while an upper lung contains more dead space from conducting airways.

2.4.2 Irradiation of heart

Another related risk factor is irradiation of heart which often falls in treatment beams' paths when a tumour is located in a lower lobe. Dose to heart was found to be the most significant univariate predictor of RP in one clinical study [42]. The

heart irradiation effect was independently tested with a rat model by van Luijk et al [43] where an increase in breathing rate was observed when a field covered the heart. They fit to this data the modified critical volume model where a functional reserve is modified as a function of heart dose. Another animal study by Ghobadi et al. [44] suggested a synergistic effect of lung and heart irradiation on cardiopulmonary functions; In their experiment, diastolic dysfunction in the left ventricle of a heart was enhanced when both lung and heart were irradiated, contributing to dyspnea and RILD.

2.4.3 Smoking status

Interestingly, in contrast to its harmful carcinogenic role, smoking was found to decrease RILD risk by several clinical studies [45] [46] as well as a pooled analysis [47]. Tucker et al. [45] found in their cohort that a rate of RP ($\geq$ 3) was lower among the patients that smoked during radiotherapy. A meta-analysis by Vogelius and Bentzen [47] obtained an odds ratio (OR) of 0.6 (p=0.008) for ongoing smoking during radiotherapy and OR of 0.7 for history of smoking (p=0.06). Mechanisms of the protective property of smoking are not clearly known. One possible cause would be smoking-induced suppression of immune response, which was shown by reduced fibrogenic markers in bronchoalveolar lavage fluid from smokers [48]. Nevertheless, relevance of smoking on RILD is still being debated. Jin et al. [46] suggested ambiguity in outcome scoring as a potential reason that the correlation is sporadically reported; Smokers are more likely to have baseline respiratory symptoms such as cough or COPD and their post-RT syndromes may not be as noticeable as for the patients with healthy baseline lung functions.

2.4.4 Chemotherapy

Use of chemotherapy is another disputable risk factor for RILD. Schedules and agents used for chemotherapy vary across institutions, which convolutes a large-scale risk analysis. There are two types of chemotherapy scheduling:

- Induction chemotherapy: For bulky tumours, chemotherapy can be given before radiotherapy to help shrink the tumour to a manageable size.
- Concurrent chemotherapy: Chemotherapy is given at the same time as radiotherapy, which was shown to improve overall survival for stage III NSCLC patients [49].

Study results on the effect of induction chemotherapy on RILD are inconclusive [50]. However, concurrent chemotherapy is considered to be add extra toxicity to patients. According to a meta-analysis by Palma et al. [51], concurrent chemotherapy with carboplatin-paclitaxel was shown to increase RP risk by more than threefold (OR = 3.33, $p < .001$). The chemotherapy agent paclitaxel has a radio-sensitizing effect, which might be beneficial for tumour control but not normal tissue toxicity [52].

2.4.5 Other factors

Older patients were thought to be more susceptible to RILD [53], especially when paclitaxel was used for chemotherapy [52] [51]. Baseline pulmonary comorbidities such as interstitial pneumonitis [54] or chronic obstructive pulmonary disease (COPD) [55] were shown to increase RP risk for SBRT regimen. However, importance of baseline lung disease is debated in CFRT cases [56].

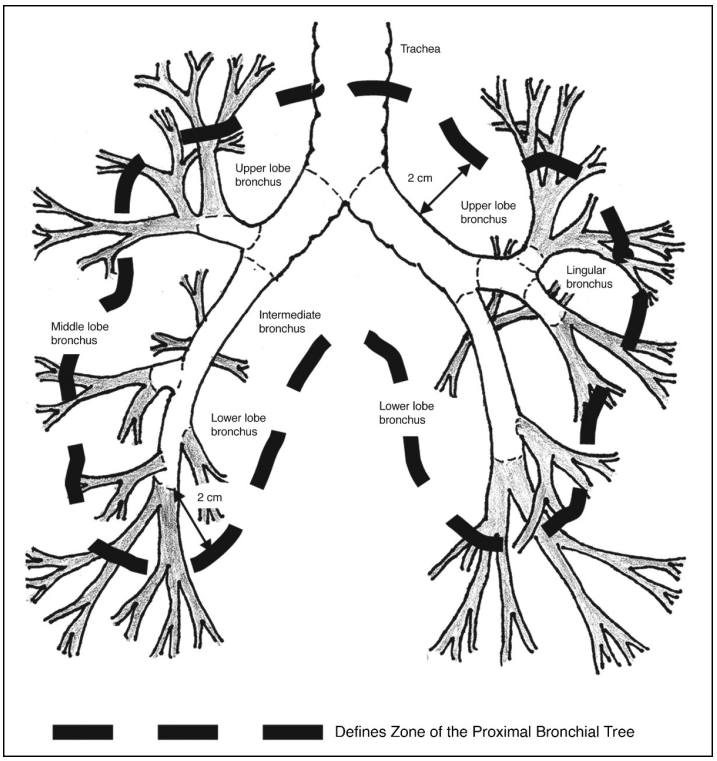

Figure 2.5: Illustration of the "no-fly zone": the anatomical region that is defined by Radiation Therapy Oncology Group protocols as a central tumour region (within the dashed area). With permission from Timmerman et al. [57]

For SBRT treatments, the patients with a tumour near central structures (e.g. hilum, central airways, primary bronchus, pericardium: see figure 2.5 for definition) are treated with special caution due to increased severe toxicity risk [57]. These tumours are treated with more fractions than peripheral tumours to reduce the BED to organs at risk. Due to this risk-adaptive dose prescription, it would be difficult to isolate from current clinical data the extra RILD risk due to a central tumour location.

There are numerous biological factors, proteins and genes, that have been proposed as a potential RILD risk modifier. Understanding the rationales behind these markers requires some background knowledge on RILD pathogenesis; A special attention will be given to the biomarkers in a chapter 3.

2.5 Lung NTCP models: sources of uncertainties

As the QUANTEC review revealed, results on dose-RILD correlation as well as non-dosimetric factors don't often agree between studies [7] [50]. There is also a large spread in the values of best-fit NTCP parameters in the literatures [58]. Besides the nature of heterogeneous response due to non-dosimetric factors, inconsistent study design and data acquisition contribute to the large uncertainty and need to be mentioned.

There is variation in the defined extent of lung, especially in which target volume (PTV, GTV, CTV) should be subtracted. Since PTV contains a margin that receives close to prescription dose, excluding PTV is expected to yield lower lung dose and thus might impact the dosimetric correlation of RILD.

Uncertainty in scoring and defining toxicity outcomes is another important factor. Building NTCP models involves conversion of toxicity grades into a binary toxicity event. However, using different toxicity grade cutoffs could lead to different dose-volume relationships [59] and significant predictors [47]. Inconsistency could also arise from choosing different toxicity grading schemes and the length of follow-up [60] [59]. Differential diagnosis of RP is not trivial because similar respiratory symptoms can be caused by other medical conditions such as viral infection, chronic obstructive pulmonary disease (COPD) and cardiac disease [61].

Dosimetric parameters are typically obtained from planning dose calculation which is performed on a free-breathing planning CT image. However, an actual dose distribution is subject to accuracy and reproducibility in patient positioning and plan delivery. This deviance from the planning dose is not always taken into account. Dose calculation algorithms could also affect the accuracy of RILD-dose relationship. The calculation algorithms have evolved over the history of computerized treatment planning from crude approximation to accurate simulation of particle transport using the Monte Carlo approach. Summary of modern dose calculation algorithms is provided in the appendix. Nielsen et al. [62] found that Burman et al.'s NTCP values could vary up to 0.23 when switching between 6 different dose algorithms from 4 commercial TPSs. Stroian et al. [33] suggested that the Monte-Carlo calculation yielded stronger correlation between local dose and probability of radiation fibrosis. Dose correction for tissue heterogeneity is an important dose calculation parameter; Lung is composed of and surround by various tissues with different density, and calculation results can change significantly if the tissue heterogeneity is not taken into account [63]. This might be a source of discrepancy between the studies before and after heterogeneity correction became commonplace in treatment planning systems.

In case of RF modelling, a post-treatment CT image have to be compared to a baseline image to assess density change. The matching between two images is made possible by a technique called the *image registration* which determines spatial transformation from one image to another. However, uncertainty in the image comparison arises from spatial inaccuracy of a registration process, which is more severe in the

presence of radiation-induced changes such as tumour regression or fibrotic changes [64].

2.6 Summary

Dose-volumetric relationships for RILD, both radiation pneumonitis and fibrosis, have been extensively studied by many researchers. The critical volume model, a semi-mechanistic model with the assumption that lung is organized with parallel architecture, is widely accepted in a clinical society; This model predicts that RILD risk is sensitive to the change in mean lung dose and irradiated volume. However, compilation of past published results shows a gradual increase in RILD risk as a function of dose or volume, without a clear threshold value below which the risk is kept minimal. There are however factors other than lung dose that were shown to modify RILD risks. Upcoming challenges are in: 1) validating these non-dosimetric factors in independent datasets, and 2) incorporating all the relevant factors into a unified NTCP model that is expected to sharpen risk stratification by the dosimetric risk models.

REFERENCES

[1] X. Allen Li, M. Alber, J. O. Deasy, et al., The use and QA of biologically related models for treatment planning: Short report of the TG-166 of the therapy physics committee of the AAPM, Medical Physics 39 (3) (2012) 1386–1409.

[2] J. T. Lyman, Complication probability as assessed from dose-volume histograms, Radiation Research Supplement 8 (1985) pp. S13–S19.

[3] G. J. Kutcher, C. Burman, Calculation of complication probability factors for non-uniform normal tissue irradiation: the effective volume method, International Journal of Radiation Oncology*Biology*Physics 16 (6) (1989) 1623–1630.

[4] C. Burman, G. Kutcher, B. Emami, M. Goitein, Fitting of normal tissue tolerance data to an analytic function, International Journal of Radiation Oncology*Biology*Physics 21 (1) (1991) 123 – 135, three-Dimensional Photon Treatment Planning Report of the Collaborative Working Group on the Evaluation of Treatment Planning for External Photon Beam Radiotherapy.

[5] B. Emami, J. Lyman, A. Brown, et al., Tolerance of normal tissue to therapeutic irradiation, International Journal of Radiation Oncology*Biology*Physics 21 (1) (1991) 109 – 122, three-Dimensional Photon Treatment Planning Report of the Collaborative Working Group on the Evaluation of Treatment Planning for External Photon Beam Radiotherapy.

[6] A. Niemierko, A generalized concept of equivalent uniform dose (EUD), Medical Physics 26 (1999) 1100.

[7] L. B. Marks, S. M. Bentzen, J. O. Deasy, et al., Radiation dose-volume effects in the lung, International Journal of Radiation Oncology*Biology*Physics 76 (3, Supplement 1) (2010) S70 – S76.

[8] V. Moiseenko, D. J., J. Van Dyk, Radiobiological modeling for treatment planning the modern technology of radiation oncology, in: J. Van Dyk (Ed.), A Compendium for Medical Physicists and Radiation Oncologists, Medical Physics Publishing, Madison, WI, 2005, pp. 185–220.

[9] Q. Wu, R. Mohan, A. Niemierko, R. Schmidt-Ullrich, Optimization of intensity-modulated radiotherapy plans based on the equivalent uniform dose, Int J Radiat Oncol Biol Phys 52 (1) (2002) 224–35.

[10] ICRU report 83, Special considerations regarding absorbed-dose and dose-volume prescribing and reporting in IMRT, Journal of the ICRU 10 (1) (2010) 27–40.

[11] H. R. Withers, J. M. Taylor, B. Maciejewski, Treatment volume and tissue tolerance, International Journal of Radiation Oncology*Biology*Physics 14 (4) (1988) 751 – 759.

[12] A. Niemierko, M. Goitein, Modeling of normal tissue response to radiation: The critical volume model, International Journal of Radiation Oncology*Biology*Physics 25 (1) (1993) 135 – 145.

[13] A. Jackson, G. J. Kutcher, E. D. Yorke, Probability of radiation-induced complications for normal tissues with parallel architecture subject to non-uniform

irradiation, Medical Physics 20 (3) (1993) 613–625.

[14] A. Jackson, R. Ten Haken, J. Robertson, M. Kessler, G. Kutcher, T. Lawrence, Analysis of clinical complication data for radiation hepatitis using a parallel architecture model, International Journal of Radiation Oncology*Biology*Physics 31 (4) (1995) 883 – 891.

[15] T. Bortfeld, R. Schmidt-Ullrichm, W. De Neve, D. Wazer, Image-guided IMRT, Springer Verlag, Heidelberg, 2005.

[16] P. Kallman, A. Agren, A. Brahme, Tumour and normal tissue responses to fractionated non-uniform dose delivery, International Journal of Radiation Biology 62 (2) (1992) 249–262.

[17] L. B. Marks, E. D. Yorke, A. Jackson, et al., Use of normal tissue complication probability models in the clinic, International Journal of Radiation Oncology*Biology*Physics 76 (3, Supplement) (2010) S10 – S19.

[18] S. L. Kwa, J. V. Lebesque, J. C. Theuws, et al., Radiation pneumonitis as a function of mean lung dose: an analysis of pooled data of 540 patients, International Journal of Radiation Oncology*Biology*Physics 42 (1) (1998) 1 – 9.

[19] M. V. Graham, J. A. Purdy, B. Emami, et al., Clinical dose volume histogram analysis for pneumonitis after 3d treatment for non-small cell lung cancer (NSCLC), International Journal of Radiation Oncology*Biology*Physics 45 (2) (1999) 323 – 329.

[20] Y. Seppenwoolde, J. V. Lebesque, K. de Jaeger, et al., Comparing different NTCP models that predict the incidence of radiation pneumonitis, International

Journal of Radiation Oncology*Biology*Physics 55 (3) (2003) 724 – 735.

[21] S. H. Benedict, K. M. Yenice, D. Followill, et al., Stereotactic body radiation therapy: the report of aapm task group 101, Med Phys 37 (8) (2010) 4078–101.

[22] R. Timmerman, J. Galvin, J. Michalski, et al., Accreditation and quality assurance for radiation therapy oncology group: Multicenter clinical trials using stereotactic body radiation therapy in lung cancer, Acta Oncologica 45 (7) (2006) 779–786.

[23] H. Yamashita, W. Takahashi, A. Haga, K. Nakagawa, Radiation pneumonitis after stereotactic radiation therapy for lung cancer, World J Radiol 6 (9) (2014) 708–15.

[24] V. Moiseenko, T. Craig, A. Bezjak, J. V. Dyk, Dose-volume analysis of lung complications in the radiation treatment of malignant thymoma: a retrospective review, Radiotherapy and Oncology 67 (3) (2003) 265 – 274.

[25] R. Vågane, Ø. S. Bruland, S. D. Fosså, D. R. Olsen, Radiological and functional assessment of radiation-induced pulmonary damage following breast irradiation, Acta Oncologica 47 (2) (2008) 248–254.

[26] E. el Khatib, S. Lehnert, Lung density change observed in vivo in rat lungs after irradiation: variations among and within individual lungs, International Journal of Radiation Oncology*Biology*Physics 16 (1989) 745–754.

[27] S. Lehnert, E. El-Khatib, The use of ct densitometry in the assessment of radiation-induced damage to the rat lung: A comparison with other endpoints, International Journal of Radiation Oncology*Biology*Physics 16 (1) (1989) 117 – 124.

[28] M. I. Koukourakis, P. G. Tsoutsou, I. Abatzoglou, Computed tomography assessment of lung density in patients with lung cancer treated with accelerated hypofractionated radio-chemotherapy supported with amifostine, Americal Journal of Clinical Oncology 32 (2009) 258–261.

[29] J. Ma, J. Zhang, S. Zhou, et al., Regional lung density changes after radiation therapy for tumors in and around thorax, International Journal of Radiation Oncology*Biology*Physics 76 (1) (2010) 116 – 122.

[30] I. I. Rosen, T. A. Fischer, J. A. Antolak, et al., Correlation between Lung Fibrosis and Radiation Therapy Dose after Concurrent Radiation Therapy and Chemotherapy for Limited Small Cell Lung Cancer, Radiology 221 (3) (2001) 614–622.

[31] D. A. Palma, J. van Srnsen de Koste, W. F. Verbakel, A. Vincent, S. Senan, Lung density changes after stereotactic radiotherapy: A quantitative analysis in 50 patients, International Journal of Radiation Oncology*Biology*Physics 81 (4) (2011) 974 – 978.

[32] Q. Diot, B. Kavanagh, T. Schefter, L. Gaspar, K. Stuhr, M. Miften, Regional normal lung tissue density changes in patients treated with stereotactic body radiation therapy for lung tumors, International Journal of Radiation Oncology*Biology*Physics 84 (4) (2012) 1024 – 1030.

[33] G. Stroian, C. Martens, L. Souhami, D. L. Collins, J. Seuntjens, Local correlation between monte-carlo dose and radiation-induced fibrosis in lung cancer patients, International Journal of Radiation Oncology*Biology*Physics 70 (3) (2008) 921 – 930.

[34] G. Ghobadi, L. E. Hogeweg, H. Faber, et al., Quantifying local radiation-induced lung damage from computed tomography, International Journal of Radiation Oncology*Biology*Physics 76 (2) (2010) 548 – 556.

[35] Z.-X. Liao, E. L. Travis, S. L. Tucker, Damage and morbidity from pneumonitis after irradiation of partial volumes of mouse lung, International Journal of Radiation Oncology*Biology*Physics 32 (5) (1995) 1359 – 1370.

[36] E. L. Travis, Z.-X. Liao, S. L. Tucker, Spatial heterogeneity of the volume effect for radiation pneumonitis in mouse lung, International Journal of Radiation Oncology*Biology*Physics 38 (5) (1997) 1045 – 1054.

[37] Y. Seppenwoolde, K. D. Jaeger, L. J. Boersma, J. S. Belderbos, J. V. Lebesque, Regional differences in lung radiosensitivity after radiotherapy for non small-cell lung cancer, International Journal of Radiation Oncology*Biology*Physics 60 (3) (2004) 748 – 758.

[38] E. D. Yorke, A. Jackson, K. E. Rosenzweig, L. Braban, S. A. Leibel, C. C. Ling, Correlation of dosimetric factors and radiation pneumonitis for non-small-cell lung cancer patients in a recently completed dose escalation study, International journal of radiation oncology, biology, physics 63 (3) (2005) 672682.

[39] A. J. Hope, P. E. Lindsay, I. E. Naqa, et al., Modeling radiation pneumonitis risk with clinical, dosimetric, and spatial parameters, International Journal of Radiation Oncology*Biology*Physics 65 (1) (2006) 112 – 124.

[40] J. D. Bradley, A. Hope, I. E. Naqa, et al., A nomogram to predict radiation pneumonitis, derived from a combined analysis of RTOG 9311 and institutional data, International Journal of Radiation Oncology*Biology*Physics 69 (4) (2007) 985

– 992.

[41] J. Bradley, M. V. Graham, K. Winter, et al., Toxicity and outcome results of RTOG 9311: A phase I-II dose-escalation study using three-dimensional conformal radiotherapy in patients with inoperable non-small-cell lung carcinoma, International Journal of Radiation Oncology*Biology*Physics 61 (2) (2005) 318 – 328.

[42] E. X. Huang, A. J. Hope, P. E. Lindsay, et al., Heart irradiation as a risk factor for radiation pneumonitis, Acta Oncologica 50 (1) (2011) 51–60.

[43] P. van Luijk, H. Faber, H. Meertens, et al., The impact of heart irradiation on dose-volume effects in the rat lung, International Journal of Radiation Oncology*Biology*Physics 69 (2) (2007) 552 – 559.

[44] G. Ghobadi, S. van der Veen, B. Bartelds, et al., Physiological interaction of heart and lung in thoracic irradiation, International Journal of Radiation Oncology*Biology*Physics 84 (5) (2012) e639 – e646.

[45] S. L. Tucker, H. H. Liu, Z. Liao, et al., Analysis of radiation pneumonitis risk using a generalized lyman model, International Journal of Radiation Oncology*Biology*Physics 72 (2) (2008) 568 – 574.

[46] H. Jin, S. L. Tucker, H. H. Liu, et al., Dose-volume thresholds and smoking status for the risk of treatment-related pneumonitis in inoperable non-small cell lung cancer treated with definitive radiotherapy, Radiotherapy and Oncology 91 (3) (2009) 427 – 432.

[47] I. R. Vogelius, S. M. Bentzen, A literature-based meta-analysis of clinical risk factors for development of radiation induced pneumonitis, Acta Oncologica

51 (8) (2012) 975–983.

[48] L. Bjermer, R. Hallgren, K. Nilsson, et al., Radiation-induced increase in hyaluronan and fibronectin in bronchoalveolar lavage fluid from breast cancer patients is suppressed by smoking, European Respiratory Journal 5 (7) (1992) 785–790.

[49] W. J. Curran, R. Paulus, C. J. Langer, et al., Sequential vs concurrent chemoradiation for stage III non-small cell lung cancer: Randomized phase iii trial rtog 9410, Journal of the National Cancer Institute 103 (19) (2011) 1452–1460.

[50] F.-M. Kong, S. Wang, Nondosimetric risk factors for radiation-induced lung toxicity, Seminars in Radiation Oncology 25 (2) (2015) 100 – 109.

[51] D. A. Palma, S. Senan, K. Tsujino, et al., Predicting radiation pneumonitis after chemoradiation therapy for lung cancer: An international individual patient data meta-analysis, International Journal of Radiation Oncology*Biology*Physics 85 (2) (2013) 444 – 450.

[52] B. Parashar, A. Edwards, R. Mehta, et al., Chemotherapy significantly increases the risk of radiation pneumonitis in radiation therapy of advanced lung cancer, Am J Clin Oncol 34 (2) (2011) 160–4.

[53] J. Dang, G. Li, S. Zang, S. Zhang, L. Yao, Risk and predictors for early radiation pneumonitis in patients with stage iii non-small cell lung cancer treated with concurrent or sequential chemoradiotherapy, Radiation Oncology 9 (1).

[54] A. Takeda, E. Kunieda, T. Ohashi, et al., Severe copd is correlated with mild radiation pneumonitis following stereotactic body radiotherapy, Chest 141 (4) (2012) 858–66.

[55] H. Yamashita, S. Kobayashi-Shibata, A. Terahara, et al., Prescreening based on the presence of CT-scan abnormalities and biomarkers (KL-6 and SP-D) may reduce severe radiation pneumonitis after stereotactic radiotherapy, Radiat Oncol 5 (2010) 32.

[56] J. Wang, J. Cao, S. Yuan, et al., Poor baseline pulmonary function may not increase the risk of radiation-induced lung toxicity, International Journal of Radiation Oncology*Biology*Physics 85 (3) (2013) 798 – 804.

[57] R. Timmerman, R. McGarry, C. Yiannoutsos, et al., Excessive toxicity when treating central tumors in a phase ii study of stereotactic body radiation therapy for medically inoperable early-stage lung cancer, J Clin Oncol 24 (30) (2006) 4833–9.

[58] J. B. S. M. Bentzen, J. Z. Skoczylas, Quantitative clinical radiobiology of early and late lung reactions, International Journal of Radiation Biology 76 (4) (2000) 453–462.

[59] S. L. Tucker, H. Jin, X. Wei, et al., Impact of toxicity grade and scoring system on the relationship between mean lung dose and risk of radiation pneumonitis in a large cohort of patients with non-small cell lung cancer, International Journal of Radiation Oncology*Biology*Physics 77 (3) (2010) 691 – 698.

[60] S. Faria, M. Aslani, F. Tafazoli, L. Souhami, C. Freeman, The challenge of scoring radiation-induced lung toxicity, Clinical Oncology 21 (5) (2009) 371 – 375.

[61] Z. Kocak, E. S. Evans, S.-M. Zhou, et al., Challenges in defining radiation pneumonitis in patients with lung cancer, International Journal of Radiation

Oncology*Biology*Physics 62 (3) (2005) 635 – 638.

[62] T. B. Nielsen, E. Wieslander, A. Fogliata, M. Nielsen, O. Hansen, C. Brink, Influence of dose calculation algorithms on the predicted dose distributions and NTCP values for NSCLC patients, Medical Physics 38 (5) (2011) 2412–2418.

[63] N. Papanikolaou, J. Battista, A. Boyer, et al., Tissue inhomogeneity corrections for megavoltage photon beams., in: AAPM Report No. 85., Medical Physics Publishing, 2004.

[64] M. Guckenberger, K. Baier, A. Richter, J. Wilbert, M. Flentje, Evalution of surface-based deformable image registration for adaptive radiotherapy of non-small cell lung cancer (NSCLC), Radiation Oncology 4 (1) (2009) 1–3.

CHAPTER 3
Biomarkers for RILD

Contents

3.1 Introduction

As seen in chapter 2, dosimetric NTCP models - especially for radiation pneumonitis - lack accuracy in an individual level due to heterogeneity in dose response. Chapter 4 demonstrated the dynamic nature of dose response for fibrosis, which does not fit the current NTCP framework with fixed parameter values for one endpoint. The next step is to move beyond the static assumption of dose response and investigate the factors behind the dynamic response, which eventually allows for more patient-specific modelling. Chapter 2 introduced some of the patient-specific RILD factors that are unrelated to lung DVH. However, biological mechanisms that link these factors to the disease are not always clear while some being debated. Meanwhile, as biological assays and functional imaging are becoming more accessible, there is a growing interest in direct measurement of patients' biological status for predicting treatment response [1].

This chapter will move on to biological factors or *biomarkers* that might help better identify the patient subgroup with higher RILD risk. Before all, it will introduce theories about RILD pathology that provide a theoretical basis for the putative biomarkers. Then, biomarkers that are responsible for different biological processes after radiation will be reviewed for their potential role as a RILD biomarker.

3.2 Pathophysiology of RILD

Detailed cellular and molecular mechanisms of RILD are complex and have not been fully established. Its complexity is beyond the *target cell hypothesis* which relates the intensity of the complication directly to the proportion of cells damaged by radiation. Several animal experiments have shown that injury is mediated by

a network of communications between inflammatory, endothelial, and parenchymal cells [2]. It is general consensus that there are two distinct phases in the progression of the injury: an early inflammatory phase leading to RP and a late fibrotic phase responsible for RF.

3.2.1 Early phase

Ionizing radiation interacts with water, creating reactive oxygen species (ROS) which damages different constituents of lung parenchyma and stroma with different consequences. One of the earliest changes is release of surfactant by type II pneumocytes into alveolar lumen, which occurs from 1 hour to 7 days after irradiation [3]. On the other hand, loss of type I pneumocytes induces proliferation and differentiation of type II pneumocytes in order to repair structural damage to alveoli. Damage to endothelium of alveolar capillaries increases the permeability of the vessels, causing accumulation of body fluids (edema) in alveolar walls and subsequent decrease in perfusion.

Release of cytokines by the irradiated cells, named as the "perpetual cytokine cascade" [4], initiates inflammatory response and persists throughout the course of RILD. *Cytokines* refer to small proteins that transmit signals to other cells through membrane receptors. The cytokine release is a defence mechanism that allows the affected cells to cooperatively resolve the injury as well as spreading "danger signals" to neighbouring cells [5]. A type of the released cytokines depends on producing cells and time after irradiation. Macrophages are known as a main production cite of the cytokines [6].

Also, reactive oxygen species (ROS), which is a direct product of interaction between ionizing radiation and water, takes part in radiation-induced signalling pathways. ROS inhibits protein tyrosine phosphatase (PTPase) activities, which activates many downstream signalling pathways through the actions of tyrosine kinases [7]. ROS are also involved in regulating expression of several cytokines and growth factors [4].

3.2.2 Late phase

The most prominent feature of late phase is fibrogenesis, proliferation of extracellular matrix and collagen. Radiation-induced fibrogenesis is thought as deregulated tissue remodelling that follows inflammation and new tissue formation as a normal wound healing process [8]. A key molecule in this process is transforming growth factor $\beta 1$ (TGFβ). This molecule is activated from its latent form by direct action of ionizing radiation or inflammatory cytokines that are released by damaged alveolar endothelial cells [2]. The active TGFβ turns on the TGFβ signalling pathway, promoting proliferation of fibroblasts and transformation of fibroblasts to myofibroblasts which is a main ingredient of fibrotic tissues. Fibrogenesis is further reinforced by epithelial-mesenchymal transition (EMT) which converts alveolar epithelium to myofibroblasts [9].

Prolonged hypoxia and oxidative stress in lung tissues are thought to be an important factor that perpetuates the fibrotic reaction [4]. A number of mechanisms induce and sustain hypoxia after irradiation: Lack of perfusion due to damaged vasculature and increased alveolar wall thickness, and increased oxygen consumption by active inflammatory cells [4]. Hypoxia stimulates macrophages to increase

the production of TGFβ [10] thereby reinforcing fibrogenic pathways. Hypoxia also promotes development of new blood vessels (angiogenesis), which attempts to compensate for reduced oxygen supply [11]. This is realized by increased production of vascular endothelial growth factor (VEGF) [10] and angiotensin II [12].

Figure 3.1 summarizes the proposed theories of interactions between biological processes and signalling molecules in the RILD pathogenesis.

3.3 What is a biomarker?

A biomarker is formally defined as "a characteristic that is objectively measured and evaluated as an indicator on normal biological processes, pathologic processes, or biological responses to a therapeutic intervention" [13]. In the context of radiotherapy, Okunieff et al. made four categories of biomarkers depending on when the signal is relevant to a phenomenon of interest [14]:

- Predictive biomarkers: They are measurable before irradiation and are associated with an event after irradiation. Depending on test results on those markers, patients could be started with different treatment arms or clinical trials.
- Prognostic biomarkers: Likewise, these markers can predict a toxicity event before it occurs, but they are available at any time after the exposure. They may not be observable at baseline and thus not useful at the treatment planning stage. However, in case of prolonged treatments such as 30-fraction radiotherapy, the markers could be measured at mid-treatment and can still open a window for plan adaptation.

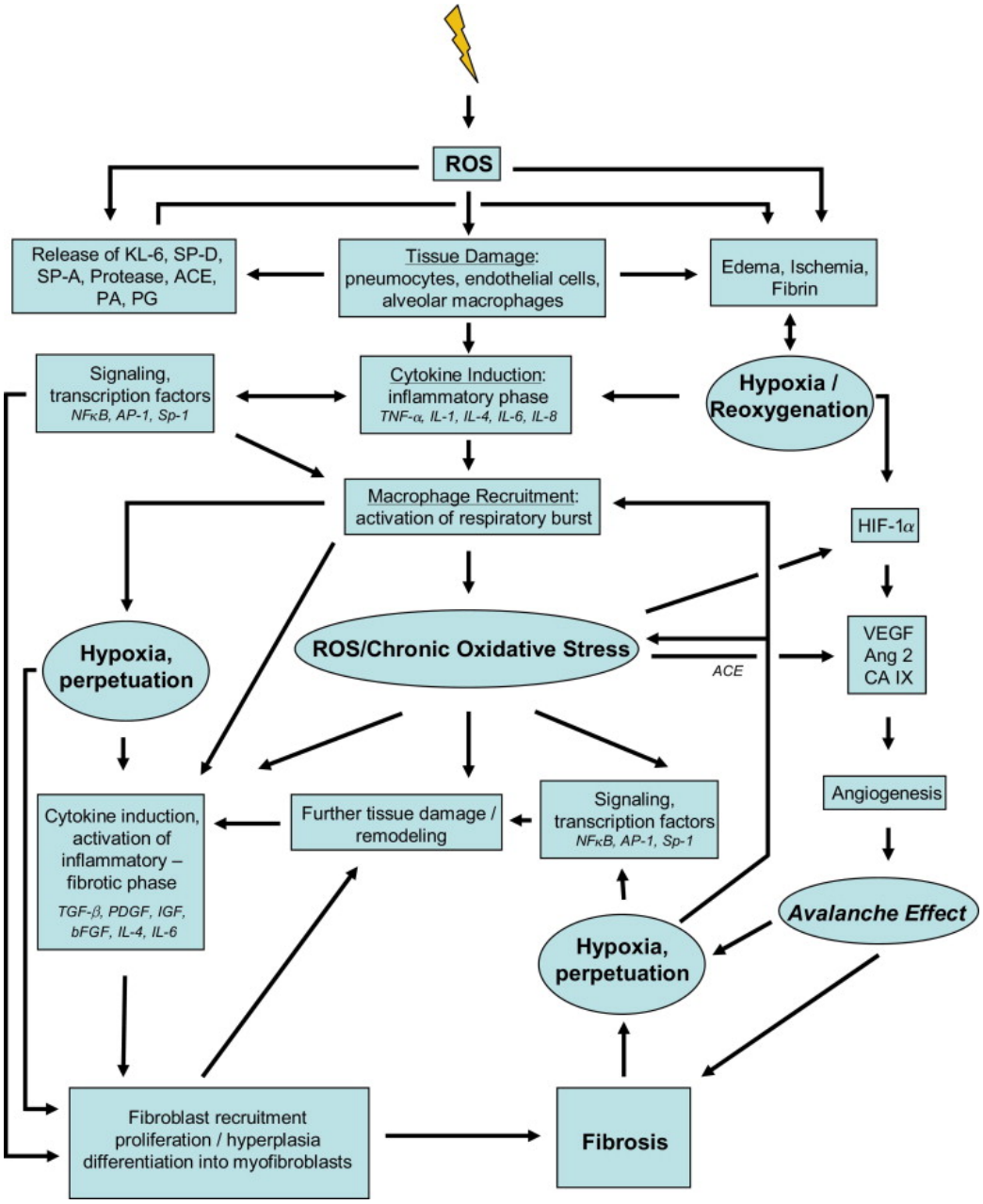

Figure 3.1: A simplified model by Fleckenstein [4] summarizing interactions between biological processes and signalling molecules following irradiation to normal lung tissues, beginning at an initial creation of reactive oxygen species (ROS) and ending at late fibrosis development. PA: plasminogen activator; PG: prostaglandin, Ang2: angiotensin II; CA IX: carboanhydrase IX.

- Diagnostic biomarkers: They are measured at the time of diagnosis of the event of interest and used to establish causality of the symptoms to irradiation.
- Dosimetric biomarkers: They are measured after irradiation and serve as a biological surrogate endpoint for dose deposition.

Among those, only predictive and prognostic biomarkers will be discussed here because they have a predictive value and thus are relevant to the context of this thesis.

3.4 RILD biomarkers

The remainder of the section will overview currently known predictive and prognostic RILD biomarkers. A list of biomarkers will consist of predominantly protein biomarkers that can be measured from patient blood (plasma or serum), which is the most relevant to this study.

3.4.1 Interleukins

Interleukin (IL) is a group of cytokines that are mainly produced by leukocytes. It plays an important role in modulating immune reactions, including immune cell migration, proliferation, and differentiation. There are several IL subtypes with different functions. IL-1α and IL-6 are known to promote inflammatory reactions, while other types such as IL-10 and IL-13 are anti-inflammatory [15]. IL-8 functions as a chemokine that attracts neutrophils to an inflammatory cite [16]. A mice study by Rubin et al. [17] suggests that an early surge in IL-1α and IL-1β right after RT might be responsible for initiation of an acute phase of RILD. In clinical investigations, IL-6 has appeared as a potential RP risk factor. Chen et al. [18] found that

plasma IL-6 level was higher for RP patients at baseline and remained high throughout the course of treatment. According to Arpin et al. [19], a decrease in serum IL-10 level combined with increasing IL-6 was the best RP predictor. However, this was not reproduced in a multi-cytokine study by Hart et al. [20], where IL-8 was the only significant factor for RP. In this study, controversially, low pre-treatment IL-8 increased RP risk, which was in contrary to the expectation that IL-8 would be pro-inflammatory.

3.4.2 Transforming growth factor β1

Transforming growth factor β1, (denoted here as TGFβ), belongs to a family of polypeptide growth factors that operates as a signalling molecule. The TGFβ signalling pathway has many downstream effects on cell growth, apoptosis, and differentiation [21]. In the context of RILD, TGFβ acts as a pro-inflammatory cytokine that promotes growth of fibroblasts and extracellular matrix as well as modulating immune reactions [4]. TGFβ is one of the most frequently studied RILD biomarkers; Its potential role as a RP risk factor was first discovered by Anscher et al. [22] where patients with RP showed an increase in plasma TGFβ from baseline to end-treatment. This result was repeatedly reproduced by the same group (Anscher and colleagues), which are summarized in table 1 of Kong et al. [23]. However, studies from other groups [24] [25] did not confirm the predictive significance of end-treatment TGFβ. Factors that might be responsible for the conflicting results are sample preparation methods [23], such as removal of platelets, and TGFβ production by tumour [21].

3.4.3 Angiotensin converting enzyme

Angiotensin converting enzyme (ACE) plays an important role in the renin-angiotensin system, a hormonal blood-pressure regulating system. It catalyzes conversion of angiotensin I to angiotensin II which acts as a vasoconstrictor, increasing blood pressure in effect. Its main production sites are vascular endothelial cells in lung and kidney epithelium. Thus, serum concentration of ACE could decrease when lung epithelium is damaged by chemo- or radiotherapy [26]. ACE could indirectly promote RILD by increasing the level of angiotensin II which enhances inflammatory and fibrotic actions via the angiotensin type 1 (AT1) receptor. This theory is supported by several studies reporting ACE-related biomarkers as a predictive factor for RILD. Lower plasma ACE level, either pre- or post- radiotherapy, was reported for patients with RP [27]. Also, the patients who were taking ACE inhibitors before RT for treating heart diseases or hypertension reportedly suffer less from RILD [28] [29]. This is consistent with findings from animal experiments [12] [30].

3.4.4 Alpha-2-macroglobulin

Alpha-2-macroglobulin (α2M) is a plasma protein that inhibits the action of proteases (enzymes that catabolize proteins). It is hypothesized to be a regulator of inflammatory activities due to its ability to bind non-specifically to many cytokines such as TGFβ1 and interleukins 1, 4, and 6 and reduce their activities [31]. Its potential as a radioprotector was shown in rats by Mihailovic et al. [32] where DNA damage was reduced by administration of α2M before radiation. Later, α2M was re-discovered as a potential biomarker for RP by a bioinformatics-based research by Oh et al. [33]. Using the mass-spectroscopy technique to detect peptides in patient

blood samples, they identified α2M as a protein that was most associated *a-priori* with previously known RP biomarkers amongst the proteins that were consistently up-regulated in RP patients. In the same study, mid- to pre-RT ratio of α2M was validated as a RP predictor in an independent cohort.

3.4.5 Osteopontin

Osteopontin (OPN) is commonly known as a linkage glycoprotein that provides mechanical support to extracellular matrix. As its prefix (osteo-) suggests, it is mainly produced in bone. Outside fibroblasts, it acts as a cytokine that influences tumour progression and immune reactions [34]. There are several evidences that an elevated OPN level in plasma indicates negative prognosis for various types of cancer albeit with a poorly understood mechanism [35]. In lung tissues, OPN is produced in epithelium or macrophages in response to stress and plays a role as a pro-inflammatory and pro-fibrotic cytokine [34]: its expression was shown to increase during bleomycin-induced fibrosis in rats [36]. OPN is also a potential endogeneous marker for hypoxia: hypoxic condition in head and neck cancer was shown in-vivo to correlate with increase OPN expression in the tumour [37]. A link between ionizing radiation and OPN has not been firmly established: a recent study by Wohlleben did not find a notable radiation effect on OPN expression in glioblastoma cell lines [38]. Unfortunately, no in-vivo study exists to date on correlating OPN plasma level to radiotherapy outcomes.

3.4.6 Other protein biomarkers

Krebs von den Lungen-6 (KL-6) is a mucin-like protein that is mainly secreted by type II pneumocytes and bronchiolar epithelium, and it has been suggested that

its serum level after radiotherapy might indicate an inflammatory state [39]. Hara et al. showed that increase in KL-6 at 2 months after single-fraction SBRT predicted grade 3 or higher RP [40]. Surfactant proteins (SP) type A and D, are another potential RILD biomarkers; These are two types of surfactant proteins that are both involved in a immune reaction by binding to pathogens and mobilizing phagocyte uptake [41]. Takahashi et al. reported high specificity of serum SP-A (83%) and SP-D (85%) increase in early diagnosis of radiation-induced fibrosis [42]. These findings inspired pre-screening of the patients with high serum level of SP-A and SP-D in the university of Tokyo hospital, which managed to decrease incidence of severe RP [43].

3.4.7 Genetic markers

Some of the variations in genotypes, especially the genes that are responsible for DNA repair, inflammation, and oxidative stress pathways, might indicate inherent radiosensitivity [23]. *Single-nucleotide polymorphism (SNP)*, alteration of one nucleotide in a minor group of population, is a frequently studied genetic pattern. For example, SNPs in 2 base excision repair genes XRCC1 and APEX1 [44] and a TGFβ1 gene [45] were found to be associated with radiation pneumonitis. According to a study by Pu et al. [46] which involved external validation, 45 SNPs on three inflammation-related genes (PRKCE, DDX58, and TNFSF7) were predictive of RP.

Although there is a growing interest in using genetic profiles for personalizing radiotherapy, the reported predictive values of SNPs have to be taken with caution. A large scale (1613 patients) radiogenomic study called RAPPER (Radiogenomics: Assessment of Polymorphisms for Predicting the Effects of Radiotherapy) failed to validate any of the previously reported associations between SNPs and late breast

toxicity [47]. This study attributed this negative result to a trade-off between allele frequency and an effect size for SNPs: low frequency SNPs are more likely to be a false positive in a small sample size, whereas high frequency SNPs tend to have low odds ratio and thus hard to detect.

3.5 RILD biomarkers: possible pitfalls

Use of biological markers to identify the risky subgroup is potentially a powerful way to improve the accuracy of population-based dose model. However, similarly to dosimetric and clinical factors, reproducibility of the RILD biomarkers are often under question. As mentioned in section 2.5, inconsistent outcome reporting could be one potential cause of the discrepancies. Another obstacle in comparing biomarker studies is difference in biological assay methodology. Communication between two research groups over significance of TGFβ serves as a good example [48] [49]. The dispute was on the difference in sample processing methods to minimize platelet degranulation affecting the TGFβ plasma level [50]. The communication also touched on the difficulty in setting a universal cutoff for a normal biomarker level; A range of TGFβ concentration from healthy patients vary widely [51], and even that is not necessarily applicable to patients with lung cancer which is known to produce TGFβ [21]. Lastly, false discovery resulting from weak statistical power is a pervasive issue in the biomarker studies with a small sample size.

A number of measures need to be taken before translating the biomarker discoveries into the clinic. Outcomes need to be scored consistently across studies while minimizing ambiguity in interpretation of symptoms. Protocols need to be established and adhered with regards to biological sample handling and assay techniques.

Finally, a novel discovery of biomarkers need to be followed by other independent investigations to replicate the finding.

3.6 Systems radiobiology approach

In biology, there is a tradition of reductionism: reducing a system into its constituents and studying the effect of individual parts on the system [52]. Under reductionism, a single molecule, gene, or a pathway is tested for its association with a phenomenon in isolation from other factors. However, this perspective is reaching its limit in explaining complex phenomenon such as radiation biology [53]. Systems biology is a paradigm shift from studying one factor individually to multiple components simultaneously and how these components interact [52]. There are two essential components to study systems biology: 1) high dimensional biological data obtained by high throughput assays, and 2) computational modelling that detects reliable patterns from the biological data.

Complexity of RILD pathogenesis suggests that it is unlikely to find a single biological factor that predicts all the disease cases, which also explains conflicting reports for many biomarkers. What adds to this complexity is strong presence of a external physical factor - radiation - which is given with varying strength and quality. Thus, unlike the previous biomarker studies, this study will consider a system that consists of physical, biological and clinical factors. Then, we model RILD as a result of interactions between such factors. Chapter 6 will discuss the computational methods that enables such modelling from high dimensional data.

3.7 Summary

Pathogenesis of radiation-induced lung disease involves many interactions and crosstalks between different signalling molecules and biological processes. In short, irradiation first triggers inflammatory and immune reactions, which evolves into a late fibrotic phase by the actions of TGFβ and prolonged hypoxia. Several proteins were proposed as predictive and prognostic biomarkers for RILD. Most of them are produced in lung tissues or macrophages and act as a cytokine during the inflammatory phase. However, discrepancies between studies exist on the predictive power of the biomarkers. The limitation of a single biomarker study leads to an alternative using systems biology where the actions/ interactions of multiple components are studied simultaneously. The presented study will investigate the possibilities that the dose-volume RILD models can be improved by combining the previously found markers.

REFERENCES

[1] C. C. Ling, X. A. Li, W. R. Hendee, Over the next decade the success of radiation treatment planning will be judged by the immediate biological response of tumor cells rather than by surrogate measures such as dose maximization and uniformity, Medical Physics 32 (7) (2005) 2189–2192.

[2] S. M. Bentzen, Preventing or reducing late side effects of radiation therapy: radiobiology meets molecular pathology, Nat Rev Cancer 6 (9) (2006) 702–713.

[3] P. Rubin, D. W. Siemann, D. L. Shapiro, J. N. Finkelstein, D. P. Penney, Surfactant release as an early measure of radiation pneumonitis, International Journal of Radiation Oncology*Biology*Physics 9 (11) (1983) 1669 – 1673.

[4] K. Fleckenstein, B. Gauter-Fleckenstein, I. L. Jackson, Z. Rabbani, M. Anscher, Z. Vujaskovic, Using biological markers to predict risk of radiation injury, Seminars in Radiation Oncology 17 (2) (2007) 89 – 98.

[5] P. G. Tsoutsou, M. I. Koukourakis, Radiation pneumonitis and fibrosis: Mechanisms underlying its pathogenesis and implications for future research, International Journal of Radiation Oncology*Biology*Physics 66 (5) (2006) 1281 – 1293.

[6] P. Rubin, J. Finkelstein, D. Shapiro, Molecular biology mechanisms in the radiation induction of pulmonary injury syndromes: Interrelationship between the

alveolar macrophage and the septal fibroblast, International Journal of Radiation Oncology*Biology*Physics 24 (1) (1992) 93 – 101.

[7] K. Valerie, A. Yacoub, M. P. Hagan, et al., Radiation-induced cell signaling: inside-out and outside-in, Mol Cancer Ther 6 (3) (2007) 789–801.

[8] C. B. Westbury, J. R. Yarnold, Radiation fibrosis–current clinical and therapeutic perspectives, Clin Oncol (R Coll Radiol) 24 (10) (2012) 657–72.

[9] C. Almeida, D. Nagarajan, J. Tian, et al., The role of alveolar epithelium in radiation-induced lung injury, PLoS One 8 (1) (2013) e53628.

[10] K. Fleckenstein, L. Zgonjanin, L. Chen, et al., Temporal onset of hypoxia and oxidative stress after pulmonary irradiation, International Journal of Radiation Oncology*Biology*Physics 68 (1) (2007) 196–204.

[11] D. Shweiki, A. Itin, D. Soffer, E. Keshet, Vascular endothelial growth factor induced by hypoxia may mediate hypoxia-initiated angiogenesis, Nature 359 (6398) (1992) 843–5.

[12] S. N. Ghosh, R. Zhang, B. L. Fish, et al., Renin-angiotensin system suppression mitigates experimental radiation pneumonitis, International Journal of Radiation Oncology*Biology*Physics 75 (5) (2009) 1528–1536.

[13] Biomarkers Definitions Working Group, Biomarkers and surrogate endpoints: preferred definitions and conceptual framework, Clin Pharmacol Ther 69 (3) (2001) 89–95.

[14] P. Okunieff, Y. Chen, A. K. Maguire, David J.and Huser, Molecuar markers of radiation-related normal tissue toxicity, Cancer and Metastasis Reviews 27 (3) (2008) 363–l374.

[15] C. A. Dinarello, Proinflammatory cytokines, Chest 118 (2) (2000) 503–508.

[16] N. Mukaida, Pathophysiological roles of interleukin-8/CXCL8 in pulmonary diseases, American Journal of Physiology - Lung Cellular and Molecular Physiology 284 (4) (2003) L566–L577.

[17] P. Rubin, C. J. Johnston, J. P. Williams, S. McDonald, J. N. Finkelstein, A perpetual cascade of cytokines post-irradiation leads to pulmonary fibrosis, International Journal of Radiation Oncology*Biology*Physics 33 (1) (1995) 99 – 109.

[18] Y. Chen, P. Rubin, J. Williams, E. Hernady, T. Smudzin, P. Okunieff, Circulating IL-6 as a predictor of radiation pneumonitis, International Journal of Radiation Oncology*Biology*Physics 49 (3) (2001) 641 – 648.

[19] D. Arpin, D. Perol, J.-Y. Blay, et al., Early variations of circulating interleukin-6 and interleukin-10 levels during thoracic radiotherapy are predictive for radiation pneumonitis, Journal of Clinical Oncology 23 (34) (2005) 8748–8756.

[20] J. P. Hart, G. Broadwater, Z. Rabbani, et al., Cytokine profiling for prediction of symptomatic radiation-induced lung injury, International Journal of Radiation Oncology*Biology*Physics 63 (5) (2005) 1448 – 1454.

[21] J.-J. Lebrun, The dual role of tgf in human cancer: From tumor suppression to cancer metastasis, ISRN Molecular Biology 2012 (2012) 28.

[22] M. S. Anscher, T. Murase, D. M. Prescott, et al., Changes in plasma TGFβ levels during pulmonary radiotherapy as a predictor of the risk of developing radiation pneumonitis, International Journal of Radiation Oncology*Biology*Physics 30 (3) (1994) 671 – 676.

[23] F.-M. Kong, S. Wang, Nondosimetric risk factors for radiation-induced lung toxicity, Seminars in Radiation Oncology 25 (2) (2015) 100 – 109.

[24] K. D. Jaeger, Y. Seppenwoolde, H. H. Kampinga, L. J. Boersma, J. S. Belderbos, J. V. Lebesque, Iplasma transforming growth factor-β levels in radiotherapy for nonsmall-cell lung cancer, International Journal of Radiation Oncology*Biology*Physics 58 (5) (2004) 1378 – 1387.

[25] A. Novakova-Jiresova, M. M. van Gameren, R. P. Coppes, H. H. Kampinga, H. J. Groen, Transforming growth factor-β plasma dynamics and post-irradiation lung injury in lung cancer patients, Radiotherapy and Oncology 71 (2) (2004) 183 – 189.

[26] B. Bénéteau-Burnat, B. Baudin, Angiotensin-converting enzyme: Clinical applications and laboratory investigations on serum and other biological fluids, Critical Reviews in Clinical Laboratory Sciences 28 (5-6) (1991) 337–356.

[27] L. Zhao, L. Wang, W. Ji, et al., Association between plasma angiotensin-converting enzyme level and radiation pneumonitis, Cytokine 37 (1) (2007) 71 – 75.

[28] S. Bracci, M. Valeriani, L. Agolli, V. D. Sanctis, R. M. Enrici, M. F. Osti, Renin-angiotensin system inhibitors might help to reduce the development of symptomatic radiation pneumonitis after stereotactic body radiotherapy for lung cancer, Clinical Lung Cancer.

[29] J. Kharofa, E. P. Cohen, R. Tomic, Q. Xiang, E. Gore, Decreased risk of radiation pneumonitis with incidental concurrent use of angiotensin-converting

enzyme inhibitors and thoracic radiation therapy, International Journal of Radiation Oncology*Biology*Physics 84 (1) (2012) 238 – 243.

[30] A. Molteni, J. E. Moulder, E. F. Cohen, et al., Control of radiation-induced pneumopathy and lung fibrosis by angiotensin-converting enzyme inhibitors and an angiotensin II type 1 receptor blocker, International Journal of Radiation Biology 76 (4) (2000) 523–532.

[31] X. Chen, X. Kong, Z. Zhang, et al., Alpha-2-macroglobulin as a radioprotective agent: a review, Chin J Cancer Res 26 (5) (2014) 611–21.

[32] M. Mihailovic, V. Milosevic, I. Grigorov, et al., The radioprotective effect of alpha2-macroglobulin: a morphological study of rat liver, Med Sci Monit 15 (7) (2009) 188–93.

[33] J. H. Oh, J. M. Craft, R. Townsend, J. O. Deasy, J. D. Bradley, I. El Naqa, A bioinformatics approach for biomarker identification in radiation-induced lung inflammation from limited proteomics data, Journal of Proteome Research 10 (3) (2011) 1406–1415.

[34] Anthony O'Regan, The role of osteopontin in lung disease, Cytokine and Growth Factor Reviews 14 (6) (2003) 479 – 488, cytokines and Lung Disease.

[35] S. R. Rittling, A. F. Chambers, Role of osteopontin in tumour progression, Br J Cancer 90 (10) (2004) 1877–81.

[36] F. Takahashi, K. Takahashi, T. Okazaki, et al., Role of osteopontin in the pathogenesis of bleomycin-induced pulmonary fibrosis, Am J Respir Cell Mol Biol 24 (3) (2001) 264–71.

[37] Q. T. Le, C. Kong, P. W. Lavori, et al., Expression and prognostic significance of a panel of tissue hypoxia markers in head-and-neck squamous cell carcinomas, Int J Radiat Oncol Biol Phys 69 (1) (2007) 167–75.

[38] G. Wohlleben, A. Scherzad, A. Güttler, et al., Influence of hypoxia and irradiation on osteopontin expression in head and neck cancer and glioblastoma cell lines, Radiation Oncology 10 (1) (2015) 1–8.

[39] N. Kohno, S. Kyoizumi, Y. Awaya, H. Fukuhara, M. Yamakido, M. Akiyama, New serum indicator of interstitial pneumonitis activity. sialylated carbohydrate antigen kl-6, Chest 96 (1) (1989) 68–73.

[40] R. Hara, J. Itami, T. Komiyama, D. Katoh, T. Kondo, Serum levels of KL-6 for predicting the occurrence of radiation pneumonitis after stereotactic radiotherapy for lung tumors, Chest 125 (1) (2004) 340–4.

[41] J. R. Wright, Pulmonary surfactant: a front line of lung host defense, J Clin Invest 111 (10) (2003) 1453–5.

[42] H. Takahashi, Y. Imai, T. Fujishima, et al., Diagnostic significance of surfactant proteins a and d in sera from patients with radiation pneumonitis, European Respiratory Journal 17 (3) (2001) 481–487.

[43] H. Yamashita, S. Kobayashi-Shibata, A. Terahara, et al., Prescreening based on the presence of CT-scan abnormalities and biomarkers (KL-6 and SP-D) may reduce severe radiation pneumonitis after stereotactic radiotherapy, Radiat Oncol 5 (2010) 32.

[44] M. Yin, Z. Liao, Z. Liu, et al., Functional polymorphisms of base excision repair genes xrcc1 and apex1 predict risk of radiation pneumonitis in patients with

non-small cell lung cancer treated with definitive radiation therapy, Int J Radiat Oncol Biol Phys 81 (3) (2011) e67–73.

[45] X. Yuan, Z. Liao, Z. Liu, et al., Single nucleotide polymorphism at rs1982073:t869c of the tgfbeta 1 gene is associated with the risk of radiation pneumonitis in patients with non-small-cell lung cancer treated with definitive radiotherapy, J Clin Oncol 27 (20) (2009) 3370–8.

[46] X. Pu, L. Wang, J. Y. Chang, et al., Inflammation-related genetic variants predict toxicity following definitive radiotherapy for lung cancer, Clinical Pharmacology and Therapeutics 96 (5) (2014) 609–615.

[47] G. C. Barnett, C. E. Coles, R. M. Elliott, et al., Independent validation of genes and polymorphisms reported to be associated with radiation toxicity: a prospective analysis study, The Lancet Oncology 13 (1) (2012) 65–77.

[48] M. S. Anscher, F.-M. Kong, In regard to De Jaeger et al.: Significance of plasma transforming growth factor-beta levels in radiotherapy for non-small-cell lung cancer: International journal of radiation oncology*biology*physics 2004;58:1378-1387), International Journal of Radiation Oncology*Biology*Physics 61 (4) (2005) 1276 – 1277.

[49] K. De Jaeger, Y. Seppenwoolde, J. V. Lebesque, H. H. Kampinga, In response to Drs. Anscher and Kong, International Journal of Radiation Oncology*Biology*Physics 63 (1) (2005) 308.

[50] N. Barthelemy-Brichant, J. L. David, L. Bosque, et al., Increased TGFβ plasma level in patients with lung cancer: potential mechanisms, European Journal of Clinical Investigation 32 (3) (2002) 193–198.

[51] D. J. Grainger, D. E. Mosedale, J. C. Metcalfe, TGF-β in blood: a complex problem, Cytokine & Growth Factor Reviews 11 (12) (2000) 133 – 145.

[52] U. Sauer, M. Heinemann, N. Zamboni, Getting closer to the whole picture, Science 316 (5824) (2007) 550–551.

[53] L. Feinendegen, P. Hahnfeldt, E. E. Schadt, M. Stumpf, E. O. Voit, Systems biology and its potential role in radiobiology, Radiat Environ Biophys 47 (1) (2008) 5–23.

CHAPTER 4
Analytical modelling of regional radiotherapy dose response of lung

Contents

4.1 Introduction

The current NTCP models, such as the Lyman-Kutcher-Burman (LKB) model, are static in design; Temporal change in dose-response relationship is neglected in model design. As explained in the previous chapter, radiation fibrosis modelling with a dose-response curve could be a useful method to gauge patient-specific dose sensitivity while preserving spatial information. This study is aimed at investigating an effect of the follow-up length on NTCP model parameters, enabling time-dependent prediction of RILD. Impacts of regional sensitivity and a dose calculation algorithm on the NTCP model were also investigated.

The study presented in this chapter was published as the following paper: "Analytical modeling of regional radiotherapy dose response of lung" authored by: Sangkyu Lee, Gabriela Stroian, Neil Kopek, Mahmood AlBahhar, Jan Seuntjens and Issam El Naqa. Phys. Med. Biol. 57(11): 3309-21, 2012.

4.2 Background

Radiation-induced lung disease (RILD) is a major constraint to radiotherapy (RT) dose escalation in the treatment of non-small cell lung cancer (NSCLC): a benefit to tumour control by dose escalation is often offset by an increased risk for RILD. In order to achieve an optimal treatment outcome, RT treatment plans have to be evaluated in terms of the patient-specific risk-benefit relationship, which is derived from accurate outcome prediction.

Three-dimensional (3D) planned dosimetric information is considered to be a primary determinant for the prediction of the risk for RILD. The 3D dose distribution is often reduced to a number of dose-volumetric parameters such as mean lung dose

(MLD) or percent lung volume receiving more than a certain dose D (VD). Several studies indicated the correlation between these dosimetric surrogates and the degree of RILD which is often measured by the frequency of symptomatic or asymptomatic radiation pneumonitis or fibrosis [1] [2] [3] [4] [5] [6].

Alternatively, radiographic endpoints such as local change in tissue density [7] [8] [9] [10] [11] [12], perfusion and ventilation [13] [14] have been used to establish local dose-response relationships. RILD results in the replacement of lung parenchyma within radiation fields by relatively dense material (exudates or fibrotic tissues), which is visible on radiographic imaging modalities such as chest X-rays or computed tomography (CT) imaging. Radiation pneumonitis, an early phase of the RILD, radiographically manifests itself by forming ground-glass opacity in an area immediately around the tumour [15]. Six to nine months after completion of RT, this pattern can evolve into radiation fibrosis which features the deposition of collagenous scar tissue [15].

Radiography-derived RILD risk modelling has several advantages over symptom-based risk modelling. First, radiographic change tends to occur more frequently than symptomatic manifestation [16], which provides larger sample size. Secondly, the benefit of investigating local dose-response by image analysis is illustrated by the variation in results in the literature on the optimal dose-volumetric parameter that predicts RP risk most accurately [5]. Although many cases of radiographic change are subclinical [17], there are several studies reporting strong correlation between increase in radiographic lung density and lung functions such as breathing rates

[18], perfusion and ventilation [13]. The listed characteristics of radiographic RILD encourage its use as a surrogate to the degree of radiation-induced lung toxicity.

A number of authors investigated local dose-response relationships from changes in CT density [7] [8] [9] [10] [11] [12]. However, analytical generalization of the local dose-response relationships based on proper radiobiological modelling has not been attempted. Furthermore, despite the temporal dynamics of RILD shown both radiographically and symptomatically, the time-dependency of RILD risk is often overlooked. Considering that RILD-related symptoms are diagnosed within a year post-treatment, we believe that long-term follow-up of the radiographic RILD over its time course may pave the way to better quantitatively assess its complication and monitor its progression allowing for better intervention instead of using arbitrarily selected time points. This work extends the methods developed by Stroian et al. [8] to quantitatively analyze follow-up CT images and model dose- and time-response of lung parenchyma using the Lyman-Kutcher-Burman (LKB) normal tissue complication probabilities (NTCP) model [19] [20] [21]. The impact of dose calculation algorithm on RILD risk models is investigated by comparing the derived dose-response obtained from Monte-Carlo (MC) dose with convolution-superposition based algorithms. In particular, we investigated the time dependency of the extracted LKB parameters with the possibility to generalize the model to account for temporal variations.

4.3 Materials and Methods

4.3.1 Patient characteristics

Twenty-one patients with stage III non-small-cell lung cancer (NSCLC) were selected for the study. The patients were treated at the Montreal General Hospital (MGH) between 2002 and 2007. The patient group consisted of 15 males and 6 females (median age: 68). 15 patients received conventionally fractionated (60 Gy in 30 fractions) 3D-conformal RT (3D-CRT) combined with chemotherapy, while 6 patients were scheduled for hypofractionation (52.5 Gy in 15 fractions) and were not eligible for chemotherapy. The RT was planned using Eclipse (Varian Medical Systems, Palo Alto, CA) treatment planning system (pencil beam convolution, no heterogeneity correction). The treatment plan consisted of coplanar 6 MV and 18 MV beams delivered by the Varian Clinac 21X linear accelerator.

The patients underwent a pre-treatment planning CT and multiple post-treatment follow-up CT or PET/CT scans. The follow-up imaging was performed at 57 different post-RT times ranging from 0 day to 626 days. The median length of the follow-up was 352 days. A total of 58 follow-up CT images were acquired from all the patients. The interval between the follow-up studies was not regular as long-term follow-up was subject to patient health or death. Planning CT scan was performed (140 kVp, mAs vary over scans) using Phillips AcQSim CT scanner. During the planning CT scan, patients were lying on a flat table in a supine and overhead arm position and allowed to breathe freely. Follow-up imaging included chest CT scans and whole-body PET/CT scans. The chest CT scan was performed while patients on a curved bed were holding their breath at maximum inhalation. Contrast agents were applied

in some cases of the chest CT scans. The whole-body PET/CT scan was taken without contrast agents under the free breathing condition. The CT component of the PET/CT scan for attenuation correction was used for the image analysis.

4.3.2 RILD scoring

Follow-up CT images were registered to the corresponding planning CT image using a 12-parameter affine (non-deformable) iterative optimization technique implemented by the registration software ANIMAL [22]. Pixel values in the registered follow-up and planning CT images were converted to physical density using a calibration curve specific to the used scanner model. The calibration curve for the particular scanner model was created by matching the average CT density values of anatomical structures in the 20 images produced by the model to reference physical density values for the corresponding sites [23]. The registered follow-up images were then subtracted from the planning CT image to obtain the change in physical density at each voxel of the planning CT image. The region of interest (ROI) for RILD scoring was confined within the lung volume outlined by the contours which were previously drawn on the planning CT to be used in treatment planning. The planning target volume (PTV) was excluded from the ROI to prevent the segmentation of any possible tumour recurrence. The CT voxels that their changes in physical density fall within a pathologically relevant range (0.123-0.799 g/cm^3) [10] were segmented as RILD distribution, which was later corrected manually by an experienced radiologist (M.B.) for artefacts that could result from registration inaccuracy such as mismatching lung contours, for instance. Severity of RILD was measured by the integrated volume of the corrected RILD distribution normalized to the lung volume. RILD

was also scored by a radiation oncologist (N.K.) according to three radiation toxicity criteria: 1) late effects in normal tissues subjective, objective, management and analytic scales (LENT-SOMA), 2) Common Toxicity Criteria for Adverse Effect version 4.0 (CTCv4), 3) toxicity criteria of the Radiation Therapy Oncology Group and the European Organization for Research and Treatment of Cancer (RTOG/EORTC) [24].

4.3.3 Dose calculation

Radiation dose distribution in the thoracic region was recalculated using the two algorithms: the Analytical Anisotropic Algorithm (AAA) with heterogeneous correction and an MC method. AAA dose calculations were performed using the Eclipse Treatment Planning Software from Varian^{TM}. MC dose calculation was performed on the McGill Monte Carlo Treatment Planning (MMTCP) platform [25] which was used to create input files for the EGSnrc-based simulation codes based on the imported RT plans (BEAMnrc and DOSXYZnrc) and control the simulation. BEAMnrc created a phase space file at 30 cm away from the isocenter using a commissioned beam model and a given accelerator head configuration. DOSXYZnrc code simulated the interaction of the particles from the phase-space file in the patient body to calculate patient dose from the delivery of a given monitor unit (MU). The dose was initially scored on a 0.5 X 0.5 X 0.5 cm^2 grid for the both algorithms, which was resampled to the same grid dimension as planning CT. In order to account for the variability in dose fractionation, the resampled voxel dose was converted to the normalized total dose (NTD) (equation 4.1):

$$NTD = \frac{nd_f(\alpha/\beta + d_f)}{\alpha/\beta + 2} \tag{4.1}$$

where n: number of fractions, d_f: dose-per-fraction and α and β are radiobiological sensitivity parameters according to the linear-quadratic (LQ) model [26]. The value of α/β was chosen to be 3 Gy [27].

4.3.4 Dose-RILD correlation

The calculated dose distribution and RILD distribution were superimposed onto the planning CT image. Differential dose-volume histograms were created for the segmented RILD volume and the lung volume in which the RILD was scored. The RILD segmentation and lung volume were divided into ipsilateral and contralateral sides. A dose-response curve at the j-th follow-up study was represented by the probability of RILD as a function of dose bin Di following the definition by Rosen [10] (equation 4.2):

$$P_j(D_i) = N(injr_j, D_i)/N(lung_j, D_i) \tag{4.2}$$

where $N(injr_j, D_i)$ and $N(lung_j, D_i)$ are the number of voxels in the RILD and lung segmentation, respectively, corresponding to the dose bin D_i. Uncertainty on the value of was estimated as independent combination of four sources of error that impact the measurement of : 1) dose calculation uncertainty, 2) matching uncertainty between planning and follow-up CT images, 3) CT number-to-density calibration. 4) Poissonian voxel counting errors [28].

4.3.5 Analytical modeling of RILD risk

The collected dose-response curves from the 21 patients were grouped into 6 follow-up periods according to the time elapsed since the completion of the radiotherapy. The time interval was 3 months except for the sixth period which included all the follow-up studies after 15 months. The Probit-like function of the LKB [19] was fit to the collection of measured complication probability data $P_j(D_i)$ given by equation 4.2, as follows:

$$\bar{P}(D) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t(D)} exp(-\frac{x^2}{2})dx \tag{4.3}$$

where:

$$t(D) = \frac{D - TD_{50}}{mTD_{50}} \tag{4.4}$$

The $P(D)$ and $\bar{P}(D)$ stand for the measured RILD probability data and the fitted function, respectively. The two parameters characterizing the model, TD50 (tolerance dose at 50% probability of complication) and m (governs the slope of the dose-response curve), were determined for each follow-up period based on maximum likelihood estimation [29]. The log-likelihood function associated with the LKB model was created with each data point weighted by the inverse-square of the uncertainty value associated with $P(D_i)$. In addition, the threshold dose for RILD was determined from the shape of the best-fit function by finding the x-intercept of the tangent line to the curve at $D = TD50$ as shown in figure 4.1.

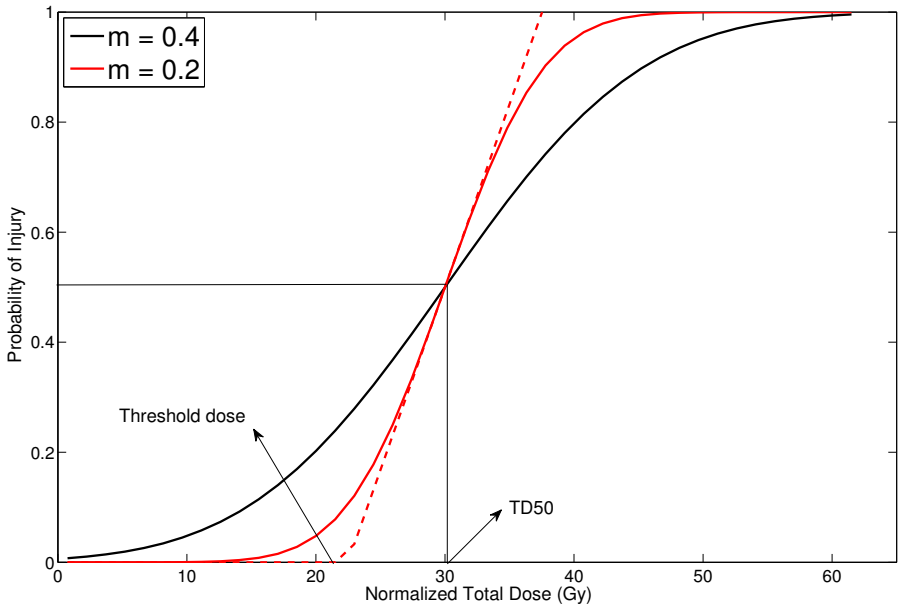

Figure 4.1: Illustration of deriving a threshold dose from the shape of the LKB function drawn for two different slope parameter values (m). Tangent line (dotted line) to the curve is drawn at dose = TD50.

4.4 Results

4.4.1 RILD scoring

Radiographic changes were confirmed in 49 out of 58 follow-up images by the radiologists who examined the images at the time of follow-up. 10 follow-up images showed the RILD both on the ipsilateral and contralateral lung and 38 follow-ups on the ipsilateral side only. Time dependency of the severity of RILD is illustrated in figure 4.2, where a significant decrease after 3 months was followed a slowly decreasing period after 12 month.

The follow-up studies were reviewed and graded for clinical toxicity by a radiation oncologist (N.K). Correlations between the normalized RILD volume and the toxicity levels graded by the 3 different scoring schemes were analyzed by Pearsons chi-square test. Due to the lack of symptomatic information at late follow-ups, only the radiological criteria was considered. Pearsons correlation (r) coefficients for the CTCv4, RTOG/EORTC, and SOMA/LENT scoring scheme were respectively 0.623, 0.406, and 0.426, which were all statistically significant ($p<0.05$), demonstrating that the quantitative CT analysis had good agreement with manual assessment by a physician.

4.4.2 Dose-RILD correlation

96% (48/49) of the follow-up studies showed the p-values for Pearsons test less than 0.05 between ipsilateral RILD probability and local dose calculated by MC. The same ratio (96%) of studies with significant correlation was shown when the MC dose was replaced by AAA dose. The dose correlation for contralateral RILD probability was weaker: 70% (7/10) of the cases showed significant correlation with MC or AAA

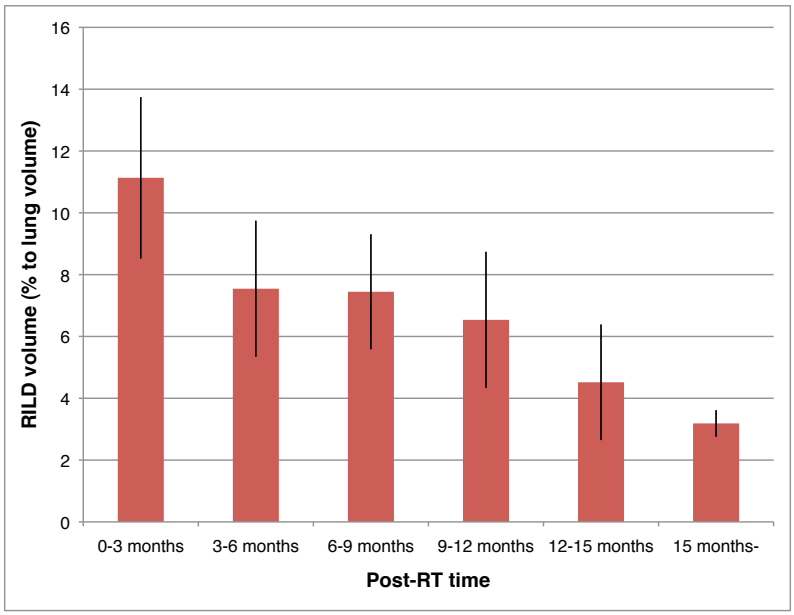

Figure 4.2: Histogram showing the population change in the scored RILD volume over time binned into 3 months periods. Error bars represent 1 standard deviation.

dose. When the contralateral and ipsilateral lung volumes are combined, the strength of correlation was slightly higher for MC (r=0.740) than AAA (r=0.735) although the differences were not statistically significant as revealed by the Student t-test (p = 0.656).

The maximum likelihood method determined optimal fit of the LKB model to the measured RILD-dose relationship for each follow-up period. Figure 4.3 illustrates the association of the fitted LKB model to the measured RILD probability data. The two best-fit parameters (TD50 and m) and the derived threshold dose for the 6 follow-up periods and the 2 dose calculation algorithms are illustrated in figure 4.4. TD50 reached its peak (AAA: 128 Gy/MC: 132 Gy), which occurred coincidently with the decrease in the segmented RILD volume. After 15 months, TD50 fell closely to its

initial level (AAA: 89 Gy/MC: 89 Gy). The value of m showed significant decrease (0.041/0.041 for AAA/MC doses, respectively) from 0.47/0.50 for the first follow-up period to 0.33/0.33 for the second follow-up period (3-6 months) after which no significant change was found. Time-change in the derived threshold dose illustrated again that 0-3 months follow-up period was statistically different from the other periods: The threshold dose (AAA: 28 Gy/MC: 27 Gy) was significantly lower than the other time periods regardless of dose calculation algorithm. The time variation in the threshold dose was not significant after 3 month post-RT.

4.4.3 Dependency on tumour position

We also investigated the influence of geographical parameters, specifically, the superior-inferior tumour position on RILD risk [30]. Tumour position was represented by the coordinates of the center-of-mass (COM) of a PTV volume. The superior-inferior position of the PTV COM (COMSI) was normalized to the extent of total lung volume on the planning CT with 0 and 1 assigned for the superior and inferior end respectively. Patient median for COMSI was 0.682. The patients with their COMSI above this value were separated from the rest of the patients to form two subgroups each of which represented superior and inferior tumour position. For each subgroup, curve fitting was performed on the dose response curve averaged over all follow-up periods. Figure 4.5 demonstrates that local RILD probability was generally higher for the patients with inferior tumour position. This difference between the two subgroups was also illustrated in best-fit parameters (table 4.1). The inferior tumour position subgroup showed significantly higher value for m (p=0.004) and slightly lower TD50 (p=0.191) and threshold dose (p=0.058) than the superior counterpart.

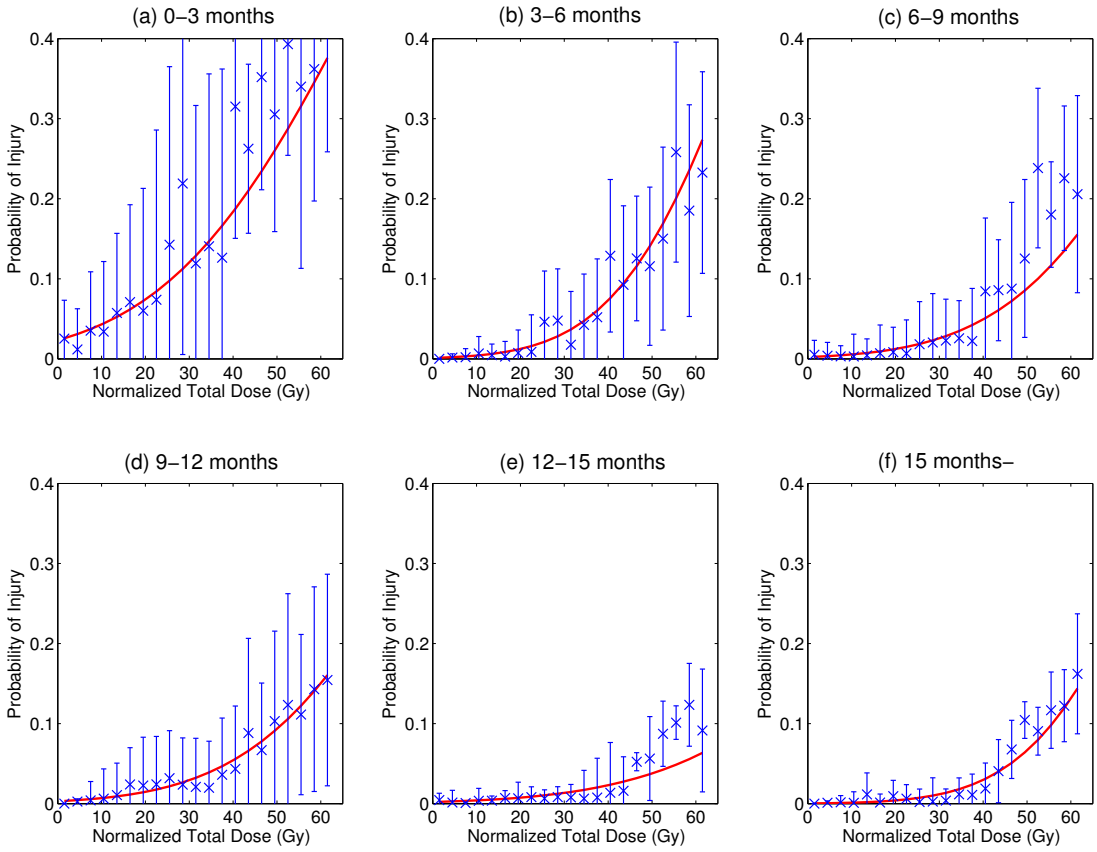

Figure 4.3: Probability of RILD as a function of MC dose for 6 follow-up periods (a)-(f). x symbols represent patient average of individual data points weighted by uncertainty values from measurements. Overlaid curves are the fitted LKB model. Error bars are showing 1 standard deviation.

Table 4.1: Best-fit parameters of the LKB models, averaged over all follow-up periods, from two patient subgroups as shown in figure 4.5. Confidence intervals are shown in parentheses.

Parameters	Patient subgroup	
	COMSI>median	COMSI<median
m	0.43 (0.39,0.46)	0.33 (0.30,0.36)
TD50 (Gy)	99.3 (86,7,112.0)	89.3 (81.2,97.4)
Threshold dose (Gy)	47.8 (43.2,52,4)	54.1 (50.3,58.0)

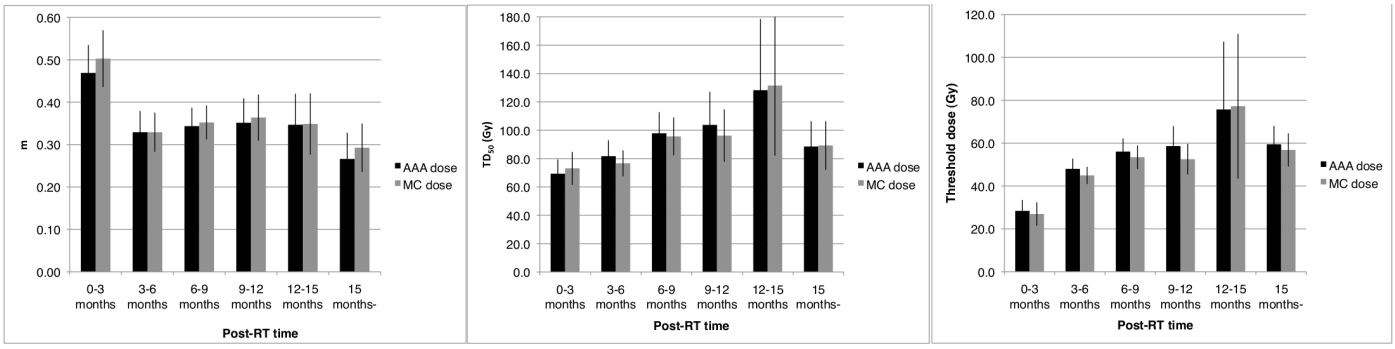

Figure 4.4: Time variation in the 3 best-fit parameters: (a) m, (b) TD50, and (c) threshold dose resulting from AAA and MC dose calculation algorithms. Error bars represent 95% confidence intervals.

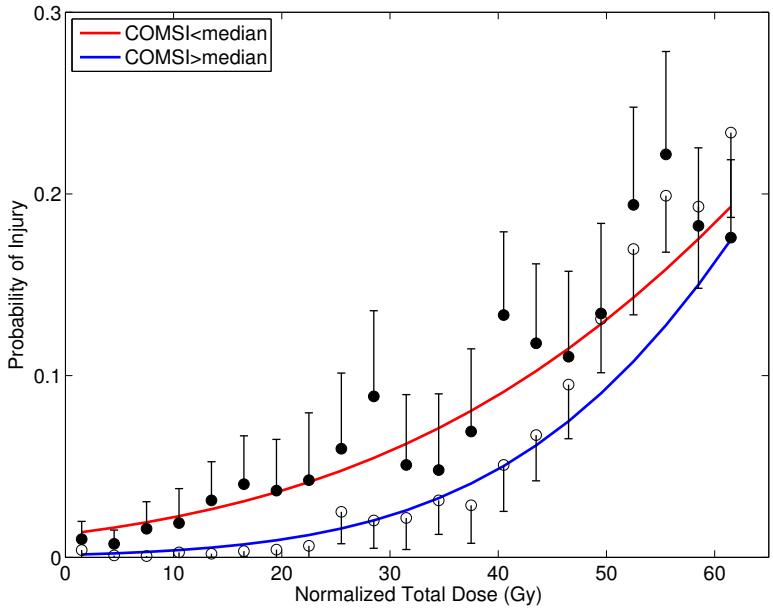

Figure 4.5: Two different dose responses for the patient subgroup with superior tumour position (COMSI>median, open circles) and inferior tumour position (COMSI<median, filled circles). Weighted average probability values for each subgroup are shown as scattered symbols which are overlaid with best-fit LKB model in each case. Error bars are symmetric (only one half is shown) with the magnitude of 1 standard error. Dose was calculated with the Monte-Carlo method.

4.5 Discussion

Fitting of analytical function to dose response has been employed by many studies which were initiated by Burman [21]. Sigmoid-type functions are widely accepted as the empirical relationship between dose and dose-induced toxicity. It is generally known that local dose positively correlates with change in lung density measured by CT imaging. To the best of our knowledge, however, application of the analytical NTCP models such as the LKB model to radiographic dose-response has not been performed explicitly so far. The proposed analytical NTCP model can be used for weighting 3D dose distribution to project the overall radiographic change associated with a designed treatment plan. Also, the time-dependent characteristics of the model can extend the model to monitor changes and better predict the long-term effects of radiation therapy on lung tissues as well as acute toxicity. However, patient-specific correlations between the degree of CT change and risk of whole-organ complication should be provided to promote clinical use of the proposed model. If future studies can reveal the clinical significance of RILD severity scored using our methods, the proposed model might be used for treatment planning optimization as a cost function for regulating lung dose.

We compared the time change in the severity of RILD with the results from other studies [7] [8] [9] [10] [11] [12]. The overall damage was most severe at the first (0-3 months) follow-up period and its partial recovery towards the baseline apparently begun at the second follow-up period (3-6 months) and continued till the last follow-up period. This result is inline with the study by Theuws et al.[31] who reported that local pulmonary injury assessed by both CT and SPECT imaging recovered between

3 and 18 months followed by a stabilization period after 18 months. Similarly, Hof et al. [12] reported that significant CT changes after stereotactic body radiotherapy (SBRT) peaked at 4 months post-RT followed by its gradual decrease. Due to the coarse binning scheme we used, we were unable to pinpoint the exact time for such a peak.

Curve fitting of dose-response curves obtained at different time points revealed that the first follow-up period (0-3 months) was significantly different from later follow-up periods, which might reflect the distinct temporal dynamics of dose-RILD correlation during the first three months. TD50 and m represent the horizontal location and the steepness of the curve, respectively. It can be inferred from the time-change of the best-fit parameters that there is an increase in TD50 and threshold dose coinciding with decrease in m. This implies that the curve shifts to higher doses while it becomes steeper. This might suggest the transition from acute damage for which the radiographic change is linear with dose to late damage for which the change is rather confined to high dose region.

It is of clinical interest to identify the minimal dose above which radiation induces change in clinical lung function. This study systematically derived corresponding threshold doses from dose-response functions at different post-RT time periods. A threshold dose could be more pertinent to the decision making in treatment planning than TD50, which in many instances exceeded the investigated dose range. The derived threshold value (PBC: 28 Gy/AAA: 28 Gy/MC: 27 Gy) at the earliest time period (0-3 months) coincides with the results from other investigations [10] [8]

where it varies around 30 Gy. The threshold was significantly elevated at later periods, again underlining possibly the distinction between early and late radiographic changes.

Impact of the accuracy of dose calculation algorithm on dose-response modeling has often been overlooked. Although the MC method is considered to be a gold standard for tissue dose calculation [32], its use for normal tissue dose response has been limited. Stroian et al [8] suggested better dose-RILD correlation with MC method than pencil-beam calculation without heterogeneity correction. In recent lung toxicity studies, however, pencil beam algorithm has broadly been replaced by more advanced algorithms such as AAA which is shown to improve accuracy in heterogeneity [33]. Thus, we extended Stroian et al.s work to include AAA algorithm in comparison with MC calculation. We found a few follow-up studies where usage of MC yielded stronger dose-RILD probability correlation than AAA, which can be seen by the higher average correlation coefficient by using MC. However, we found only a slight increase in the coefficient for MC when it was averaged over the entire population. Furthermore, replacing AAA by MC resulted in no significant difference in the shape of the best-fit dose-response functions. This is possibly because of inherent patient heterogeneity in dose response that could have washed out the difference in dose distributions. The patient-dependent discrepancy between these algorithms and MC suggests that MC should still be considered as a primary choice for dose calculation algorithms when patient-specific prediction is to be made. Moreover, steeper dose-response relationships of RILD for late follow-up periods stress the importance of the accuracy of local dose calculation.

Among the recognized factors that affect RILD risk are the superior-inferior tumour volume location and chemotherapy agents. We investigated the impact of chemotherapy on RILD risk by performing separate fitting for the subgroup which received chemotherapy (n=15) and the subgroup which did not (n=6). Although we found significantly higher TD50 and threshold dose from the chemotherapy-less subgroup, the fitting error was too high for that subgroup due to its small sample size. Moreover, it could not be ruled out that different fractionation schemes might have contributed to the differences between the subgroups despite NTD normalization. There are studies showing increased risk for radiation pneumonitis for patients who have tumours located in the base of the lung [6] [3]. Our results indicate that irradiation of lower lung is associated with increased incidence for radiographic RILD. Difference between subgroup means was not significant at 95% confidence level due to high patient heterogeneity. However, we found that dose response was more linear when the tumour was inferiorly located, giving rise to increased risk in the intermediate dose region (10-40 Gy). This might be another supporting result for the argument that there exists regional variation in radiosensitivity in lung. We still have to investigate the impact of respiratory motion on patient dose distribution and the resulting RILD-dose correlation, which is expected to be more pronounced in the inferior lung. Differences between the superior and inferior tumour position groups in the intermediate dose region might be attributed to the fact that dose calculation based on a static planning CT set underestimates the lung volume receiving a mid-range dose compared to the dose accumulated through the entire breathing cycle [34].

Clinical importance of the presented RILD dose-response model is mainly limited by the relatively small size of patients group (n=21). Scarcity of late follow-up imaging resulting from small sample size as well as limited survival of the NSCLC patients was an obstacle to keep the 3-months time resolution after 15 months post-RT. Limited availability of homogeneous patients accounts for high uncertainty on likelihood fitting, which is specifically clear in the 12-15 months follow-up period. The major factor contributing to the heterogeneous response might be inclusion of the hypofractionated patient group. Although the LQ model was used to correct for difference in fractionation, the alpha-beta ratio we used was not precisely determined. Apart from expanding the patient dataset, we are planning to apply this method to other imaging modalities such as single-photon emission computed tomography (SPECT) or hyperpolarized magnetic resonance imagimg (MRI) [35] to complement the limitation of CT to anatomical representation. We have not tested the applicability of the model to SBRT-induced toxicity. Analysis of follow-up CT images from SBRT-treated patients for the verification of the model is underway.

4.6 Conclusion

In summary, the presented work characterized local dose response of lung tissue after radiotherapy by fitting the LKB model to complication probability data derived from quantitative CT image analysis and Monte-Carlo dose calculation. Dependency of the extent of RILD on local dose, post-treatment time, and tumour SI position was indicated. Time-variability in the best-fit parameters indicated a clear distinction between early- (0-3 months) and late- (3 months-) responding behaviors

of RILD. After validation on an expanded patient dataset, the proposed analytical dose-response can potentially be used as a dose-constraining function for inverse treatment planning optimization schemes.

4.7 Acknowledgements

This study was supported by Canadian Institute of Health Research (CIHR) grant #MOP: 102550 and #MOP: 114910.

REFERENCES

[1] M. L. Hernando, L. B. Marks, G. C. Bentel, et al., Radiation-induced pulmonary toxicity: a dose-volume histogram analysis in 201 patients with lung cancer, International Journal of Radiation Oncology*Biology*Physics 51 (3) (2001) 650 – 659.

[2] D. Oetzel, P. Schraube, F. Hensley, G. Sroka-Pérez, M. Menke, M. Flentje, Estimation of pneumonitis risk in three-dimensional treatment planning using dose-volume histogram analysis, International Journal of Radiation Oncology*Biology*Physics 33 (2) (1995) 455 – 460.

[3] M. V. Graham, J. A. Purdy, B. Emami, et al., Clinical dose volume histogram analysis for pneumonitis after 3d treatment for non-small cell lung cancer (NSCLC), International Journal of Radiation Oncology*Biology*Physics 45 (2) (1999) 323 – 329.

[4] M. K. Martel, R. K. Ten Haken, M. B. Hazuka, A. T. Turrisi, B. A. Fraass, A. S. Lichter, Dose-volume histogram and 3D treatment planning evaluation of patients with pneumonitis, International Journal of Radiation Oncology*Biology*Physics 28 (3) (1994) 575 – 581.

[5] G. Rodrigues, M. Lock, D. D'Souza, E. Yu, J. V. Dyk, Prediction of radiation pneumonitis by dose-volume histogram parameters in lung cancer-a systematic review, Radiotherapy and Oncology 71 (2) (2004) 127 – 138.

[6] E. D. Yorke, A. Jackson, K. E. Rosenzweig, et al., Dose-volume factors contributing to the incidence of radiation pneumonitis in non-small-cell lung cancer patients treated with three-dimensional conformal radiation therapy, International Journal of Radiation Oncology*Biology*Physics 54 (2) (2002) 329 – 339.

[7] F. B. Geara, R. Komaki, S. L. Tucker, E. L. Travis, J. D. Cox, Factors influencing the development of lung fibrosis after chemoradiation for small cell carcinoma of the lung: Evidence for inherent interindividual variation, International Journal of Radiation Oncology*Biology*Physics 41 (2) (1998) 279 – 286.

[8] G. Stroian, C. Martens, L. Souhami, D. L. Collins, J. Seuntjens, Local correlation between monte-carlo dose and radiation-induced fibrosis in lung cancer patients, International Journal of Radiation Oncology*Biology*Physics 70 (3) (2008) 921 – 930.

[9] J. Ma, J. Zhang, S. Zhou, et al., Regional lung density changes after radiation therapy for tumors in and around thorax, Int J Radiat Oncol Biol Phys 76 (1) (2010) 116–22.

[10] I. I. Rosen, T. A. Fischer, J. A. Antolak, et al., Correlation between Lung Fibrosis and Radiation Therapy Dose after Concurrent Radiation Therapy and Chemotherapy for Limited Small Cell Lung Cancer, Radiology 221 (3) (2001) 614–622.

[11] D. A. Palma, J. van Srnsen de Koste, W. F. Verbakel, A. Vincent, S. Senan, Lung density changes after stereotactic radiotherapy: A quantitative analysis in 50 patients, International Journal of Radiation Oncology*Biology*Physics 81 (4) (2011) 974 – 978.

[12] H. Hof, J. Zgoda, S. Nill, et al., Time- and dose-dependency of radiographic normal tissue changes of the lung after stereotactic radiotherapy, International Journal of Radiation Oncology*Biology*Physics 77 (5) (2010) 1369–1374.

[13] L. J. Boersma, E. M. F. Damen, R. W. de Boer, et al., Estimation of overall pulmonary function after irradiation using dose-effect relations for local functional injury, Radiotherapy and Oncology 36 (1) (1995) 15 – 23.

[14] L. B. Marks, M. T. Munley, G. C. Bentel, et al., Physical and biological predictors of changes in whole-lung function following thoracic irradiation, International Journal of Radiation Oncology*Biology*Physics 39 (3) (1997) 563 – 570.

[15] T. R. Koenig, R. F. Munden, J. J. Erasmus, et al., Radiation injury of the lung after three-dimensional conformal radiation therapy, Am. J. Roentgenol. 178 (6) (2002) 1383–1388.

[16] B. Movsas, T. A. Raffin, A. H. Epstein, J. Charles J. Link, Pulmonary radiation injury, Chest 111 (1997) 1061–1076.

[17] S. L. Kwa, J. V. Lebesque, J. C. Theuws, et al., Radiation pneumonitis as a function of mean lung dose: an analysis of pooled data of 540 patients, International Journal of Radiation Oncology*Biology*Physics 42 (1) (1998) 1 – 9.

[18] M. Takahashi, G. Balazs, Y. Pipman, et al., Radiation-induced lung injury using a pig model: Evaluation by high-resolution computed tomography, Investigative Radiology 30 (2) (1995) 79–86.

[19] J. T. Lyman, Complication probability as assessed from dose-volume histograms, Radiation Research Supplement 8 (1985) pp. S13–S19.

[20] G. J. Kutcher, C. Burman, Calculation of complication probability factors for non-uniform normal tissue irradiation: the effective volume method, International Journal of Radiation Oncology*Biology*Physics 16 (6) (1989) 1623–1630.

[21] C. Burman, G. Kutcher, B. Emami, M. Goitein, Fitting of normal tissue tolerance data to an analytic function, International Journal of Radiation Oncology*Biology*Physics 21 (1) (1991) 123 – 135, three-Dimensional Photon Treatment Planning Report of the Collaborative Working Group on the Evaluation of Treatment Planning for External Photon Beam Radiotherapy.

[22] D. L. Collins, A. C. Evans, Animal: Validation and applications of nonlinear registration-based segmentation, Int J Pattern Recognition Artificial Intelligences 11 (8) (1997) 1271–1294.

[23] D. R. White, H. Q. Woodard, S. M. Hammond, Average soft-tissue and bone models for use in radiation dosimetry, Br J Radiol 60 (717) (1987) 907–13.

[24] J. D. Cox, J. Stetz, T. F. Pajak, Toxicity criteria of the radiation therapy oncology group (RTOG) and the european organization for research and treatment of cancer (EORTC), International Journal of Radiation Oncology*Biology*Physics 31 (5) (1995) 1341 – 1346.

[25] A. Alexander, F. Deblois, G. Stroian, K. Al-Yahya, E. Heath, J. Seuntjens, Mmctp: a radiotherapy research environment for monte carlo and patient-specific treatment planning, Physics in Medicine and Biology 52 (13) (2007) N297–308.

[26] B. G. Douglas, J. F. Fowler, The effect of multiple small doses of x rays on skin reactions in the mouse and a basic interpretation, Radiation research 66 (2) (1976) 401–426.

[27] B. Dubray, M. Henry-Amar, J. H. Meerwaldt, et al., Radiation-induced lung damage after thoracic irradiation for hodgkin's disease: the role of fractionation, Radiotherapy and Oncology 36 (3) (1995) 211–217.

[28] S. Lee, Image-based dose correlation studies on radiation- induced lung injury, Ph.D. thesis, McGill University, Montreal, Quebec (2011).

[29] J. Thames, H. D., M. E. Rozell, S. L. Tucker, K. K. Ang, D. R. Fisher, E. L. Travis, Direct analysis of quantal radiation response data, Int J Radiat Biol Relat Stud Phys Chem Med 49 (6) (1986) 999–1009.

[30] A. J. Hope, P. E. Lindsay, I. E. Naqa, et al., Modeling radiation pneumonitis risk with clinical, dosimetric, and spatial parameters, International Journal of Radiation Oncology*Biology*Physics 65 (1) (2006) 112 – 124.

[31] J. C. M. Theuws, Y. Seppenwoolde, S. L. S. Kwa, et al., Changes in local pulmonary injury up to 48 months after irradiation for lymphoma and breast cancer, International Journal of Radiation Oncology*Biology*Physics 47 (5) (2000) 1201–1208.

[32] I. J. Chetty, B. Curran, J. E. Cygler, et al., Report of the aapm task group no. 105: Issues associated with clinical implementation of monte carlo-based photon and electron external beam treatment planning, Medical Physics 34 (12) (2007) 4818–53.

[33] A. Van Esch, L. Tillikainen, J. Pyykkonen, et al., Testing of the analytical anisotropic algorithm for photon dose calculation, Vol. 33, AAPM, 2006, 11.

[34] M. Ding, F. Newman, L. Gaspar, et al., A 4d treatment planning tool for the evaluation of motion effects on lung cancer treatments, Journal of Physics: Conference Series 102 (1) (2008) 012008.

[35] C. W. Hodge, W. A. Tom, S. B. Fain, S. M. Bentzen, M. P. Mehta, On the use of hyperpolarized helium mri for conformal avoidance lung radiotherapy, Medical Dosimetry 35 (4) (2010) 297–303.

CHAPTER 5
Texture analysis on intra-treatment CT images and and its implication on RILD heterogeneity

Contents

5.1 Background

Chapter 4 modelled patients' radiation response from post-treatment CT density changes. However, predictive potential of such model is limited because the images were acquired after completion of RT delivery. Also, the DRCs from individual patients showed a high degree of heterogeneity, which was partially explained by tumour position. Research in this chapter is intended to address two questions:

- Can the radiologic response heterogeneity be further explained by biological factors?
- Can the heterogeneity be captured early on, before completion of RT, to predict treatment outcome?

In addition, this work explores an emerging area in radiotherapy outcome modelling: *Radiomics*. Radiomics refers to extraction of a large number of medical imaging features for quantifying and monitoring tumour characteristics [1]. It involves not only intensity of each pixels in images, but also how these intensity values are spatially distributed and correlated which is specifically named as *textures*. Refer to the appendix of this text or a paper by Haralick et al. [2] for mathematical methods of quantifying such patterns.

The presented work is considered preliminary. Preparation of a manuscript for publication is underway while methodology is being refined and more patient data is being added. The participating authors are: N. Ybarra, K. Jeyaseelan, M. AlBahaar, S. Faria, N. Kopek, P. Brisebois, and I. El Naqa.

5.2 Introduction

Radiation induced lung disease (RILD) is a dose-limiting side effect from radiotherapy on thoracic sites. It accompanies clinical symptoms such as radiation pneumonitis (RP) and subclinical radiographic changes. [3]: For years, prediction of RP has been challenged by many researchers using lung dose volumetric histogram (DVH) parameters such as mean lung dose (MLD) or Vx (percentage lung volume receiving x Gy or higher). However, these dosimetric models show inconsistency with regards to the best metric [4] and their prediction accuracy is limited [5].

In order to improve prediction accuracy of the conventional RILD models, various clinical and biological factors have been suggested as a patient specific radiosensitivity factor [6]. Among those, serum/plasma measurements of the proteins playing a role in early inflammatory phase of RILD were suggested to have predictive values. Examples of such proteins include transforming growth factor(TGF)-β1 [7] [8], Interleukin(IL)-6 [9], alpha-2-macroglobulin(α2M) [10], angiotensin converting enzyme(ACE) [11] [12]).

High-throughput imaging modalities are another emerging biomarkers for predicting treatment outcomes. Among all, computed tomography (CT) imaging, the most ubiquitous diagnostic tool in radiation oncology, can deliver a high amount of quantitative data extracted from spatial arrangement of grey levels also known as "textures" [2]. CT image features, including textures, can be used for outcome modeling in two ways: First, they can provide a quantitative and objective method of measuring patient-specific toxicity response to compensate for ambiguity of symptom-based scoring [13] [14]. Secondly, CT features acquired before RT completion could

have predictive potential, which is of a prime interest for prediction of lung cancer prognosis [15] [16] [1].

In the context of RILD, a series of works by Cunliffe and colleagues intended to establish a set of the CT features that can faithfully detect clinically significant radiation induced changes in lung tissues [17] [18] and also possesses predictive value for RP [19]. However, the imaging information was acquired after RT completion, which limits its practical value as a predictive factor. Bertelsen et al. [20] investigated intra-fraction CT density change from daily cone-beam CT images and detected a high degree of patient specificity in the magnitude of the change. This result suggests early CT change as an implication of radio-sensitivity which might be related to treatment outcome. In order to establish CT features as a true "biomarker", however, a causal link between CT and biological changes needs to be investigated. Correlation between CT features and direct measurement of biological states, in terms of protein or gene expression, would be a useful exploratory study for this purpose.

This work was conducted to further characterize early CT changes before RT completion using texture analysis. Patient specific patterns in the early texture changes were studied for possible indication of subclinical biological changes and clinical symptomatic outcomes.

5.3 Materials and Methods

5.3.1 Patient characteristics

22 stage III non-small cell lung cancer (NSCLC) patients were prospectively recruited in the period of 2011-2014 according to a data collection protocol approved

by institutional review boards. Patients were considered for this study under the conditions of: 1) survival of at least 6 months, 2) no history of previous lung irradiation, and 3) baseline Karnofsky performance status (KPS) of equal to or greater than 70. All the patients were treated with 60 Gy/ 30 fractions radiotherapy with induction and concomitant chemotherapy. A 3-dimensional (3D) conformal technique was used for RT delivery except one patient treated with Tomotherapy. The data collection protocol involved CT scanning and blood sample collection at 3 time points: day of CT simulation, the 16th and the last fraction. The patients were scored for RP using the Common Toxicity Criteria for adverse events (CTCAE) version 4. Detailed cohort characteristics are shown in table 5.1.

5.3.2 Imaging and biomarker data acquisition

All the CT scans at the 3 time points were performed without contrast on a CT simulator (Philips Brilliance^{TM}, Amsterdam, The Netherlands). During the scan, the patients were positioned on a flat bed without breathing control. The images were acquired in a 4-dimensional mode, but an average image was used for further analysis. Image acquisition protocol was fixed at a routine chest protocol for CT simulation (120 kVp, 275 mA, 3 mm axial resolution). There was slight variation in in-plane resolution between the images due to different field of view: median (range) of a pixel size was 0.98 (0.82,1.17), 0.96 (0.79,1.17) and 0.98 (0.84,1.17) for baseline, mid- and end-RT images.

The collected patient blood samples were analyzed for concentration of protein biomarkers using enzyme-linked immunosorbent assay (ELISA). The following biomarkers were chosen due to their proposed roles in RILD progression [21]:

interleukin(IL)-6, IL-8, angiotensin converting enzyme (ACE), alpha-2-macroglobulin (α2M), and transforming growth factor (TGF)-β1 and plasma concentration of osteopontin (OPN). The former 5 biomarkers were measured from serum while plasma was used for OPN.

5.3.3 Image pre-processing and texture analysis

The baseline, mid- and end-RT images were pre-processed to produce 2-dimensional regions of interests (ROIs) in lung tissues at anatomically matching locations in each image. The selected ROIs were analyzed for the changes in textures with reference to the corresponding ROI in the baseline. The study closely followed the methodology by Cunliffe et al. [19]. To summarize the described procedures:

1. Image registration and adaptive contouring of lung regions: Mid- and end-RT images were registered to the planning CT using the Plastimatch software with the deformable fast symmetric forces demons algorithm [22]. Using the resulting registration, Contours for lung volumes minus planning target volume (PTV), drawn in the baseline image for treatment planning, was deformed to adaptively contour the lungs in the mid- and end-RT images.
2. Determination of ROI locations: Candidate locations for the ROIs in the baseline image were set as a grid of points with a uniform grid spacing of 32x32 pixels. Forty 32x32 pixel ROIs in the baseline image were formed at the subset of candidate points that satisfy the following conditions: i) at least 85% of the ROI pixels are located within contoured lung for every time points, ii) Distribution of the locations in the baseline CT are uniform in all 3 directions, and iii) There is no overlap between the ROIs at the transformed locations in the

Table 5.1: Patient cohort characteristics. *: heterogeneity-corrected, PTV excluded

	Patient count (%)
Cohort size	22
RP grade	
0	11 (50)
1	1 (5)
2	6 (27)
3	4 (18)
4 or 5	0 (0)
2 and above	10 (45)
Mean lung dose*	
median	15.0
range	4.8-20.9
V20 (%)*	
median	25.3
range	9.5-35
Smoking status	
current	16 (73)
previous	6 (27)

mid- and end-RT images. The deformation vector field from step 1 was used to determine the matching locations of the 40 ROIs in the mid- and end-RT images.

3. Texture analysis: Twenty CT texture features with reported stability in the absence of pathologic change [23], were calculated at the ROI regions. The texture features consisted of 8 first order, 5 grey level co-ocurrence matrix, 4 Laws' filter features, and 3 fractal features (table 5.2). Mathematical definitions and details are shown in the appendix The pixels that fell outside the lung contours were not included for calculation of the textures.

Table 5.2: Twenty candidate RILD features introduced in [23]. See appendix for more explanation.

Name	Abbreviation	Note
First order features		
Mean	HU_mean	
Median	HU_median	Statistics from the
70% quantile	HU_Q70	histogram of
30% quantile	HU_Q30	Hounsfield unit
5% quantile	HU_Q5	distribution in the
Minimum	HU_min	ROIs.
Binned entropy	entropy_binned	
Unbinned entropy	entropy_unbinned	
Fractal features		
Brownian motion	frac_brownian	Self-similarity of image patterns and its scale dependence.
Box counting	frac_coarse	
Fine box counting	frac_fine	
GLCM features		
Sum average	GLCM_SM	
Sum of squares var.	GLCM_Var	Spatial dependence of co-ocurrence of gray levels [2]
Sum entropy	GLCM_SE	
Difference entropy	GLCM_DE	
Entropy	GLCM_E	
Laws' filter features		Output of 4 filters
R5L5 entropy	R5L5	that detects 4 distinct
S5L5 entropy	S5L5	patterns: level (L),
E5L5 entropy	E5L5	edge (E), spot (S),
W5L5 entropy	W5L5	ripple (W) [24]

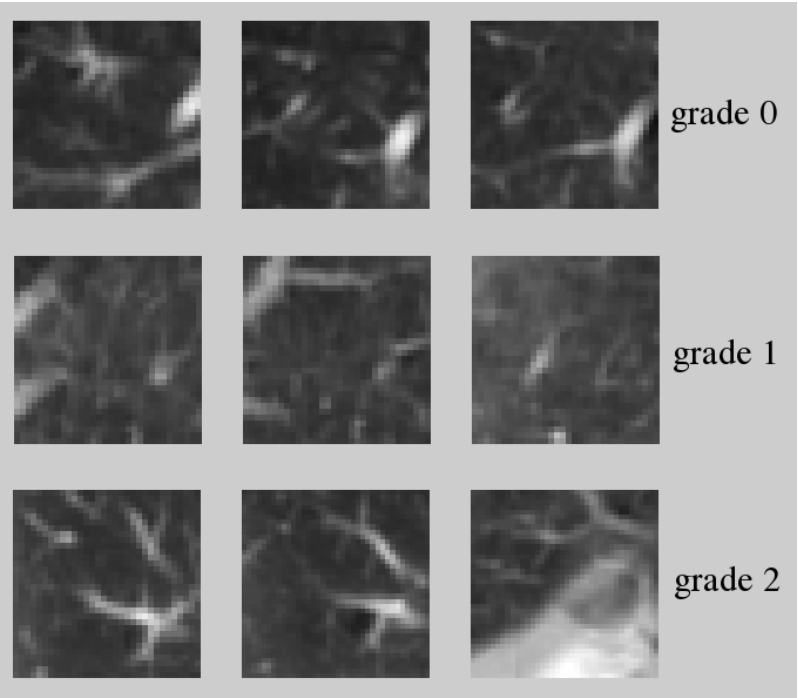

Figure 5.1: Longitudinal radiographic changes in 3 regions of interests (rows) and corresponding radiologist's grading at end-RT. Column 1: baseline, column 2: mid-RT, column 3: end-RT.

5.3.4 Damage scoring and correction of locations for ROIs

The ROI sets were reviewed by an experienced radiologist who compared pairs of ROIs between baseline and mid-/end-RT images and: i) corrected for any misalignment of ROI locations in the mid-/end-RT images, and ii) graded the severity of a radiographic change from baseline into 4 levels (0: no change, 1: mild change, 2: moderate change, 3: severe change) (figure 5.1).

5.3.5 Statistical analysis

Importance of the 20 CT features were measured in terms of the correlation of the feature signals to the radiologist's grading results. The feature signal was defined

as difference from the baseline as a percentage of the baseline level. The area under the receiver operating curve (AUC) was used for measuring how well the feature scores can separate the ROIs that were graded as no change and those that received a grade 1 or higher. The number of features were reduced to a smaller subset that showed consistently high AUC with respect to the radiologist's grades.

Each of the selected features was further modelled for their dose response using linear mixed model. For an ROI from a patient i at location $\mathbf{x}$ and a given time point t (mid- or end-RT), change in a feature T from baseline ($\Delta T(t, \mathbf{x})$) was modelled as a linear function of average dose with a slope $\alpha_{i,t}$ and an intercept $I_{i,t}$:

$$\Delta T(\mathbf{x}, i, t) \sim \alpha_{i,t} D(\mathbf{x}) + I_{i,t} \tag{5.1}$$

The mixed model uses a *fixed* component to describe patient average relationship, and adds a *random* component to take into account patient-specific variability in the relationship. In order to test patient heterogeneity in dose response, we compared three following nested models (difference between the three approaches is illustrated in figure 5.2):

- Random intercept, fixed dose effect: The intercept was allowed to vary between patients, but the slope was assumed to the same for every patients. Thus, $\alpha_{i,t}$ in equation 5.1 was reduced to α_t.
- Random intercept, random dose effect: Both the intercept and slope were allowed to vary between patients.
- Random intercept, random dose effect with an interaction term between them.

A choice between the models were made using the likelihood ratio test [25].

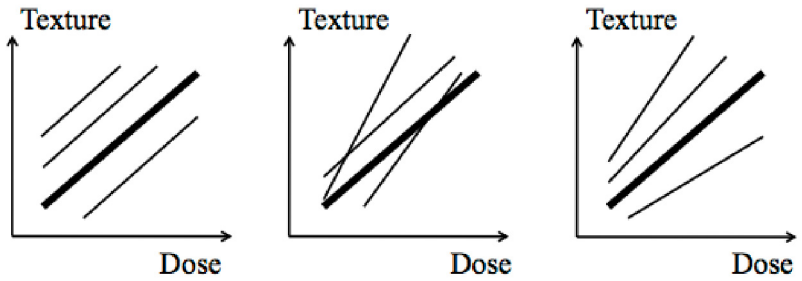

Figure 5.2: Illustration of varying relationships assumed by 3 types of nested mixed models. Patient average relationship is shown in bold and patient-specific relationships in three thinner lines. Left: Random intercept & fixed slope - every patient is assumed to have a same slope with varying intercepts. Middle: Random intercept & random slope - each patient is allowed to have an individual slope and an intercept. Right: Random intercept & random slope with an interaction term - Distribution of patient-specific slopes is correlated with intercepts (in this case, patients with a lower intercept has a lower slope as well).

5.4 Results

5.4.1 Agreement of the CT features to visual grading

At mid-RT, 2% (18/879) of the ROIs were classified by the radiologist as grade 1 injury, while no ROI was graded beyond 1. At end-RT, 6% (53/879) of the ROIs received a grade 1 or higher, while 0.8% (7/879) were grade 2. No grade 3 was observed at both time points.

Signals from the 20 features were used for classifying the ROI with an injury grade 1 or higher. AUC of the classification showed variability between features, while it generally improved at end-RT. Figure 5.3 shows that the features can be clustered for their patterns of AUC values at two time points: The first cluster occupying the top half of the heatmap, consists mainly of first-order and GLCM features. This group of features showed higher AUC at both time points than the

cluster at the bottom which are mostly fractal and Laws' filter features. Among all, 30% quantile (HU_Q30) and median (HU_median) of the HU histogram recorded the highest AUC at both time points. These two features were highly correlated ($r^2 = 0.93$); In order to reduce redundancy, HU_median was dropped in favour of HU_Q30 on the ground of lower r^2 value with the radiologist's score (0.30 vs. 0.31 for HU_Q30).

5.4.2 Dose response modelling of the features

Dose response of the feature HU_Q30 at mid treatment was best described by a fixed slope and random intercept. The identified fixed slope, 0.21 HU/Gy (95% confidence interval: 0.07, 0.32) indicated a significant ($p = 0.001$) dose effect from the population average response (table 5.3). The random intercepts varied from -24 to 26 HU (figure 5.4). The patient-specific intercept did not correlate with RP (Pearson's test r = 0.15, p = 0.51).

At end-treatment, the best dose model for HU_Q30 was a random intercept/ random slope model without interaction. The patient average (fixed) slope was 0.25 HU/ Gy (95% confidence interval: -0.05, 0.55) which was not significant (p=0.11) (table 5.3). However, unlike mid-treatment, the patients showed significant ($p < 0.001$) heterogeneity in the slope of the response. The random slope showed negative but non-significant (r= -0.42, p = 0.05) relationship with RP, while the intercept was positively correlated with RP (r = 0.40, p=0.06).

The fitted dose response functions for both time points, along with raw data points, can be shown in figure 5.6.

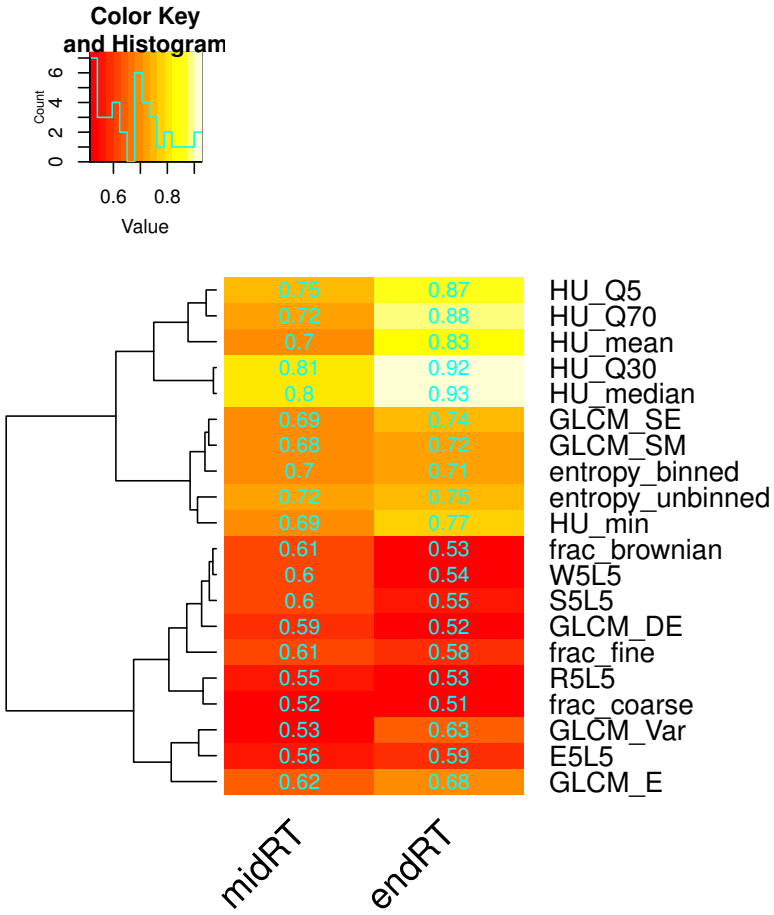

Figure 5.3: Heatmap of AUC for the 20 CT features, shown in rows, for classifying ROIs of injury grade 1 or higher assigned by the radiologist. Results of hierarchical clustering of the 20 features are shown on the left and side as a dendogram.

Table 5.3: Patient average (fixed component) dose response for HU_Q30 at mid- and end-RT. Numbers in parentheses: 95% confidence invervals.

	Time points	
	mid-RT	end-RT
Slope (HU/ Gy)	0.21 (0.07,0.32)	0.25 (-0.05,0.55)
Intercept (HU)	-0.33 (-7.25,6.59)	10.13 (-3.71,24.0)

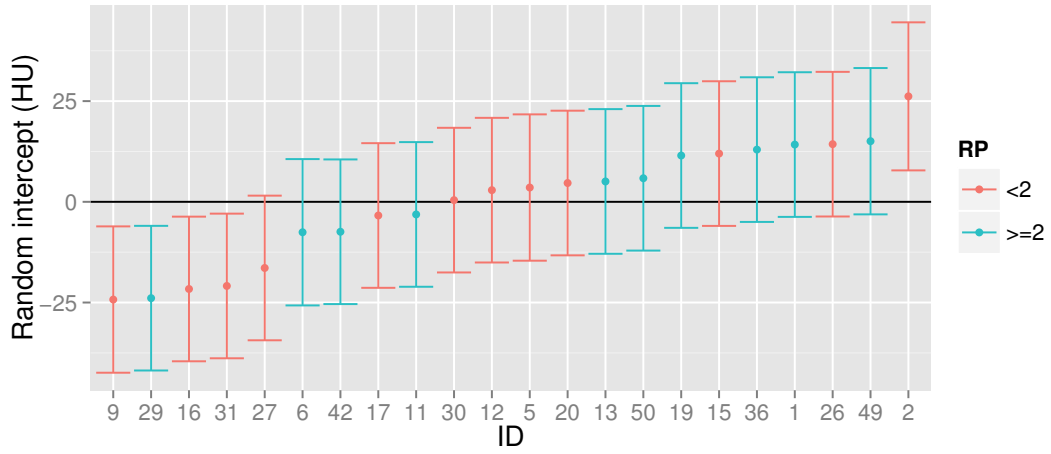

Figure 5.4: Random intercepts of dose response of the CT feature HU_Q30 at mid-treatment. Error bars: 95% confidence intervals.

Table 5.4: Pearson's r between biomarkers and the patient-specific parameters of the mixed models for HU_Q30. P-values in parentheses.

	Intercept (midRT)	Intercept (endRT)	Slope (endRT)
OPN	-0.26 (0.25)	0.09 (0.68)	0.01 (0.95)
IL8	0.12 (0.60)	0.03 (0.89)	0.20 (0.38)
ACE	-0.11 (0.64)	0.30 (0.18)	0.24 (0.28)
IL6	-0.36 (0.10)	0.52 (0.01)	-0.09 (0.68)
a2M	0.06 (0.80)	-0.09 (0.69)	-0.03 (0.88)
TGFb	0.20 (0.38)	-0.09 (0.69)	0.11 (0.64)

5.4.3 Correlation of dose response model parameters with biomarkers

The three parameters that characterize patient specific dose response - random slope at mid-RT, random slope and intercept at end-RT - was further investigated further for their biological relevance. No relationship stood significant after Bonferroni correction on the significance level ($\alpha = 2.7 * 10^{-3}$). However, table 5.4 demonstrated notable correlation between IL6 and the random intercept at mid- and end-RT.

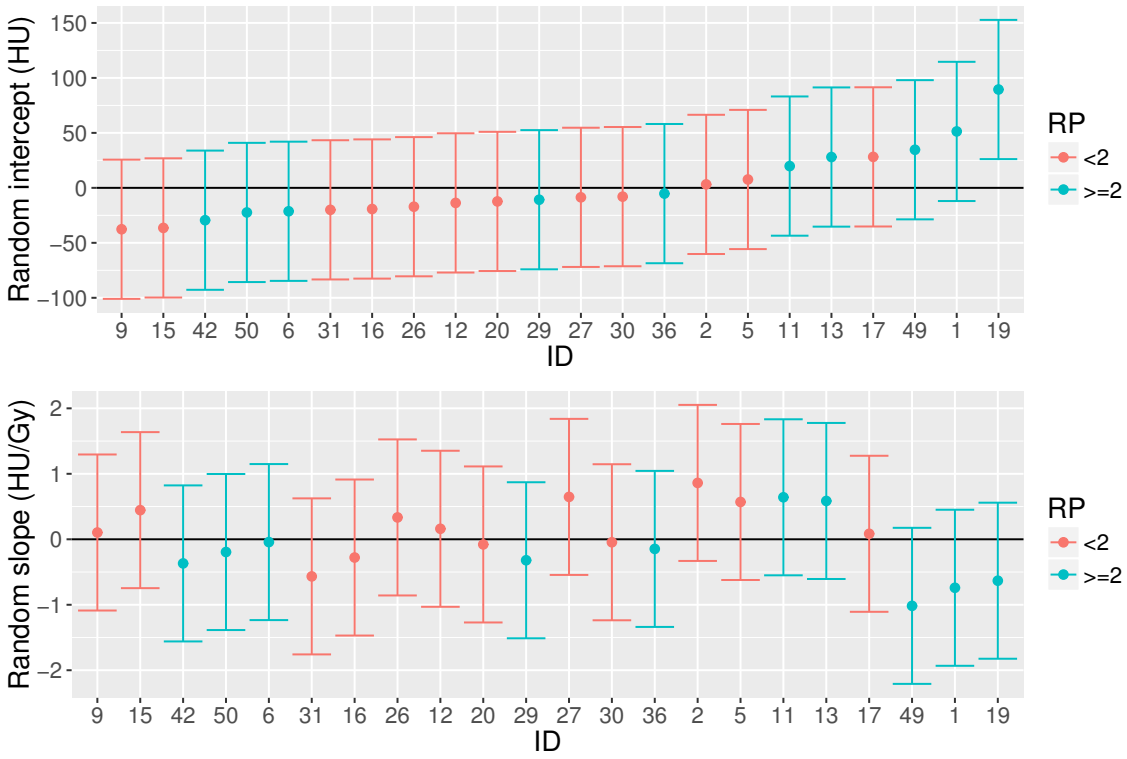

Figure 5.5: Random intercepts and slopes of dose response of HU_Q30 at end-treatment. Error bars: 95% confidence intervals.

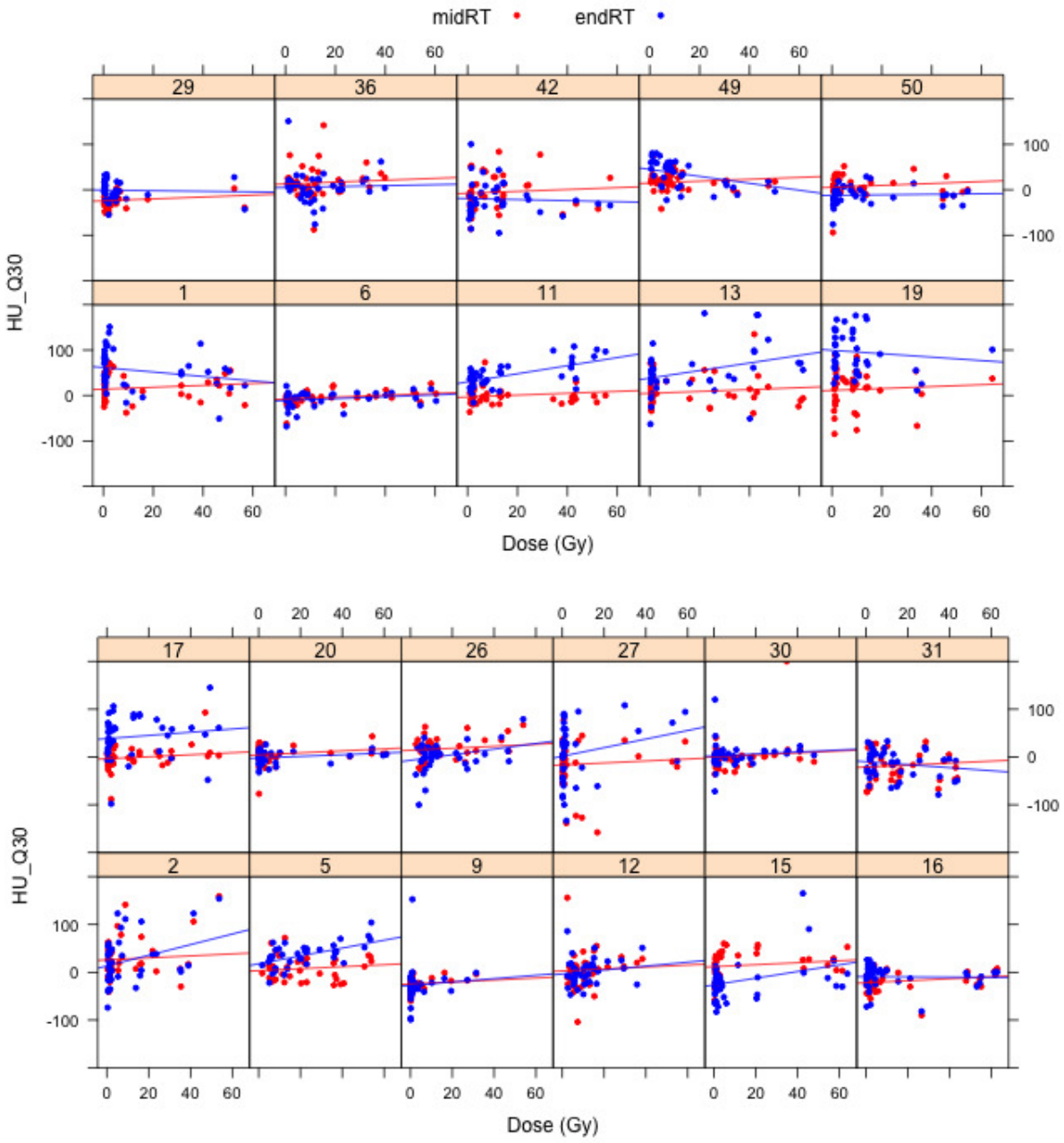

Figure 5.6: Intra-treatment changes in HU_Q30 from baseline for all ROIs as a function of dose. The fitted mixed effect models (red and blue lines) are superposed. Panels for each patients were divided into grade ≥ 2 symptomatic (top) and asymptomatic (bottom) RP groups.

5.5 Discussion

We investigated textural changes in CT images before completion of radiotherapy to search for their value as biological surrogates or predictors for clinical outcomes. Using linear mixed models, we specifically studied dose or patient dependence on such changes and their longitudinal patterns. Dose dependence on radiographic changes after radiotherapy is strongly established by many studies [26] [27]. The remaining question is in explaining patient heterogeneity in such radiographic response [28] [29]. If such heterogeneity can be captured before RT completion, it could open an avenue towards biological adaptive RP using CT biomarkers.

Previously, Bertelsen et al. [20] detected CT density increase during the course of radiotherapy from the analysis of daily CBCT images. They also reported development of dose response towards the end of radiotherapy. Our study refines their methods in many aspects: First, all the images were obtained helically and suffered less image degradation due to scatter than cone beam acquisition. Secondly, raw mid- and end-RT images were used for texture analysis without deformation which might affect robustness of the textures [17]. Lastly, noise in textures due to registration error was reduced due to an extra correction of ROI location by an experienced radiologist.

At mid-treatment, we observed significant dose dependency in the CT change represented by 30% quantile of the Hounsfield unit histogram. However, the slopes of the dose response did yet to vary significantly between patients. Patient heterogeneity was more observable at the end of radiotherapy where we found significant random effects on both a slope and an intercept. This is thought to be a result of

dose accumulation and resulting increase in CT signals as confirmed by increased fraction of the ROIs classified as abnormal from 2% at mid-RT to 6% at end-RT. However, we do not rule out other pathological changes that might influence the CT analysis results. For example, 4 out of the 10 patients with RP (study ID 1, 6, 11, and 19) were marked as a possible case of pneumonia by a referring physician. This might be responsible for the observed increase in intercept for some of these patients (figure 5.5), which indicates the CT change that is not caused by dose deposition. However, differential diagnosis was not performed objectively for every patients, and thus this hypothesis could not be tested in this study.

We performed rudimentary variable selection on the 20 candidate CT features based on correlation with the radiologist's assessment. This choice could be biased by assessment of only one person. In the future, inter-observer variability in visually grading ROIs needs to be taken into account. Nevertheless, the pattern we found in the AUC across the 20 features (figure 5.3) agrees with the findings by Cunliffe et al [18]. They also reported higher AUC from first order and GLCM features when the lowest injury grade was used as a cutoff. Thus, it demonstrates that those two types of textures could be used as a surrogate for the radiologists' grading scheme.

We failed to identify predictive value or biologically relevance of the dose-response model parameters. This study is in progress and awaits addition of more patient data to confirm the discussed findings. The study methods, especially in the selection of CT features, need to be further refined. The results are limited to one of the first order features, HU_Q30. Such histogram based features discard spatial

distribution of grey levels and thus limits the quantitative potential of texture analysis. Nevertheless, the number of features in models have to be tightly regulated in awareness of false positive probability due to multiple comparison. Future direction lies in investigation of different feature selection techniques on this data, which was previously explored by Parmar et al. for their lung cancer study [1]. Another challenge remains in including into this analysis framework post-treatment scans taken at a different scanner and protocols some of which includes contrast agents. Sensitivity of textures to such changes, as already addressed by [30] [31], needs to be address before extending this modelling framework to the post-treatment scans.

5.6 Conclusion

We analyzed texture changes in CT images during RT delivery by comparing the matching anatomical regions with a baseline image. In our cohort, we found that 30% quantile of Hounsfield unit histogram agreed most with visual assessment by a radiologist. This feature showed significant dose dependency at mid-treatment. Patient heterogeneity on the dose response relationship, in terms of slope and intercept, was stronger at end-treatment. The presented method for patient-specific modelling of dose response of intra-treatment CT changes could identify potential biomarkers for biological plan adaptation.

REFERENCES

[1] C. Parmar, R. T. Leijenaar, P. Grossmann, et al., Radiomic feature clusters and prognostic signatures specific for lung and head & neck cancer, Sci Rep 5 (2015) 11044.

[2] R. M. Haralick, K. Shanmugam, I. Dinstein, Textural features for image classification, IEEE Transactions on Systems, Man, and Cybernetics SMC-3 (6) (1973) 610–621.

[3] P. Rubin, L. Constine, L. Marks, ALERT: Adverse Late Effects of Cancer Treatment: Volume 2: Normal Tissue Specific Sites and Systems, Medical Radiology, Springer, 2013.

[4] L. B. Marks, S. M. Bentzen, J. O. Deasy, et al., Radiation dose-volume effects in the lung, International Journal of Radiation Oncology*Biology*Physics 76 (3, Supplement 1) (2010) S70 – S76.

[5] C. Dehing-Oberije, D. D. Ruysscher, A. van Baardwijk, S. Yu, B. Rao, P. Lambin, The importance of patient characteristics for the prediction of radiation-induced lung toxicity, Radiotherapy and Oncology 91 (3) (2009) 421 – 426.

[6] F.-M. Kong, S. Wang, Nondosimetric risk factors for radiation-induced lung toxicity, Seminars in Radiation Oncology 25 (2) (2015) 100 – 109.

[7] M. S. Anscher, L. B. Marks, T. D. Shafman, et al., Using plasma transforming growth factor beta-1 during radiotherapy to select patients for dose escalation,

Journal of Clinical Oncology 19 (17) (2001) 3758–3765.

[8] L. Zhao, K. Sheldon, M. Chen, et al., The predictive role of plasma tgf-β during radiation therapy for radiation-induced lung toxicity deserves further study in patients with non-small cell lung cancer, Lung Cancer 59 (2) (2008) 232 – 239.

[9] Y. Chen, P. Rubin, J. Williams, E. Hernady, T. Smudzin, P. Okunieff, Circulating IL-6 as a predictor of radiation pneumonitis, International Journal of Radiation Oncology*Biology*Physics 49 (3) (2001) 641 – 648.

[10] J. H. Oh, J. M. Craft, R. Townsend, J. O. Deasy, J. D. Bradley, I. El Naqa, A bioinformatics approach for biomarker identification in radiation-induced lung inflammation from limited proteomics data, Journal of Proteome Research 10 (3) (2011) 1406–1415.

[11] L. Zhao, L. Wang, W. Ji, et al., Association between plasma angiotensin-converting enzyme level and radiation pneumonitis, Cytokine 37 (1) (2007) 71 – 75.

[12] S. N. Ghosh, R. Zhang, B. L. Fish, et al., Renin-angiotensin system suppression mitigates experimental radiation pneumonitis, International Journal of Radiation Oncology*Biology*Physics 75 (5) (2009) 1528–1536.

[13] Z. Kocak, E. S. Evans, S.-M. Zhou, et al., Challenges in defining radiation pneumonitis in patients with lung cancer, International Journal of Radiation Oncology*Biology*Physics 62 (3) (2005) 635 – 638.

[14] S. L. Tucker, H. Jin, X. Wei, et al., Impact of toxicity grade and scoring system on the relationship between mean lung dose and risk of radiation pneumonitis in a large cohort of patients with non-small cell lung cancer, International Journal

of Radiation Oncology*Biology*Physics 77 (3) (2010) 691 – 698.

[15] B. Ganeshan, S. Abaleke, R. C. Young, C. R. Chatwin, K. A. Miles, Texture analysis of non-small cell lung cancer on unenhanced computed tomography: initial evidence for a relationship with tumour glucose metabolism and stage, Cancer Imaging 10 (2010) 137–43.

[16] H. J. Aerts, E. R. Velazquez, R. T. Leijenaar, et al., Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach, Nat Commun 5 (2014) 4006.

[17] A. R. Cunliffe, H. A. Al-Hallaq, Z. E. Labby, et al., Lung texture in serial thoracic ct scans: Assessment of change introduced by image registration, Medical Physics 39 (8) (2012) 4679–4690.

[18] A. R. Cunliffe, r. Armato, S. G., C. Straus, R. Malik, H. A. Al-Hallaq, Lung texture in serial thoracic ct scans: correlation with radiologist-defined severity of acute changes following radiation therapy, Phys Med Biol 59 (18) (2014) 5387–98.

[19] A. Cunliffe, r. Armato, S. G., R. Castillo, N. Pham, T. Guerrero, H. A. Al-Hallaq, Lung texture in serial thoracic computed tomography scans: correlation of radiomics-based features with radiation therapy dose and radiation pneumonitis development, Int J Radiat Oncol Biol Phys 91 (5) (2015) 1048–56.

[20] A. Bertelsen, T. Schytte, S. M. Bentzen, O. Hansen, M. Nielsen, C. Brink, Radiation dose response of normal lung assessed by cone beam ct - a potential tool for biologically adaptive radiation therapy, Radiother Oncol 100 (3) (2011) 351–5.

[21] K. Fleckenstein, B. Gauter-Fleckenstein, I. L. Jackson, Z. Rabbani, M. Anscher, Z. Vujaskovic, Using biological markers to predict risk of radiation injury, Seminars in Radiation Oncology 17 (2) (2007) 89 – 98.

[22] G. C. Sharp, N. Kandasamy, H. Singh, M. Folkert, Gpu-based streaming architectures for fast cone-beam ct image reconstruction and demons deformable registration, Phys Med Biol 52 (19) (2007) 5771–83.

[23] A. R. Cunliffe, r. Armato, S. G., X. M. Fei, R. E. Tuohy, H. A. Al-Hallaq, Lung texture in serial thoracic ct scans: registration-based methods to compare anatomically matched regions, Med Phys 40 (6) (2013) 061906.

[24] K. Laws, Textured image segmentation, Report, University of Southern California (1980).

[25] B. M. Bolker, M. E. Brooks, C. J. Clark, et al., Generalized linear mixed models: a practical guide for ecology and evolution, Trends in Ecology & Evolution 24 (3) (2009) 127 – 135.

[26] J. Ma, J. Zhang, S. Zhou, et al., Regional lung density changes after radiation therapy for tumors in and around thorax, International Journal of Radiation Oncology*Biology*Physics 76 (1) (2010) 116 – 122.

[27] D. A. Palma, J. van Srnsen de Koste, W. F. Verbakel, A. Vincent, S. Senan, Lung density changes after stereotactic radiotherapy: A quantitative analysis in 50 patients, International Journal of Radiation Oncology*Biology*Physics 81 (4) (2011) 974 – 978.

[28] D. De Ruysscher, H. Sharifi, G. Defraene, et al., Quantification of radiation-induced lung damage with ct scans: the possible benefit for radiogenomics,

Acta Oncol 52 (7) (2013) 1405–10.

[29] G. Defraene, W. van Elmpt, W. Crijns, P. Slagmolen, D. De Ruysscher, Ct characteristics allow identification of patient-specific susceptibility for radiation-induced lung damage, Radiother Oncol 117 (1) (2015) 29–35.

[30] V. Kumar, Y. Gu, S. Basu, et al., Radiomics: the process and the challenges, Magn Reson Imaging 30 (9) (2012) 1234–48.

[31] D. Mackin, X. Fave, L. Zhang, et al., Measuring computed tomography scanner variability of radiomics features, Invest Radiol 50 (11) (2015) 757–65.

CHAPTER 6
Building the RILD model: computational methodology

Contents

6.1 Introduction

Radiation oncologic data is highly complex and heterogeneous; Patients are treated with different radiotherapy plans tailored to their anatomy. The complexity inflates when patient-specific clinical and biological information are added. When building a model to predict radiotherapy outcomes, we are often left without enough prior knowledge that allows us to design the specifics of the model. For example, we do not know how many variables to include in our model. Thus, a classical way of fitting an analytical function with a pre-determined form is no longer relevant in this modern era of "big data" [32]. We have already seen this from the application of the LKB model in chapter 4 where the number of model parameters grew rapidly as more factors (e.g. time progression, tumour position) are included. *Machine learning* is an emerging discipline that focuses on finding useful patterns from complex data. Instead of fitting a function with a pre-determined form to data, machine learning techniques try to establish which form of function can describe the observed patterns in data. Another important aspect of machine learning is to ensure that the learned pattern is also applicable to the world outside the data upon which the pattern is learned.

This chapter will provide brief introduction to machine learning and specific techniques that will be used in the later chapters. Then, it will describe a computational pipeline that was devised for building and validating a RILD model from dosimetric, clinical and biological data. Some materials in this chapter were published in: supplementary material for Lee et al., Bayesian network ensemble as a multivariate strategy to predict radiation pneumonitis risk. *Medical Physics*, 42(5):2421-2430

and: Lee, S. and El Naqa, I. (2015). Machine learning methodology. In El Naqa, I., Li, R., and Murphy, M. J., editors, *Machine Learning in Radiation Oncology*, pages 21-39. Springer International Publishing.

6.2 Brief introduction to machine learning

6.2.1 Basic terminology

Each example, or a patient in our case, can be characterized by a vector $\mathbf{x}$ that consists of p variables $(x_1, x_2, ..., x_p)$. Each variable in the vector, x_i, is called a *feature*[1] , and a vector $\mathbf{x}$ is called *a feature vector*. Data as a collection of n examples can be written as a matrix $\mathbf{X}$ where examples are arranged in rows and features are organized column-wise.

$$\mathbf{X} = (\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n})^T \tag{6.1}$$

Machine learning techniques can be broadly classified into the following two:

- Supervised learning: Data X is given with a *target vector* $\mathbf{y}$ where each element y_i is matched with a feature vector $\mathbf{x_i}$. A learning problem is to train a function f that relates the features to the target $\mathbf{y} = f(\mathbf{X})$ using *training data* $\{\mathbf{X^t}, \mathbf{y^t}\}$. Once the training is done, the function f can be applied to a new set of data called a *test set*. Intuitively, it can be seen as learning with a "teacher" that provides the answers (target vectors) to a problem.

[1] In this text, the terms "feature" and "variable" will be used interchangeably.

- Unsupervised learning: Training data is given without a target $\mathbf{y}$. Instead, the function f unveils hidden structures in data. An example of this learning is clustering or estimation of probability distribution.

Supervised learning is again classified into two types of learning:

- Classification: A target vector $\mathbf{y}$ takes discrete values or *labels*. Radiotherapy outcome modelling with endpoints in a discrete (e.g. toxicity grade 1-5) or binary format (e.g.toxicity grade 2 or higher) falls into this category.
- Regression: A target vector is in a continuous value. One example in radiotherapy would be xerostomia (dry mouth due to damage to salivary gland from radiotherapy) which can be measured by saliva flow.

6.2.2 Bias-variance tradeoff

Bias and variance are two important components of the performance of supervised machine learning models. Mathematical expressions for bias and variance are derived by decomposing a *generalization error* which is defined as the deviation of trained function h from a true target value y (equation 6.2). Note that the expectation value is taken to average the error over the distribution of testing examples $\{\mathbf{x}^*, y^*\}$

$$e = E[(y^* - h(x^*))^2] \tag{6.2}$$

Then, we assume that there is a true function $f(x)$ that describes y^* with a noise σ such that $y^* = f(x^*) + \epsilon$. The error can be decomposed into the following components (see [1] for derivation):

$$
\begin{aligned}
E[(y^* - h(x^*))^2] &= E[(h(x^*) - \overline{h(x^*)})^2] \\
&+ (\overline{h(x^*)} - f(x*))^2 \\
&+ E[(y^* - f(x^*))^2]
\end{aligned}
$$

The first component, $E[(h(x^*) - \overline{h(x^*)})^2]$, is called *variance*; It measures the deviation of the trained functions around its mean $\overline{h(x^*)}$ resulting from different training examples. In other words, it measures sensitivity of the learned function to changes in the training data. The second term $(\overline{h(x^*)}^2 - f(x*))^2$, *bias*, indicates how well the trained function $h(x)$ fits the true function $f(x)$ on average. The last component $E[(y^* - f(x^*))^2]$ represents the deviation from the true function due to noise inherent in observation of y.

The bias and variance are both closely related to model complexity. In general, the bias can be reduced by increasing model complexity (e.g. adding more parameters). However, this typically leads to increasing variance or so called *overfitting* where the model fits well in the training set but performs worse in unseen examples. This relationship, called *bias-variance tradeoff* is very important consideration when designing a model. Efforts should made during training to fit the data reasonably well but at a right level of complexity where bias and variance are in a good balance (figure 6.1). An approach called *model selection* advocates the choice of a single model from candidate models that is expected to have the least generalization error. This approach has an old historical root, dating from the 14th century principle

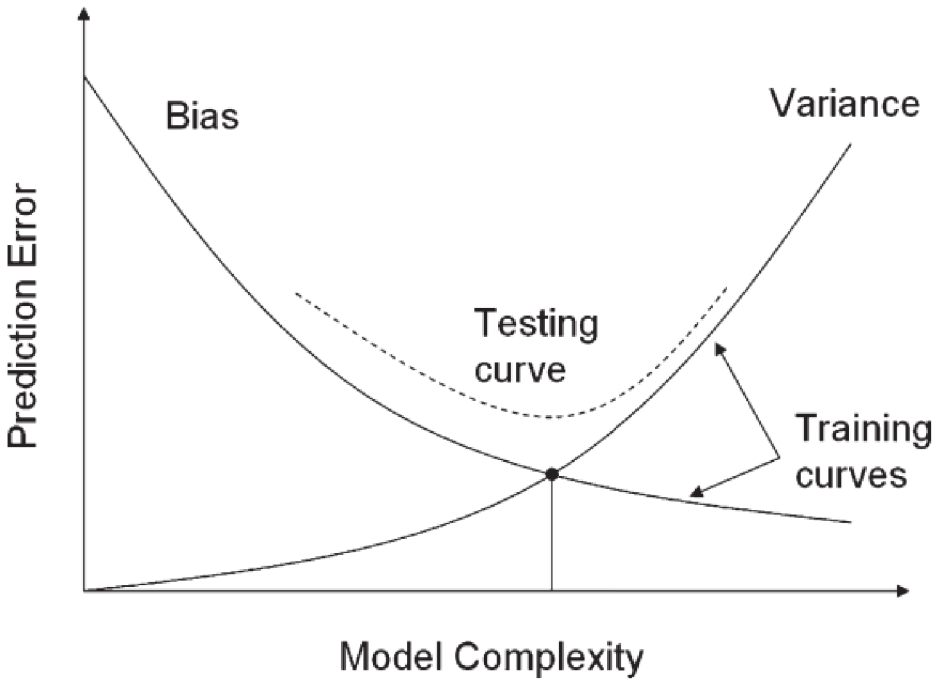

Figure 6.1: Illustration of the bias-variance tradeoff in machine learning. As model complexity increases, With permission from Deasy and El Naqa [2]

"Occam's razor": It states that amongst several hypotheses, the one that makes the least assumptions is the most plausible.

There are two ways of finding the bias-variance balance. The first way is to control model complexity in a form of a *regularization parameter* which imposes a penalty for higher model complexity. Then, we repeat the model training with different values of the regularization parameters, and evaluates the generalization error of the resulting functions in the subset of the training set reserved for tuning the parameter values i.e. not used during the training process (cross validation: section 6.2.3). An alternative way, based on Bayesian statistics, controls the model complexity in a more implicit way, which will be introduced in section 6.3

6.2.3 Model validation

Validation of a classifier on the examples unseen during a training phase is important for evaluating its generalizability. A testing set can be collected independently from a training set and reserved solely for a validation purpose. In the absence of such disjoint data, a single dataset needs to be used both for training and testing a classifier. Various resampling techniques, such as K-fold cross-validation and bootstrapping, are in practice to divide data into training and testing sets. In a K-fold cross-validation setting, data is partitioned into K folds of equal size and a classifier is trained using K-1 folds and validated on the 1 remaining fold. This step is repeated K times so that every fold (and thus every example) is used only once for validation (figure 6.2).

Bootstrapping, devised by Efron and Tibshirani [4], randomly samples from the original dataset with replacement in order to generate a"new" dataset of the same size. Allowing replacement results in the samples that are not chosen into the bootstrap replicate (out-of-bag(OOB) samples). One way of evaluating a classifier using bootstrapping is to train it on a bootstrap replicate and test it on the OOB samples. However, this leads to overly pessimistic result due to the fact that OOB samples does not fairly represent a random sample drawn from the true population [5]. The 0.632+ bootstrap [6], one of the variants of bootstrap methods, takes as a bootstrap performance ($\hat{f}$) a weighted average of the following two components: fitting performance on the entire dataset $f(\mathbf{x}, \mathbf{x})$ and the performance of the model trained in a bootstrap replicate ($\mathbf{x}^{*\mathbf{b}}$) and tested on the OOB samples ($f(\mathbf{x}^{\mathbf{b}^*}, \mathbf{x}^{\mathbf{b}^*}(0))$).

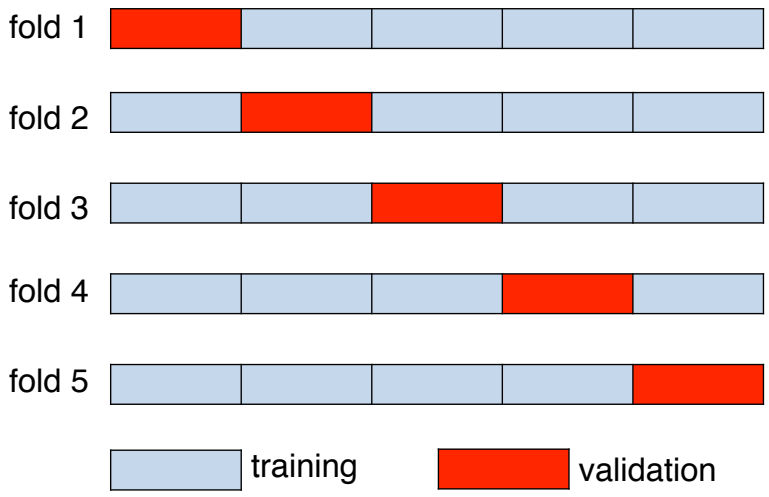

Figure 6.2: Illustration of 5-fold cross validation. Data is divided into 5 disjoint sets, and at each fold one of the sets is reserved for validation while the rest is used for training.

$$\hat{f} = \frac{1}{B}\sum_{i=1}^{B}[\alpha(b)f(\mathbf{x}^{\mathbf{b}^*}, \mathbf{x}^{\mathbf{b}^*}(0)) + (1-\alpha(b))f(\mathbf{x},\mathbf{x})] \tag{6.3}$$

Where $\alpha(b)$ is a weighting factor defined as a following:

$$\alpha(b) = \frac{0.632}{1 - 0.368 \cdot R(b)} \tag{6.4}$$

$R(b)$, a relative overfitting rate, is defined as the ratio of difference between

$$R(b) = \frac{f(\mathbf{x}^{\mathbf{b}^*}, \mathbf{x}^{\mathbf{b}^*}(0)) - f(\mathbf{x},\mathbf{x})}{f(\mathbf{x}^{\mathbf{b}^*}, \mathbf{x}^{\mathbf{b}^*}(0)) - \gamma} \tag{6.5}$$

Where γ denotes a no-information error rate that applies when a true class label is independent of predictions by a model. For example, when f = AUC, $\gamma = 0.5$. Introducing R(b) effectively gives more weight to the OOB performance for the bootstrap replicates that a model overfits.

At the expense of of heavier computation, Bootstrapping has several advantages over cross validation as a classifier testing method. First, it can reduce variability in prediction error estimates that is prominent in cross-validation based methods [7]. This property, especially the one of 0.632+ bootstrap, was demonstrated in small datasets [8]. Moreover, it allows us to assess variability not only in prediction performances but also the parameters that constitute the model.

6.3 Bayesian approach to machine learning

Bayesian statistics emphasizes subjectivity in interpreting probability of an event. According to Bayesian philosophy, our understanding on an event a lies in the combination of two components: what we already understand about a, and gathered

observation D on a. This principle is epitomized by the following equation known as the *Bayes' law* (equation 6.6):

$$\begin{aligned} p(a|D) &= \frac{p(D|a)p(a)}{P(D)} \\ &\propto p(D|a)p(a) \end{aligned}$$

The first component of the right hand side, $p(D|a)$ is called *likelihood*. The second component, $p(a)$ is *prior probability*, or simply *prior*, and represents our belief on the event a *before* we observe the data D. Product of the two is proportional to $p(a|D)$, called *posterior probability*, with a normalization factor $P(D)$ that is constant with respect to a.

When applied to machine learning, the Bayesian approach finds the model m^* that maximizes the posterior $p(m|D)$, called *maximum a-posteriori*(MAP), given data D and a prior distribution on parameters $p(m)$. Choice of the prior can be subjective and beyond the scope of this text (see Kass and Wasserman [9] for more details). In the absence of strong prior beliefs, it is common to use the non-informative prior which gives equal weights to every models (uniform $p(\theta)$) [10]. In that case, the model selection is about maximizing the likelihood obtained by marginalizing over all the possible parameter values that specify the model:

$$p(m|D) \propto p(D|m) = \int p(D|\theta)p(\theta|m)d\theta \tag{6.6}$$

This quantity, $p(D|m)$, is called *marginal likelihood* or *evidence*, and serves as an important criterion for Bayesian model selection. A choice between a model m_1

and m_2 is made based on the *Bayes factor* (K) which is defined as a ratio of marginal likelihood between the two models:

$$K = \frac{P(D|m_1)}{P(D|m_2)} = \frac{p(D|m_1,\theta)p(\theta|m_1)d\theta}{p(D|m_2,\theta)p(\theta|m_2)d\theta} \tag{6.7}$$

Information criteria, another Bayesian metric for model selection, was suggested as an alternative to maximum likelihood estimation (MLE) to handle its overfittng tendency [1], Bayesian information criteria (BIC) [3] is one variant of this quantity, defined as:

$$\ln p(D) \sim \ln p(D|\theta^*) - \frac{1}{2}M \ln N \tag{6.8}$$

where N is the number of examples in the training set, M is the number of parameters in θ, and θ^* is the optimal parameter set. As seen in the second term, the log likelihood in BIC linearly decreases with the number of parameters, which regularizes against larger model order.

The Bayesian model selection is different from the cross-validation based selection where parameter values are explicitly tuned in a reserved dataset to minimize the cross validation error. Marginalizing over parameters does away with having to set aside a part of data for the tuning. Complex models with more parameters in the model generally have lower marginal likelihood for model, providing protection against overfitting [11]. The rationale behind this is illustrated in the figure 6.3. Complex models (H_2) can explain a wide variety of data, compared to the simpler one (H_1) which is focused on the limited range of data (C_1). However, marginal likelihood needs to be normalized (e.g. the area under the curve for $P(D|H)$ should

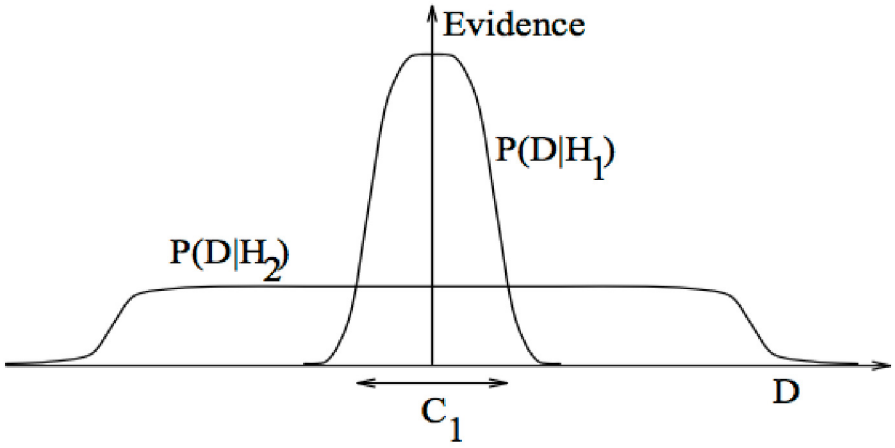

Figure 6.3: Illustration of how Bayesian model selection prefers a simpler model. Marginal likelihoods (evidences) of two models H_1 and H_2 are plotted in a space of possible data sets D. In the observed region C_1, a simpler model H_1 has a higher likelihood. With permission from [11]

be 1) and thus the curve for H_2 is spread out, which makes it a less powerful model than H_1 in terms of predicting C_1.

6.3.1 Bayesian model averaging

The Bayesian approach provides an alternative to model selection: Instead of selecting one model, we consider a batch of models with different posterior probabilities. Then, when it comes to making a decision, we obtain a composite result by averaging out the results from all the models in the batch weight by respective posterior. This inference method, referred to as *Bayesian model averaging (BMA)* [12], can be formulated in terms of n models $\{m_1, m_2, ..., m_K\}$, a quantity of interest c, and data D as the following:

$$p(c|D) = \sum_{k}^{K} p(c|M_k, D)p(M_k|D) \tag{6.9}$$

BMA can overcome the uncertainty in model selection in case there are multiple models that fit the data almost equally well. Madigan and Raftery [13] provided empirical evidences that averaging over models can improve prediction accuracy. They addressed the issue with exhaustive summation in equation 6.9 by considering only the models with sufficient parsimony and prediction accuracy. Note that BMA should be distinguished from bagging or boosting where an ensemble of models are trained from different training sets, usually obtained by bootstrapping from the original training data.

6.3.2 Monte-Carlo posterior sampling

Bayesian inference requires knowledge of a posterior $p(D|m)$. One way is to approximate the distribution as superposition of gaussians (Laplace approximation) centred around the values determined from observations. Another popular method of approximation is Markov Chain Monte Carlo (MCMC) sampling [14]. MCMC is a class of stochastic algorithms that draw samples from a multivariate probability distribution by generating a trajectory of random walks (chain) across different states in a probability space. A random walk is carried out based on Markov chain mechanism which can be stated as: probability of a current state depends only on the previous state. Mathematically, a length-i chain of states $\{x^{(1)}, x^{(2)}, ..., x^{(i)}\}$ is called a Markov chain when:

$$p(x^{(i)}|x^{(i-1)}, ..., x^2, x^1) = p(x^{(i)}|x^{(i-1)}) = T(x^{(i)}; x^{(i-1)}) \tag{6.10}$$

The function T, called a transition matrix, governs the probabilities for different possible transitions. The chain is said to be at a stationary state when a transition

probability is independent of a chain length i. When the chain converges to a stationary state, the sample distribution in a chain loses its dependence on the initial state and can be used as a fair approximation of the target probability distribution. See the work by Cowles and Carlin [15] that reviewed different convergence tests. MCMC can be well integrated into Bayesian statistics because a converged Markov chain can be approximated as an actual joint posterior distribution [15].

There is a variety of sampling techniques for generating Markov chains. Among those, the Metropolis-Hastings algorithm is one of the most elementary methods that has created many derivatives. The algorithm introduces a proposal density to draw a candidate sample, and accept or reject it as a Markov chain depending on the magnitude of the proposal density and posterior probability relative to the last sample in the chain. The algorithm can be summarized by the following steps:

1. Start at the last sample of the chain x_{old}.
2. Draw a candidate sample x_{new} from the proposal density $Q(x_{new}; x_{old})$
3. Compute an acceptance ratio A which is a product of the ratio of the posterior probability (A_1) and proposal the density(A_2):

$$A = A_1 A_2, A_1 = \frac{P(x_{new}|D)}{P(x_{old}|D)}, A_2 = \frac{Q(x_{old}; x_{new})}{Q(x_{new}; x_{old})} \tag{6.11}$$

4. x_{new} is accepted as a valid sample with probability $\min(1, A)$

The proposal density is customizable to a specific application, but its choice might affect the speed of convergence. For detailed explanation on Metropolis-Hastings as well as other sampling algorithms, consult the handbook by Brooks et al.[16]

6.4 Supervised machine learning techniques

6.4.1 Logistic regression

Multivariate logistic regression maps a real-valued input vector into a probability value in $[0, 1]$ via a logistic sigmoid function. Logistic regression belongs to a family of *generalized linear model* which relates a target value with linear combination of input variables via a link function $\sigma(a)$ (equation 6.12)

$$p(y|\mathbf{x}) = \sigma(\mathbf{w}^T\mathbf{x})) \tag{6.12}$$

Where $\mathbf{x} = (1, x_1, x_2, ..., x_n)$, a feature vector with a constant element for determining the model offset, and a vector $\mathbf{w} = (w_0, w_1, w_2, ..., w_n)$ has regression coefficients for input variables as well as an intercept (w_0).

Logistic regression is characterized by an inverse logit (logistic) link function (equation 6.13):

$$\sigma(a) = \frac{1}{1 + exp(-a)} \tag{6.13}$$

Dimensionality of the logistic model is equal to the number of variables in $\mathbf{x}$ + 1 (including the intercept), as each variable has one adjustable parameter in $\mathbf{w}$. The weights $\mathbf{w}$ can be determined by maximum likelihood fitting; Although there is no closed solution, $\mathbf{w}$ can be computed by an iterative method called the Newton's method [17]. However, maximum likelihood method generally has low-bias and large variance, which is more prone to overfitting [1]. In logistic regression fitting, the size of the weight vector $\mathbf{w}$, called a p-norm, is introduced as a regularization method. The p-norm of $\mathbf{w}$, $\|\mathbf{w}\|_p$, is defined as the following:

$$\|w\|_p = (\sum_{i=1}^{N} \|w_i\|^p)^{1/p} \tag{6.14}$$

This form of regularization implies that the model with larger magnitudes of coefficients is considered more complex and subject to higher penalty.

An objective function $O(\mathbf{w}|\mathbf{x}, y)$ is now redefined as:

$$O(\mathbf{w}|\mathbf{x}, y) = LL(\mathbf{x}, y, \mathbf{w}) + \lambda\|w\|_p$$

where a hyperparameter λ governs the strength of regularization. Regularization effectively reduces the absolute values of the coefficients $\mathbf{w}$. Preventing large values in $\mathbf{w}$ can control over-fitting by making the best-fit function "smoother", which increases the bias in effect; Without regularization, the maximum likelihood method tends to prefer a solution with large variations in an effort to fit every points in the data. The hyperparameter λ is typically optimized in a cross-validation setting: The best value λ is chosen so that the resulting model $f(\mathbf{x}|\mathbf{w}, \lambda)$ gives the best average performance in the validation sets.

Choice of the norm of regularization, p, also has a great impact on the fitted function. For example, L_2 regularization or *ridge regression* [18], prefers a solution that minimizes the squared sum of the coefficients. On the other hand, L_1 norm imposes a penalty to the sum of the absolute values of the coefficients. Solutions under L_1 regularization tend to be *sparse*; many parameters in the optimized $\mathbf{w}$ are set to zero. This property can be used as a variable selection strategy (see 6.5).

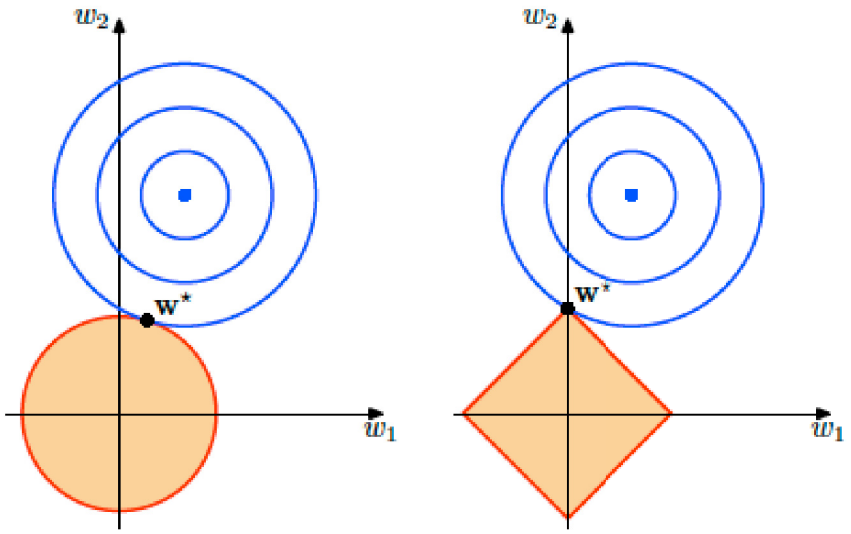

Figure 6.4: Demonstration of how L_1 and L_2 regularization determines the optimal solution in a two-dimensional parameter space (w_1,w_2). Blue and red contours represent iso-likelihood and iso-regularization surfaces, respectively. The optimum parameters $\mathbf{w}^*$ are given by the intersection between the two surfaces. Due to its shape in the parameter space, L_1 regularization is more like to produce sparse solution where $w_1 = 0$. With permission from Bishop et al. [1]

6.4.2 Bayesian Network

Bayesian Belief Network, or Bayesian Network, is designed to model probabilistic relationships amongst a set of random variables. A key feature of Bayesian Network is graphical representation of the relationships via a directed acyclic graph (DAG) which encodes the presence and direction of influence between variables. In a DAG, each variable is assigned to a node and connected to each other via an edge (vertex) which originates from a variable (parent) that influences the probability of the variable it's connected to (child). Thus, probability of a random variable is set to be conditional upon its parent variable(s). The connectivity information in a DAG derives conditional independence relationships that can be stated as: random variables X and Y are conditionally independent given another variable set $Z_1, Z_2, ..., Z_n$ if and only if:

$$P(X|Y, Z_1, Z_2, ..., Z_n) = P(X|Z_1, Z_2, ..., Z_n) \tag{6.15}$$

A set of conditional independence relationships specified in a DAG greatly simplifies computation of probability distributions by use of this convenient property: joint probability distribution between the entire variable set, $X = X_1, X_2, ..., X_n$, can be obtained by taking the product of all the conditional probabilities for each parents-child set (*the chain rule for Bayesian Networks* [19]). Figure 6.5 demonstrates a network of local control of non-small-cell lung cancer (LC) in relation to the following clinical and dosimetric variables: age(A), GTV volume (G), PTV coverage (V75, V60), pre-treatment chemo (P) [20]. Using the chain rule, a joint probability can be factorized into:

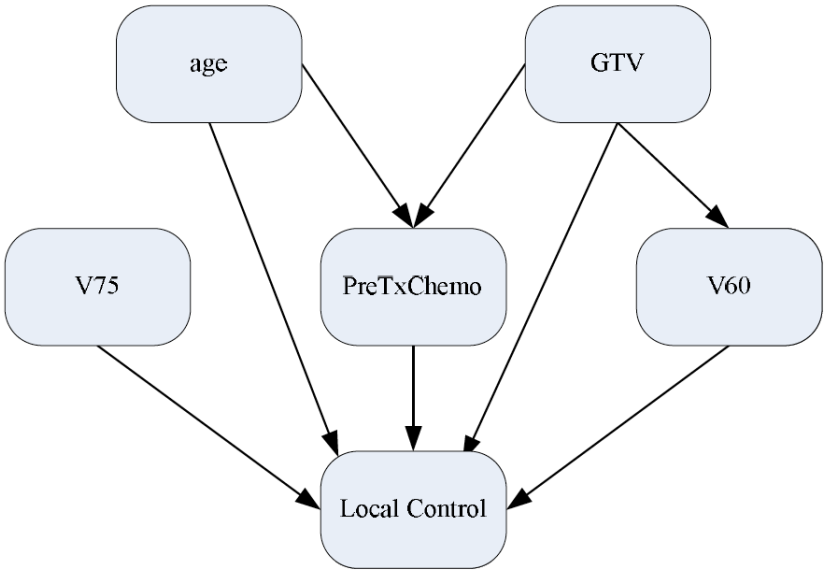

Figure 6.5: A Bayesian Network DAG for predicting local control of NSCLC using radiotherapy and clinical variables. The DAG was trained from clinical data by Oh et al. [20]. (With permission from Oh et al. [20])

$$P(LC, A, G, V75, V60, C)$$
$$= P(A)P(G)P(V75)P(C|A, G)P(V60|G)P(LC|A, G, C, V75, V60)$$

Conditional probability values are often referred to as the "parameters" of Bayesian Network. The parameters can be trained from data as a maximum likelihood estimate or maximum a posteriori (MAP) which incorporates a prior probability with the likelihood obtained from observations.

A DAG can be constructed using prior knowledge on the study domain. When the domain knowledge is not sufficient, observational data can be used to search

for the DAG that can best describe the data. DAG searching can be solved as an optimization problem where a predefined a scoring function is maximized over a space of possible DAG configurations. Searching algorithms can vary according to a choice of the scoring function and searching procedures. Widely used scoring functions include a marginal likelihood (Bayesian) score and a Bayesian Information Criteria (BIC) score. Both scores aim at achieving a balance between the fitness to data (an edge is more likely to be formed between the variables with stronger correlation in data) and complexity of a graph (quantified by the number of edges or parameters), although difference exists in a degree to which complexity is penalized. Mathematical details can be consulted in a primer by Koller and Friedman [19].

Since the number of possible DAGs grows super-exponentially with the number of variables, it is impractical to search exhaustively over the entire graph space for the highest-scoring DAG. Various heuristic approaches have been suggested to reduce a computational cost. For example, a greedy search algorithm begins with the empty graph and keeps adding on edges only when it leads to a higher graph score. Also, constraints on graph topology can be imposed to the search algorithm in order to confine a search domain. For example, the search can be restricted to tree-like structures (Chow-Liu trees) [21] or a certain variable ordering that permits only the edges between the variables in descending order (K2 algorithm) [22]. High-scoring DAGs can be discovered by a sampling method such as Markov Chain Monte Carlo (MCMC) [23]. The MCMC algorithm generates samples of DAGs encountered during a random walk over the graph space (Markov chain), which can be approximated as a posterior distribution of DAGs upon convergence of a chain.

The Probabilistic approach of BN makes it suitable for handling uncertainties. Especially in a medical domain, missing records or test results could have a negative impact on prediction performance. Bayesian Network does not require the full observation on its features for prediction, as it is capable of building and marginalizing joint probability using the conditional dependence relationships between the features. This advantage, in comparison to non-probabilistic classifiers such as the support vector machine (SVM), was shown in survival prediction of lung cancer patients by Jayasurya et al. [24]. Other applications of the BN in radiation oncology include a prognostic network for prostate cancer [25] and lung cancer [20].

6.5 Variable selection

It is often necessary to reduce data dimension in machine learning for three purposes: First, it helps interpretability; It is easier to describe or visualize the data with fewer variables. Second, it is more difficult to train optimal classifiers with the number of variables far exceeding the number of examples due to overfitting. Third, computation and data storing could be hampered by too many variables. Dimensionality reduction can be performed in two ways:

- Feature transformation: A new set of features are created as a result of transforming the data into lower dimensions. Principal component analysis (PCA), projection of data into principal component vectors, is one example.
- Feature selection: A subset of variable is chosen for the purpose of building a data model such as classifiers.

Feature selection methods can be roughly cast into the three categories (while other variants are not mentioned here):

- Filter: Each variable is "scored" individually, independently of a method of inference (e.g. linear regression, graphical models...) and variables above a threshold score are chosen. One simple example would be to rank features in terms of their correlation coefficients to a target variable and choose the significant ones.
- Wrapper: Variables are "wrapped" in a target modelling method for which its predictive performance (cross validation error or bayesian criteria) is obtained. Different combinations of variables are explored until the performance of the resulting model reaches its optimum. When data dimensionality is large, repeating model training with different variable subsets could be computationally demanding; This issue can be solved by the *greedy search algorithm* which arrives at the optimal subset over iterations by making an incremental change on the subset from the previous iteration. *Forward selection* starts at an empty set and progressively add variables, while *backward elimination* starts at a full variable set and keeps removing variables that contribute less to the prediction.
- Direct objective optimization: The number of variables is included in an optimization function as a parameter. The optimizer then tries to improve the goodness of fit while keeping the variable space the smallest.

This section will cover two methods of feature selection that will be used for later chapters. Description of the other feature selection methods can be seen in a review paper by Guyon and Elisseeff [26].

6.5.1 LASSO

Least absolute shrinkage and selection operator (LASSO) [27] is a type of direct objective optimization method that uses regularized logistic regression as an objective function (section 6.4.1). It capitalizes on the property of L_1 regularization in inducing sparsity (variables with zero coefficients). Compared to ridge regression, LASSO has a parsimonious advantage, which can improve prediction accuracy in the presence of irrelevant features [28]. Moreover, compared to stepwise (greedy search) or best subset selection, LASSO provides a more smooth form of solution [29]. However, LASSO requires tuning of a regularization parameter λ (equation) which is typically optimized in a cross-validation setting.

6.5.2 Koller-Sahami filtering

The Koller-Sahami (KS) variable filter [30] is an archetype of the family of Markov-blanket based variable filtering methods. A Markov blanket with respect to a class variable (C) refers to a set of variables (M) that makes all the other variables (B) independent of the class conditioned on the variables in the blanket:

$$P(C|M,B) = P(C|M) \tag{6.16}$$

In other words, once the values of the blanket are all known, the variables outside the blanket become superfluous when it comes to classifying a target variable. This concept of the Markov blanket is applied to finding the smallest subset of the variables that can reconstruct the probability distribution for a class variable. Information theory based metrics such as entropy or mutual information are used as criteria for determining which variables should be added to a blanket. Information theory,

conceived by Claude Shannon in his seminal paper in 1948 [31], measures information of a random variable as the minimum number of bits required to encode it. This property is called *entropy* and mathematically defined with respect to a probability distribution $p(x)$ as:

$$H(x) = -\sum_{x} p(x) \log_2 p(x) \tag{6.17}$$

A goal of supervised learning is to approximate the probability distribution of a class variable. *Kullback-Leibler (KL) divergence*, a measure of difference between two distinct distributions, can be used to indicate the goodness of approximation. Let us say that we use another distribution $q(x)$ as an approximation for the class $p(x)$. KL divergence of $q(x)$ to $p(x)$, denoted as $D(p||q)$, is defined as:

$$D(p||q) = \sum_{x} p(x) \log_2(p(x)/q(x)) \tag{6.18}$$

The value of the KL divergence is smaller when $q(x)$ approximates $p(x)$ better.

The quantity of interest for variable selection is difference between $P(c|\mathbf{G})$ and $P(c|\mathbf{G}^*)$ where $\mathbf{G}$ and $\mathbf{G}^*$ are respectively an original and reduced feature set. The difference between $P(c|\mathbf{G})$ and $P(c|\mathbf{G}^*)$, measured in terms of the KL divergence $D(P(c|\mathbf{G})||P(c|\mathbf{G}^*))$, indicates how well the class can be approximated by reduced features. The KS filter performs a greedy search for the best feature set $\mathbf{G}^*$; It starts at a full variable set and iteratively eliminates the least scoring variables until the optimal model order is reached (backward elimination). At each round of elimination, the following steps are taken:

1. For each variable F_i, a Markov blanket M_i of size n is formed; It consists of n variables $\{F_j\}$ in the remaining variable list with the smallest KL divergence $D(P(c|F_i)||P(c|F_j))$
2. Compute the *information gain* of F_i which is defined as: $D(P(c|M_i \cup F_i)||P(c|M_i))$. In other words, it measures how much information it contributes to the class given a Markov blanket M_i.
3. Remove from the feature list the variable F_i with the least information gain.

The size of the blanket M_i can be tuned and smaller size is preferred for modest size datasets in order to prevent inaccurate conditional probability values. According to the original paper, there is no firmly quantitative criteria for stopping the elimination rounds although a sudden increase in information gain might be used as an indication.

The advantage of Markov blanket based methods, including the KS filter, is that selection of variables is not biased to a certain classification algorithm that ensues due to its information theoretic approach [26].

6.6 Computational approach for RILD modelling

Ever since the LKB model, logistic regression has been a popular choice of a multivariate method for many radiotherapy outcome modelling problems; This choice roots from the observation that dose response usually follows an S-shaped curve [32]. Moreover, logistic regression is an intuitive way of investigating the effects of multiple factors on an outcome; The fitted coefficients and p-values from the likelihood ratio test[2] indicates the relative importance of each variable to prediction

[33]. The logistic regression based computational framework for radiotherapy outcome modelling was implemented by El Naqa and colleagues in a software package Dose Response Explorer System (DREES) [34]. The software also has features for model order determination and variable selection based on information criteria and forward selection. The DREES was applied to prediction of pneumonitis by Bradley et al. [35] and Hope et al. [36]. The former investigators created a two-parameter RP model (P(RP)) which was selected from bootstrap retraining:

$$P(RP) = \sigma(-1.5 + 0.11\text{MLD} - 2.8\text{GTV\_SIposition}) \tag{6.19}$$

where σ is a logistic function.

However, logistic regression is limited by its simplistic assumption that every predictor contributes to a dependent variable independently to each other. It is still possible to assign interaction terms, but such modification needs to be determined *a-priori*. Bayesian network is a desirable alternative to logistic regression, not only for modelling of RILD but also other medical conditions, in a number of ways:

- Interpretability and hypotheses generation: Combining covariates in a graphical way gains not only in prediction but also in subject-matter knowledge. The trained graph might point to the relationships that were not previously known but potentially important in understanding disease onset. This explicit

[2] The likelihood ratio test derives a p-value from the ratio of the likelihood of the model that contains the variable of interest to the model without it.

modelling of interaction blends well with the paradigm of systems radiobiology (section 3.6).

- Probabilistic property: Practically, it is often difficult to collect all the information on predictive features; Missing data is not uncommon in medical data collection (e.g. patients missing a blood test). Probabilistic inference by BN provides flexibility of the model to operate without complete information.
- Intrinsic variable selection: Unlike logistic regression, not all the variables are used for classification. Topology of a DAG determines the most important variables for prediction (Markov blanket). This property reduces effective dimensionality of the model with regards to a class variable.

The remainder of the study will use this concept of Bayesian network to create a data model that explains RP onset using biological, clinical and dosimetric variables. Predictive potential of the BN approach will be tested in reference to the logistic regression model. However, choice of variables might favour one method over another. For example, the variables selected by the LASSO filter are already optimized for the best predictive performance with a logistic function. We hypothesize that the Koller-Sahami method provides a selection of features independent of a particular machine learning method due to its information theoretic approach.

The Bayesian ensemble approach will attempt to overcome the difficulty in finding the optimal Bayesian network structure; Posterior distribution of BN models will be approximated form graph samples obtained by the Markov-Chain Monte Carlo method. Prediction of RP will be made by averaging the prediction results from individual models weighted by posterior. I implemented this computational

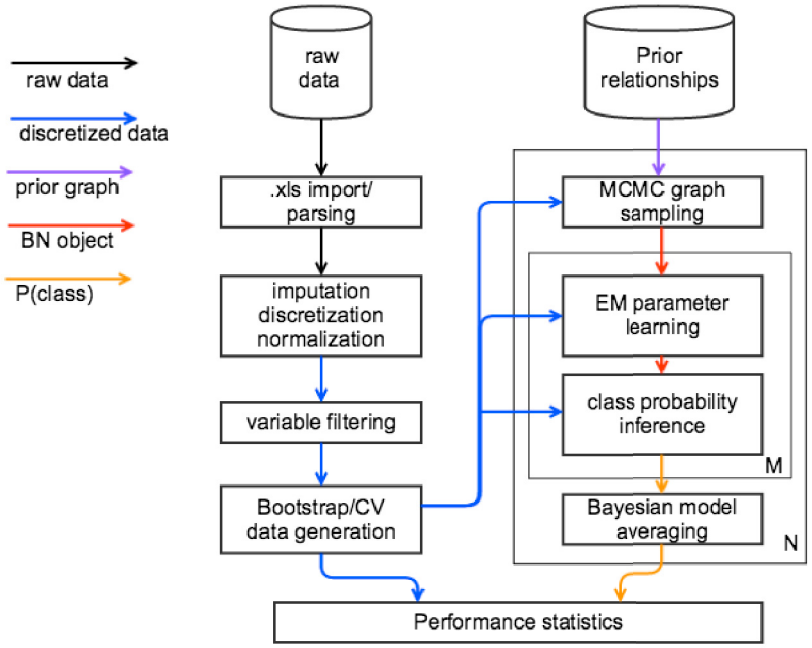

Figure 6.6: Schematic diagram of the modules constituting the MCMCBNE package. M: number of graphs in one ensemble, N: number of bootstrap/CV datasets

strategy into a MATLAB-based code package Markov Chain Monte Carlo sampling for Bayesian Network Ensemble (MCMCBNE) which is released in: `https://github.com/meson200/MCMCBNE`. Figure 6.6 summarizes the modules consisting the code system and data flow.

REFERENCES

[1] C. M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[2] J. O. Deasy, I. E. Naqa, Radiation Oncology Advances, Springer US, Boston, MA, 2008, Ch. Image-Based Modeling of Normal Tissue Complication Probability for Radiation Therapy, pp. 211–252.

[3] G. Schwarz, Estimating the Dimension of a Model, The Annals of Statistics 6 (2) (1978) 461–464.

[4] B. Efron, R. Tibshirani, Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy, Statistical Science 1 (1) (1986) 54–75.

[5] B. Efron, Estimating the error rate of a prediction rule: Improvement on cross-validation, Journal of the American Statistical Association 78 (382) (1983) 316–331.

[6] B. Efron, R. Tibshirani, Improvements on cross-validation: The 632+ bootstrap method, Journal of the American Statistical Association 92 (438) (1997) 548–560.

[7] B. Efron, R. Tibshirani, An Introduction to the Bootstrap, Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Taylor & Francis, 1994.

[8] B. Sahiner, H. P. Chan, L. Hadjiiski, Classifier performance prediction for computer-aided diagnosis using a limited dataset, Med Phys 35 (4) (2008) 1559–70.

[9] R. E. Kass, L. Wasserman, The selection of prior distributions by formal rules, Journal of the American Statistical Association 91 (435) (1996) 1343–1370.

[10] L. Wasserman, Bayesian model selection and model averaging, Journal of Mathematical Psychology 44 (1) (2000) 92 – 107.

[11] D. J. C. MacKay, Information Theory, Inference & Learning Algorithms, Cambridge University Press, New York, NY, USA, 2002.

[12] J. A. Hoeting, D. Madigan, A. E. Raftery, C. T. Volinsky, Bayesian model averaging: a tutorial, Statistical Science 14 (1999) 382–417.

[13] D. Madigan, A. E. Raftery, Model selection and accounting for model uncertainty in graphical models using occam's window, Journal of the American Statistical Association 89 (428) (1994) 1535–1546.

[14] J. Geweke, Bayesian inference in econometric models using monte carlo integration, Econometrica 57 (6) (1989) 1317–39.

[15] M. K. Cowles, B. P. Carlin, Markov chain monte carlo convergence diagnostics: A comparative review, Journal of the American Statistical Association 91 (1996) 883–904.

[16] S. Brooks, A. Gelman, G. Jones, X. Meng, Handbook of Markov Chain Monte Carlo, Chapman & Hall/CRC Handbooks of Modern Statistical Methods, CRC Press, 2011.

[17] R. Fletcher, Practical Methods of Optimization; (2Nd Ed.), Wiley-Interscience, New York, NY, USA, 1987.

[18] A. E. Hoerl, R. W. Kennard, Ridge regression: Biased estimation for nonorthogonal problems, Technometrics 12 (1) (1970) 55–67.

[19] D. Koller, N. Friedman, Probabilistic Graphical Models: Principles and Techniques - Adaptive Computation and Machine Learning, The MIT Press, 2009.

[20] J. H. Oh, J. M. Craft, R. Townsend, J. O. Deasy, J. D. Bradley, I. El Naqa, A bioinformatics approach for biomarker identification in radiation-induced lung inflammation from limited proteomics data, Journal of Proteome Research 10 (3) (2011) 1406–1415.

[21] C. Chow, C. Liu, Approximating discrete probability distributions with dependence trees, IEEE Trans. Inf. Theor. 14 (3) (2006) 462–467.

[22] G. F. Cooper, E. Herskovits, A bayesian method for the induction of probabilistic networks from data, Mach. Learn. 9 (4) (1992) 309–347.

[23] D. Madigan, J. York, D. Allard, Bayesian Graphical Models for Discrete Data, International Statistical Review 63 (2) (1995) 215–232.

[24] K. Jayasurya, G. Fung, S. Yu, et al., Comparison of bayesian network and support vector machine models for two-year survival prediction in lung cancer patients treated with radiotherapy, Medical Physics 37 (4) (2010) 1401–1407.

[25] W. P. Smith, J. Doctor, J. Meyer, I. J. Kalet, M. H. Phillips, A decision aid for intensity-modulated radiation-therapy plan selection in prostate cancer based on a prognostic bayesian network and a markov model, Artificial Intelligence in Medicine 46 (2) (2009) 119 – 130.

[26] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, J. Mach. Learn. Res. 3 (2003) 1157–1182.

[27] R. Tibshirani, Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society (Series B) 58 (1996) 267–288.

[28] A. Y. Ng, Feature selection, l1 vs. l2 regularization, and rotational invariance, in: In the 21st International Conference on Machine Learning, 2004, p. 78.

[29] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer, 2009.

[30] D. Koller, M. Sahami, Toward optimal feature selection, in: In 13th International Conference on Machine Learning, 1995, pp. 284–292.

[31] C. E. Shannon, A mathematical theory of communication, SIGMOBILE Mob. Comput. Commun. Rev. 5 (1) (2001) 3–55.

[32] S. Lee, I. El Naqa, Machine learning methodology, in: I. El Naqa, R. Li, M. J. Murphy (Eds.), Machine Learning in Radiation Oncology, Springer International Publishing, 2015, pp. 21–39.

[33] L. Breiman, Statistical modeling: The two cultures, Statistical Science 16 (3) (2001) 199–231.

[34] I. E. Naqa, J. Bradley, A. I. Blanco, et al., Multivariable modeling of radiotherapy outcomes, including dose volume and clinical factors, International Journal of Radiation Oncology*Biology*Physics 64 (4) (2006) 1275 – 1286.

[35] J. D. Bradley, A. Hope, I. E. Naqa, et al., A nomogram to predict radiation pneumonitis, derived from a combined analysis of RTOG 9311 and institutional data, International Journal of Radiation Oncology*Biology*Physics 69 (4) (2007) 985

– 992.

[36] A. J. Hope, P. E. Lindsay, I. E. Naqa, et al., Modeling radiation pneumonitis risk with clinical, dosimetric, and spatial parameters, International Journal of Radiation Oncology*Biology*Physics 65 (1) (2006) 112 – 124.

CHAPTER 7
Bayesian Network ensemble as a multivariate strategy to predict radiation pneumonitis risk

Contents

7.1 Introduction

Chapter 2 and 3 introduced different RILD factors from dosimetric, clinical and biological domains. These factors are often studied in isolation from other factors i.e. their individual odds ratio on RILD risk. There is a practical need for creating a multivariate RILD model that considers the proposed factors from all the domains to exploit their predictive potential.

However, practical concerns arise from creating a RILD model that spans multiple knowledge domains. Chapter 6 discussed the bias-variance tradeoff, where predictive performance in unseen data can decrease with the complexity of a model. Moreover, too many variables in a model can compromise its interpretability. The same chapter also introduced the basic principles of Bayesian network and why it is a good candidate multivariate method for prediction of medical conditions. In this chapter, radiation pneumonitis was modelled using Bayesian network in which interactions between RP risk factors were detected and used for prediction of RP.

The study presented in this chapter was published as the following paper: "Bayesian network ensemble as a multivariate strategy to predict radiation pneumonitis risk" authored by: Lee, S., Ybarra, N., Jeyaseelan, K., Faria, S., Kopek, N., Brisebois, P., Bradley, J. D., Robinson, C., Seuntjens, J., and El Naqa, I. Med. Phys. 42(5): 2421-30, 2015.

7.2 Background

Radiotherapy (RT) is considered a standard of care for medically inoperable locally-advanced non-small-cell lung cancer (NSCLC) [1]. Despite rapid advances in radiation delivery techniques during the last two decades, prognosis of NSCLC remains poor, with 5-year survival stalled at 18.2% [2]. RT dose escalation can possibly benefit marginal survival outcome with chemoradiation [3]. However, the main clinical obstacle to dose escalation is excessive normal tissue toxicity and especially the risk of radiation pneumonitis (RP). In this regard, accurate prediction of RP risk may be useful for cancer cure as well as enhancing the quality of life for patients receiving radiation treatment [4].

Current RP prediction models in clinical use rely almost entirely on dosimetric parameters such as average dose to lung or percentage of irradiated lung volume [5]. However, it has been reported that patients display heterogeneity in normal tissue radiosensitivity given identical radiotherapy regimen [6]. Identifying over- or under-responding patients has been attempted on the basis of underlying biological reactions responsible for the endpoint of interest. Fleckenstein et al. [7] describes the pathogenesis of radiation-induced lung disease as multiple inter-reacting cellular activities such as hypoxia, fibrogenesis, inflammation, and angiogenesis. This theory has been corroborated by several independent studies that reported association between RP risk and biomarkers for those processes (Transforming growth factor(TGF)-β1 [8] [9], Interleukin(IL)-6 [10], alpha-2-macroglobulin(α2M) [11], Angiotensin converting enzyme(ACE) [12] [13]). However, significance of those proteins

as an individual predictive biomarker is still debating, and multivariate analysis on those biomarkers for prediction has yet to be investigated.

Some investigators combined dosimetric and clinical RP risk factors through multivariate logistic modelling to enhance the predictive performance of univariate models [14] [15]. However, logistic regression is limited in its flexibility to embrace dependence among multiple features. There is growing emphasis from systems biology perspective, as proposed by a few authors, [16] [4] to better understand structures of dependence between biophysical variables such as hierarchy or modulation. We propose Bayesian Network (BN) as a framework to render such arrangements possible. BN is a graphical model designed for modelling joint probability distribution among random variables. Calculation of joint probability is greatly facilitated by conditional independence relationships encoded in a directed acyclic graph (DAG). In this way, we can derive a probabilistic classifier from BN by estimating the probability of a class conditional on the covariates with known values. A DAG also provides visual representation of causal relationships, which is appealing to end-users who wish to interpret inference results intuitively.

A DAG can either be specified by experts or learned from observational data. Learning DAGs from data is known to be computationally complex due to super-exponentially growing possibilities of possible graphs with the number of variables [17]. Such complexity of the model poses a big challenge to many model selection schemes which select the single best-fitting model based on a predefined score function. However, the schemes do not always lead to an optimal classification on unseen data due to the way that the score function is designed [18]. An alternative to model

selection is retaining multiple models in an ensemble and taking the average of prediction results by the ensemble members weighted by their posterior probabilities. This approach, called Bayesian model averaging [19], has been shown experimentally to improve performance of complex models [20], including the multivariate RP prediction models by Das et al. [21].

Bayesian Network has been in use for radiotherapy outcome prediction, mainly for tumour control [22] [23] [24]. To our knowledge, it has not been applied to radiation induced toxicity and especially RP where univariate predictors achieved limited success. In this paper, we intended to elaborate on the methods of BN ensemble learning as a proof of concept that could in the future improve risk prediction and also generate hypotheses on the pathophysiology of radiation-induced diseases.

7.3 Materials and Methods

7.3.1 Patient cohort

Fifty four stage III NSCLC patients were included in this study according to the following criteria: 1) received standard 2-Gy-per-fraction 3D conformal radiotherapy with curative intent, 2) have no history of previous lung irradiation, and 3) baseline Karnofsky performance status (KPS) of equal to or greater than 70. Chemotherapy was also given neoadjuvantly or concurrently except 2 patients. Blood samples for biomarkers were first acquired on the CT simulation day (pre-treatment) for baseline and the 15th fraction (mid-treatment). Post-radiotherapy toxicity was quantified using Common Toxicity Criteria for adverse events (CTCAE) version 3. Incidence of RP, classified as CTCAE toxicity grade 2 or higher, was reported for 35% (19/54) of the patients. Detailed cohort characteristics can be seen in the table 7.1.

Table 7.1: Patient cohort characteristics. *: heterogeneity-corrected, PTV excluded

	Patient count (%)
Cohort size	32
RP grade	
0	34 (63)
1	1 (2)
2	11 (20)
3	6 (11)
5	2 (4)
2 and above	19 (35)
Gender	
male	23 (43)
female	31 (57)
Prescription dose (Gy)	
median	60.0
range	56.0-70.0
Mean lung dose*	
median	15.0
range	4.8-24.0
Chemotherapy	
No	2 (4)
Concurrent	20 (37)
Neoadjuvant	1 (2)
Neoadjuvant+Concurrent	29 (54)
Adjuvant	2 (4)

7.3.2 Candidate variable list

We collected from blood samples and archived radiation oncology database all the information on candidate RP covariates with reported association in the literature (see table 7.2 for references). From each blood sample, serum concentration of the following candidate RP biomarkers were measured by enzyme-linked immunosorbent assay (ELISA): transforming growth factor (TGF)-β1, interleukin(IL)-6, angiotensin converting enzyme (ACE), and α2M. In addition, the following dose-volumetric parameters were extracted from RT treatment plans based on Anisotropic Analytical Algorithm (AAA) heterogeneity correction: mean lung dose (MLD), mean heart dose (MHD), percentage volume of lung receiving 20 Gy/30 Gy or less (V20/V30), PTV volume, and superior-inferior position of PTV normalized to apical-basal extent of lung (PTVCOMSI) [25]. In order to account for dosimetric uncertainty due to tumour motion, PTV volume was excluded from lung dose calculation. In summary, the raw dataset contained the RP incidence and 16 RP covariates in the four main categories shown below:

- Baseline concentration (_pre) of TGF-β1, IL-6, ACE, and α2M.
- Mid-treatment concentration (_ratio) of TGF-β1, IL-6, ACE, and α2M, taken as a percentile increase from baseline.
- RT plan parameters: MLD, MHD, V20, V30, PTV volume, PTVCOMSI.
- Clinical patient information: age, smoking status.

7.3.3 Data pre-processing

For Bayesian Network modelling, all the covariates were discretized into 2 bins (high/low) because finer discretization led to sparsity in conditional probability estimation. The binning was performed using the K nearest neighbour clustering algorithm [26] which grouped the data points into two clusters based on a distance (figure 7.1). The Koller-Sahami (KS) variable filter [27] was used for reducing the number of covariates. The KS filter iteratively removes the variables that give little or no additional information (entropy) to a class in the presence of other variables (blankets). For our application, a blanket size of 1 was chosen because larger size resulted in inaccurate entropy values in our data. The filtering was implemented in the following three sequential st.: the KS filter shrunk the data down to 2 variables (the smallest number allowed with a blanket size = 1) and the change in the entropy of the removed along the 14 elimination rounds was recorded. The optimal dimensionality of data was defined as the number of variables with the cross-entropy higher than the value of a dummy variable filled with random values. Then, the KS filter was reapplied to the full dataset down to the determined dimensionality. This was repeated in 1000 bootstrap samples and the variables that were selected most frequently were retained for the next steps.

7.3.4 Bayesian Network learning

The Metropolis-Hastings MCMC sampling for Bayesian Network graphs [20], implemented in the Bayesian Network Toolbox (BNT) [28], was chosen as a graph searching algorithm. An acceptance ratio for drawing a Markov chain sample was determined by the ratio of a marginal likelihood graph score, $P(G|D)$, and a proposal

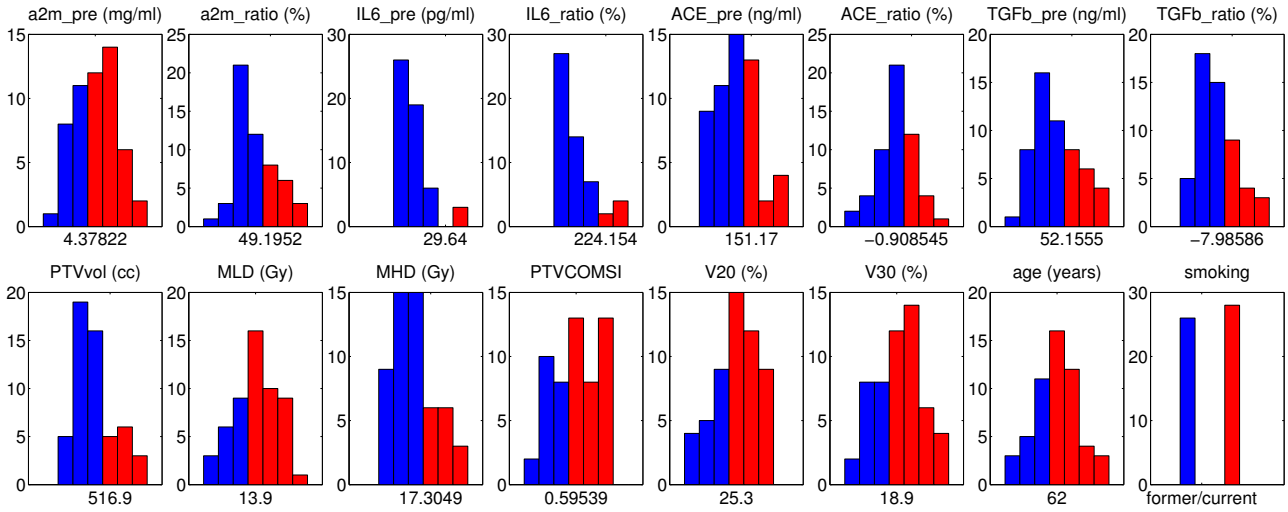

Figure 7.1: KNN clustering discretization of the 16 candidate covariates into high (red) and low (blue) bins. The raw values were binned into a histogram with 7 equally spaced intervals ranging ±2 standard deviation (except smoking). The high-to-low bin boundaries were shown on the x-axis.

distribution $Q(G_{new}; G_{old})$ between the new graph sample G_{new} and the last sample in the chain G_{old}:

$$A = A_1 A_2, A_1 = \frac{P(G_{new}|D)}{P(G_{old}|D)}, A_2 = \frac{Q(G_{old}; G_{new})}{Q(G_{new}; G_{old})} = \frac{N(G_{old})}{N(G_{new})} \tag{7.1}$$

A uniform distribution $Q(G_{new}; G_{old}) = \frac{1}{N(G_{old})}$ was used for the proposal density, where $N(G_{old})$ is the number of the valid graphs that can be created from G_{old} satisfying the conditions given below:

- Should be able to be created from G_{old} by modification on a single edge (addition of a new edge or deletion/reversal of an existing edge)
- The graph after modification has to satisfy the following constraints:
  1. No loop is formed by edges (acyclicity)

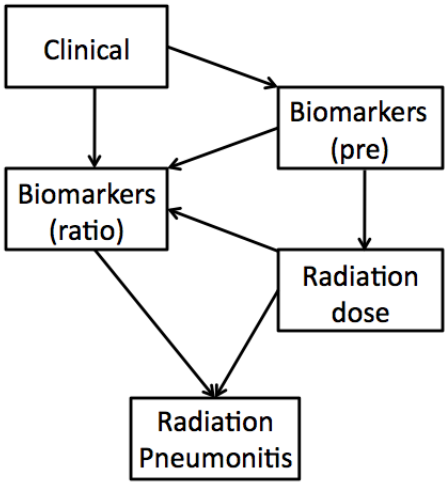

Figure 7.2: Diagram of allowed causal links between variable categories used for accepting/rejecting graph samples during MCMC simulation.

2. The number of parents for any node is no more than 3 (maximum fan-in)
3. Every edge is in a causal direction (causality prior)

A causality prior imposes categorical restrictions on the presence or the direction of edges between nodes in order to reduce the search space. The allowed connections between the four categories of RP covariates (see the methods section) and an RP node (figure 7.2) are compatible with temporal order (e.g. baseline biomarkers $\rightarrow$ mid-treatment biomarkers $\rightarrow$ RP) or cause-effect relationship (e.g. radiation $\rightarrow$ mid-treatment biomarkers).

Twenty-five Markov chains were created with random initial graphs and grown until convergence. Convergence of the chains was checked at every 10000 MCMC samples. At each checkpoint, the samples were histogrammed into a graph posterior distribution which consist of distinct graphs in the chains and the frequencies of

their occurrence. Convergence was judged by sufficient consistency in the posterior distribution over time. After the convergence, the first 10000 samples in the chains were discarded due to strong correlation with arbitrary initial condition (burn-in period) and did not contribute to estimating posterior distribution. For every graphs in the posterior, Maximum a Posteriori (MAP) CPTs for parent-child pairs in the DAG were learned from data with equivalent sample size $\alpha = 1$. The expectation-maximization (EM) algorithm was applied in conjunction with MAP in order to handle missing data which constituted 9.8% of the data.

7.3.5 Ensemble Bayesian Network classifier

A Bayesian Network structure with a complete graph and CPTs is capable of making inference on the probability of RP. The inference was performed by the junction tree algorithm implemented in the BNT toolbox [28] which is capable of dealing with variables with an unknown value in a query. Probability of RP was computed for each Bayesian Network graph in an ensemble which consisted of the MCMC graph samples with N highest posterior probability $P(i|D)$. These probability estimates were averaged over the ensemble weighted by the posterior probability of the graphs to yield the final estimate of the RP probability:

$$P(RP) = \frac{\sum_{i=1}^{N} P_i(RP)P(i|D)}{\sum_{i=1}^{N} P(i|D)} \tag{7.2}$$

Predictive performance of the BN ensemble model was quantified by and receiver operating characteristics (ROC) and a reliability plot . An ROC curve plots true positive rates against false positive rates while varying classification threshold on the

estimated $P(RP)$. The measured ROC metrics include the area under the ROC curve (AUC), sensitivity, specificity and accuracy at the optimal operating threshold that maximizes the sum of sensitivity and specificity according to Youden's J statistics. While ROC measures discriminative ability of a model as a classifier, a reliability plot shows accuracy of probability estimates. In order to create a reliability plot, the patients were distributed into 5 risk groups based on the $P(RP)$ computed by the BN models and their actual risk (an event rate in each group) was plotted against the modelled risk (an average $P(RP)$ in each group).

7.3.6 Bootstrap validation of the Bayesian Network model

Graphical features and prediction power of the BN ensemble were validated in 200 non-parametric bootstrap datasets, which was intended to test robustness of the BN learning methods to limited size of data. Confidence in the presence of a certain link in a graph (f), denoted as $P(f)$, was estimated from M bootstrap replicates according to the following formula (equation 7.3) [29]:

$$P(f) = \sum_{i=1}^{M} Pr(f|D_i) \tag{7.3}$$

Where $Pr(f|D_i)$ denotes the probability of detecting f from Bayesian Network graph training on a bootstrap replicate D_i, and can be approximated by the posterior-weighted average occurrence of the link in the ensemble ($\mathcal{G}$) [30]:

$$Pr(f|D_i) \sim \frac{\sum_{G \in \mathcal{G}} Pr(G|D_i) f(G)}{\sum_{G \in \mathcal{G}} Pr(G|D_i)} \tag{7.4}$$

Classification accuracy of the BN model was also validated using the same 200 bootstrap replicates with the 0.632+ bootstrap error calculation method [31].

7.3.7 Comparison with other RP predictors

Predictive performance of the BN ensemble model was compared with univariate predictors in the dataset and a multivariate logistic regression predictor which was built on the covariates selected by the KS filter. For these predictors, unlike for the BN modelling, the input data retained its continuous scale after standard normalization. Missing data was filled in by K nearest neighbour imputation prior to the training. In order to minimize overfitting, a L2 regularizing term was added to a mean-square error for the objective function. The coefficient of the regularization term was tuned by 10-fold cross validation repeated 100 times with randomly assigned folds. ROC metrics using these predictors on the 200 bootstrap replicates were compared with the results from the BN model in a pairwise fashion by the Wilcoxon's signed rank test with $\alpha = 0.05$.

7.4 Results

7.4.1 Variable selection

The KS filtering on the full dataset resulted in the optimal dimensionality of 6 (figure 7.3) and the following chosen variables: mid-treatment α2M, 2) pre-treatment IL-6, 3) mid-treatment ACE, 4) MHD, 5) V20, and 6) PTVCOMSI (table 7.2). The selected variables displayed a varying degree of univariate correlation with RP, with V20 at the weakest (odds ratio = 1.32) and MHD at the strongest (odds ratio = 2.46).

Table 7.2: Candidate RP covariates. The underlined variables were included into the BN model via KS filtering.
† Bootstrap frequency of 16 RP covariates selected by the KS variable filtering at a model order 6.

variable	odds ratio	selection frequency†
α2M_pre [11]	0.58 (0.30, 1.10)	0.321
$\underline{\alpha 2M\_ratio}$ [11]	1.68 (0.93, 3.01)	0.545
IL6_pre [10]	1.89 (0.81, 4.40)	0.404
IL6_ratio [10]	1.33 (0.76, 2.34)	0.3
ACE_pre [12]	0.76 (0.42, 1.38)	0.205
ACE_ratio [12]	2.17 (1.01, 4.67)	0.382
TGFβ_pre [8]	0.92 (0.52, 1.63)	0.357
TGFβ_ratio [8]	0.99 (0.56, 1.74)	0.204
MLD [32]	1.35 (0.75, 2.44)	0.254
MHD [33]	2.46 (1.26, 4.83)	0.696
V20 [34]	1.32 (0.74, 2.36)	0.647
V30 [34]	1.60 (0.88, 2.93)	0.207
PTV volume [34]	0.99 (0.56, 1.74)	0.21
PTVCOMSI [25]	0.63 (0.35, 1.14)	0.613
age [35]	0.95 (0.54, 1.67)	0.366
smoking [35]	1.21 (0.68, 2.13)	0.289

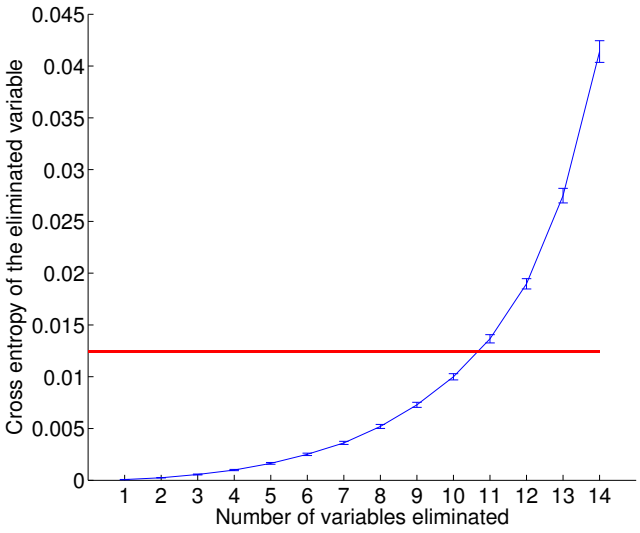

Figure 7.3: Cross entropy of the variables removed at each round of KS backward elimination. A red area indicates the range of cross entropy from a dummy variable. Error bars: 95% confidence interval from bootstrapping.

7.4.2 Bayesian Network learning with MCMC

Figure 7.4 demonstrates that likelihood of the sampled graphs rose from the initial suboptimal locations to reach an oscillatory equilibrium shortly after 1000 runs. However, equilibrium on estimating posterior distribution of graphs did not occur until at least 50000 runs (figure 7.5). The convergence of the posterior distribution was attributed to a restricted search space imposed by the causality prior, as the chain failed to converge within reasonable time in the absence of such conditions.

7.4.3 Confidence in BN graphical features

When the ensemble size of 200 was used, a median confidence level of the 23 possible connections allowed by the causal prior was 0.27. 10 out of 23 achieved the

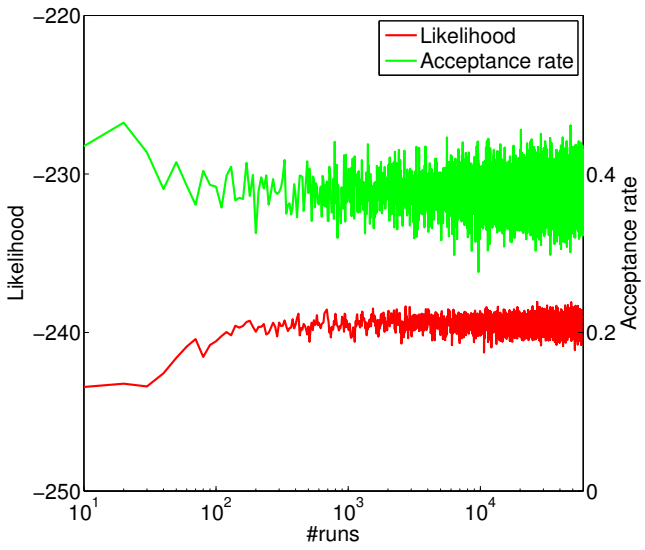

Figure 7.4: Change in likelihood graph score and acceptance rate over 60000 MCMC iterations shown on log scale. 25 Chains with random initialization were averaged.

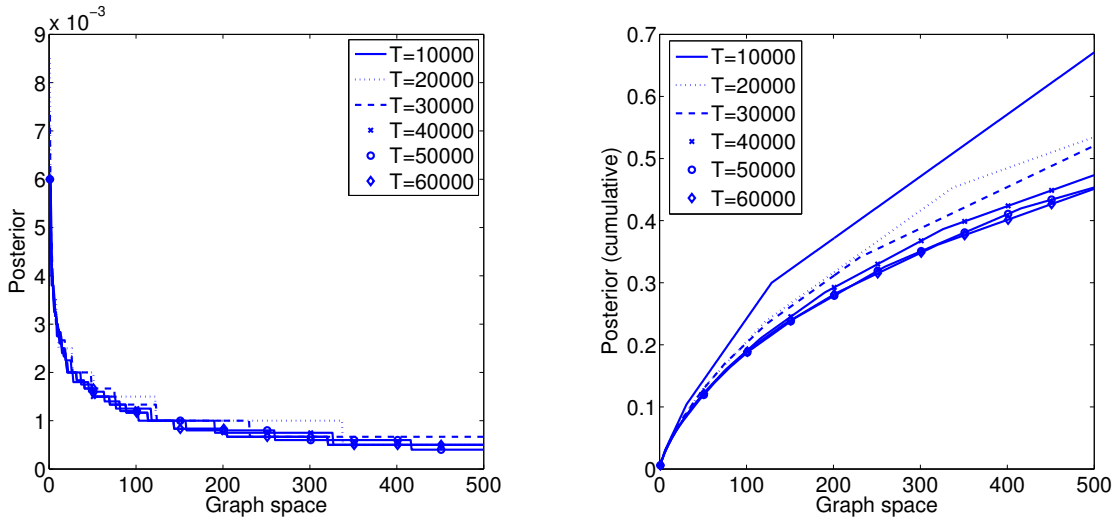

Figure 7.5: Monte-Carlo posterior estimation of graphs over MCMC runs up to 60000 iterations (T). Left: differential, right: cumulative.

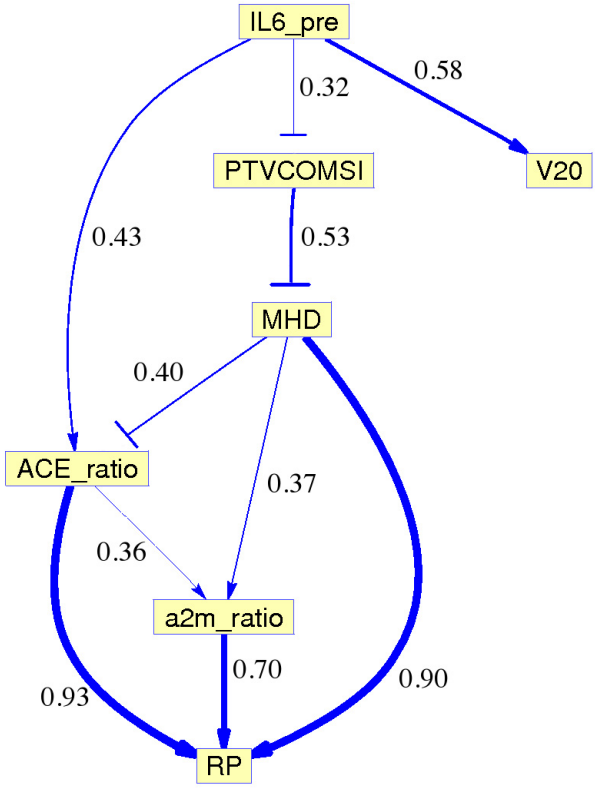

Figure 7.6: Variables connected by directed edges with a confidence level higher than random. Edge thickness is proportional to its confidence level which is also written on the edge (table 7.3). Arrow-headed and bar-headed edges are assigned to positive and negative correlation, respectively.

confidence level higher than chance (0.29) [36] (table 7.3, figure 8.2). The highest value was recorded for the connection from ACE_ratio to RP (0.93).

7.4.4 Prediction of RP risk

Predictive power of the Bayesian Network method was compared with that of individual biomarkers and dosimetric RP models in the literature [37] citeNiemierko97 [38]. As seen by changes in AUC and specificity (figure 7.8), prediction with the

Table 7.3: Bootstrap estimated confidence of the connections shown in the figure 8.2.

parent	child	confidence
ACE_ratio	RP	0.93
MHD	RP	0.90
α2M_ratio	RP	0.70
IL6_pre	V20	0.58
PTVCOMSI	MHD	0.53
IL6_pre	ACE_ratio	0.43
MHD	ACE_ratio	0.40
MHD	α2M_ratio	0.37
ACE_ratio	α2M_ratio	0.36
IL6_pre	PTVCOMSI	0.32

BN method greatly improved as the ensemble size grew up to 200 beyond which the increase tailed off. For any size of the ensemble, the AUC for the BN was larger than any univariate classifiers (figure 7.7). The highest AUC from univariate biomarkers and dosimetric models was recorded by mid-treatment ACE (0.66) and MHD (0.69), which was significantly lower than the BN model at an ensemble size N=200 (0.83). Given the same set of variables, multivariate logistic regression was shown to be less effective in classification than BN at any ensemble size, as seen from significantly lower AUC and specificity at the optimal operating threshold (figure 7.7, table 7.4).

Accuracy of probability estimates of the multivariate models was evaluated via a reliability plot (figure 7.9). The accuracy was measured as a goodness of fit to a perfect probability estimate which corresponds to a diagonal line in the plot. Similarly to ROC metrics, the Bayesian Network model with larger ensemble size was shown to give more accurate estimates of RP risks.

Table 7.4: Comparison of RP prediction performance in ROC metrics. (Parenthesis: 95% confidence interval) *: ensemble size 200, **: at the probability cutoff at the maximum Youden index.

	AUC	SE**	SP**		AUC	SE	SP
BayesNet*	0.83 (0.82,0.85)	0.80 (0.78,0.83)	0.81 (0.78,0.83)	Biomarkers			
				α2M_pre	0.63 (0.62,0.65)	0.74 (0.71,0.77)	0.66 (0.65,0.67)
Logistic	0.77 (0.76,0.78)	0.81 (0.78,0.83)	0.73 (0.70,0.75)	α2M_ratio	0.63 (0.61,0.64)	0.67 (0.65,0.70)	0.75 (0.74,0.76)
Dosimetry							
MLD	0.62 (0.61, 0.63)	0.75 (0.72, 0.79)	0.62 (0.61,0.64)	IL6_pre	0.57 (0.57,0.61)	0.59 (0.56,0.62)	0.77 (0.74,0.79)
V20	0.58 (0.59,0.62)	0.76 (0.71,0.79)	0.57 (0.55,0.61)	IL6_ratio	0.51 (0.50,0.53)	0.63 (0.59,0.67)	0.65 (0.63,0.67)
V30	0.60 (0.59,0.62)	0.60 (0.57,0.62)	0.78 (0.77,0.80)	ACE_pre	0.57 (0.56,0.58)	0.72 (0.70,0.76)	0.61 (0.59,0.62)
gEUD [39]	0.63 (0.61, 0.64)	0.76 (0.73, 0.80)	0.62 (0.60,0.64)	ACE_ratio	0.66 (0.65,0.67)	0.70 (0.68,0.73)	0.74 (0.72,0.75)
Bradley et al. [38]	0.64 (0.63,0.65)	0.81 (0.78, 0.83)	0.63 (0.62,0.64)	TGF_pre	0.50 (0.49,0.51)	0.82 (0.79,0.85)	0.43 (0.42,0.45)
MHD	0.69 (0.68,0.70)	0.72 (0.70,0.74)	0.78 (0.76,0.79)	TGF_ratio	0.51 (0.51,0.53)	0.65 (0.62,0.70)	0.58 (0.56,0.60)
Clinical							
age	0.53 (0.51, 0.54)	0.48 (0.44, 0.52)	0.76 (0.74,0.78)				
smoking	0.54 (0.53, 0.55)	0.41 (0.37, 0.46)	0.66 (0.65,0.68)				

In order to examine the ability of the Bayesian Network ensemble to perform probabilistic inference under missing data, the probability of RP was estimated in the absence of intra-treatment biomarker data. AUC values decreased significantly as a result, but with varying degrees depending on ensemble sizes. For larger ensemble sizes above 200, the reduction in AUC was smaller (0.03) than for sizes less than 50 (0.05) (figure 7.8).

7.5 Discussion

High input dimensionality is certainly one of the biggest challenges in multivariate disease modelling, especially in radiotherapy induced sequelae where dosimetric, biological and patient specific clinical parameters all contribute to extra complexity. Dependence between risk factors is inevitable in radiotherapy informatics and has been neglected by many multivariate methods such as logistic regression in the name of the "black box" approach. Naive independence assumption of such models may

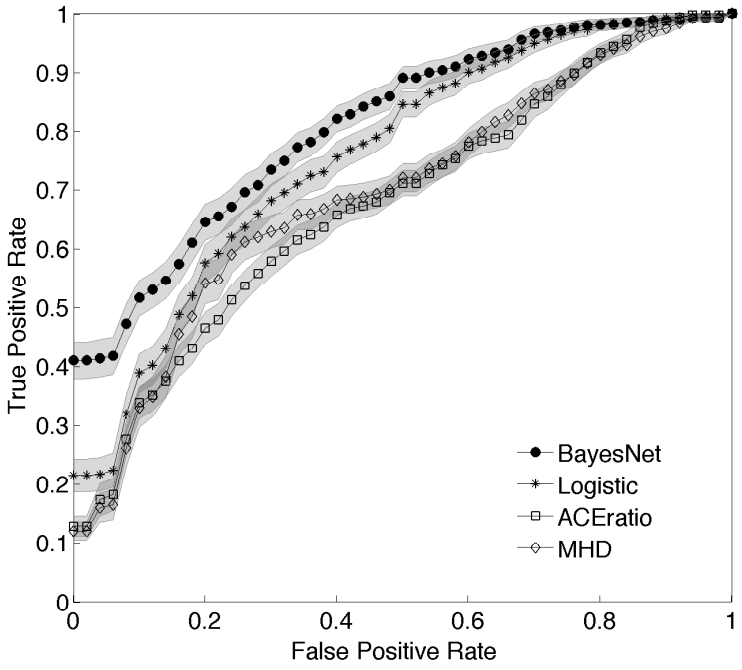

Figure 7.7: ROC curves for RP prediction using Bayesian Network at ensemble size 200, logistic regression and 2 best univariate (mid-treatment ACE, MHD) models. Shaded regions: bootstrap estimated 95% confidence bands.

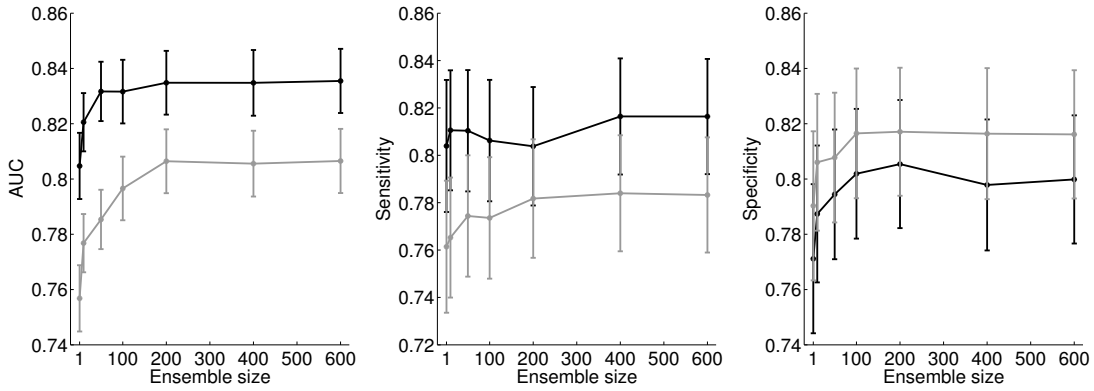

Figure 7.8: ROC metrics using the Bayesian Network model ensemble with a varying size from 1 to 600. Black: prediction with a complete dataset, Grey: prediction without intra-treatment biomarker measurements. Error bars: bootstrap-estimated 95% confidence intervals.

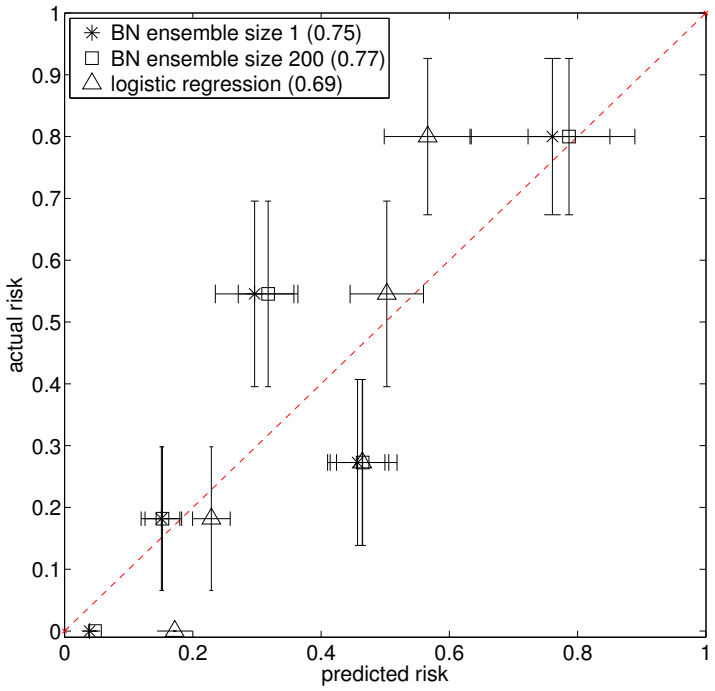

Figure 7.9: Reliability plot of 3 probabilistic classifiers. Numbers in parentheses indicate the r^2 value with respect to an ideal probability estimator (dashed red line). Standard errors, shown in error bars, were computed from binomial distribution (actual risk) or quadrature summation of variance of probability scores in a group and bootstrap-estimated model uncertainty (predicted risk).

lead to increased bias in probability estimation and consequently poor inference with unseen data. Alternatively, this variety of factors could be structured into systems of different hierarchical levels between which connections could be established by fusing observation with prior knowledge. Bayesian Network provides the statistical platform in which this idea of systems biology can be implemented and applied for disease prediction. However, the amount of current expert knowledge in this area does not fully cover all the putative RP risk factors. Thus, feasibility of learning BN structures from radiation oncology data needs to be addressed.

Bayesian Network is intrinsically a complex modelling technique with a multitude of parameters. Several measures were taken in training and testing the BN prediction model to address the issue of overfitting: a causality prior and a fan-in bound imposed restrictions on the number of possible graphs to narrow down the search space. In addition, marginal likelihood as a graph scoring function penalizes larger number of edges and helps find a balance between the complexity and fitness of data [40]. Generalizability of the BN classifier was validated in a 0.632+ bootstrap setting which was shown to be the closest to the true population estimate in small training sample size and high model orders [41]. However, even these efforts do not remove the uncertainty of a single BN model learned from a modest size dataset. We observed from MCMC simulation the presence of multiple high-scoring graphs also known as likelihood equivalence [42]. We addressed this issue by adopting a Bayesian approach which embraces an ensemble of models rather than a single one. Having a bag of models compromises graphical interpretation of a single model but overcomes

the uncertainty in model selection and improves the performance on unseen data [19], which was shown in our cohort by larger AUC with increasing ensemble size.

We observed superiority of the Bayesian Network method over logistic regression in classifying RP events (figure 7.7). The difference could be attributed to a probabilistic approach of Bayesian Network. It models a joint probability distribution $p(\mathbf{x}, y)$ for covariates $\mathbf{x}$ and a class y and for classification computes $p(y|\mathbf{x})$ using the Bayes' rule, in contrary to logistic regression which is limited to computing $p(y|\mathbf{x})$. The advantage of the former, already reported by Ng and Jordan [43] in several datasets with a modest size, might explain more accurate RP probability estimates by BN models in our data (figure 7.9). Also, the probabilistic nature of BN enables inference under incomplete data. For non-probabilistic models such as logistic regression, missing data have to be filled in by imputation, which inevitably corrupts the integrity of data. Furthermore, we were looking into capitalizing on this feature to attempt "hastened" prediction before receiving mid-treatment biomarkers. Although some reduction in AUC was observed, the model to a certain degree compensated for the missing information by using graphical relationships and parameters obtained with complete training data. This is due to the role that ensemble plays (figure 7.8) and its resulting high AUC values (0.76-0.81) compared to MHD (0.70) which was the only variable connected to RP other than mid-treatment biomarkers (figure 8.2).

A Markov-Chain Monte Carlo method was used for exploring a model space and approximating posterior probability. As seen from figure 7.5, the graph samples from MCMC showed power law-like distribution which consists of a small number of high-posterior graphs and a long tail of low posterior. It turned out that only the

graphs around 200, constituting the peak and a small part of the "tail" at cumulative probability less than 30%, is necessary for the optimal RP classification. This result agrees well with previous studies showing that high posterior models are indeed beneficial for classification tasks as well [30]. We also derived from those graphs the confidence levels of graph features through repeated training in bootstrap replicates. Edges with high confidence might imply existence of novel causal relationships. Although we did not conduct any independent bench experiment to validate the links, this feature of the BN model is definitely an advantage in generating new hypotheses compared to many "black-box" type multivariate models.

Precautions need to be taken, however, in interpreting the results of this study for clinical application of the presented Bayesian Network methodology. Small number of RP events due to modest dataset size and low-risk nature implies higher uncertainty in predicting positive instances than the negatives. This was demonstrated in figure 7.9 showing worse precision of probability estimates for higher risk patients. Another limitation, inherent to BN structure learning, is that the connections between variables that we identified from BN graphs may not always indicate direct causality. Although non-causal links were ruled out during the structure learning, it is possible that a pair of variables without any direct causality are seemingly correlated through any confounding factor hidden from the model. Due to computational burden from MCMC and a hazard of overfitting under a limited sample size, input dimension had to be tightly constrained by use of variable selection. The KS filtering scheme was shown to be effective in removing redundant variables [27]. In this study,

it was useful in reducing the dimensionality of a dose volumetric domain that contained highly correlated variables. However, this type of filtering might have dropped the variables that could have been more useful in encoding causal relationships.

Future development of this study will address the effect of different fractionation schemes on variable selection, types of biophysical interactions identified by BN, and RP prediction by applying the described method on a SBRT cohort. We are also developing a method for systematically encoding domain knowledge in radiobiology into the graph learning. In addition to the causality constraints, this is expected to identify true biological relationships with high confidence and thus improve the generalizability of the model. Lastly, although the model was validated internally using 0.632+ bootstrap, an external source of data with matching characteristics would be necessary to add more clinical relevance to our findings.

7.6 Conclusion

We applied a Bayesian Network framework in conjunction with Bayesian statistics for constructing interaction graphs of biological, dosimetric and clinical covariates for radiation pneumonitis. Markov Chain Monte Carlo sampled high posterior graphs which were used as an ensemble to estimate radiation pneumonitis risk. We have shown that a certain size of the ensemble was sufficient to perform optimal classification. Bootstrap-validated predictive power of the Bayesian Network ensemble was superior to any univariate predictors or multivariate logistic regression. Statistical confidence in graph features in the ensemble obtained by bootstrapping can potentially identify novel biophysical relationships. After validation on larger dataset, the

presented modelling strategy could be useful for estimating normal tissue complication probability (NTCP) for various endpoints.

7.7 Acknowledgments

We acknowledge the consults from Dr. Jung Hun Oh and Dr. Natalie Japkowicz. We also thank Dr Christina Speirs for providing us a part of clinical data. The computational work was enabled in part by computer resources provided by WestGrid (www.westgrid.ca) and Calcul Quebec (http://www.calculquebec.ca/). This research was funded by the Canadian Institute of Health Research (CIHR) grant MOP-114910, Excellence in Radiation Research for the 21st Century (CIHR, Canada) and the CREATE Medical Physics Research Training Network grant (NSERC, Canada).

REFERENCES

[1] National Cancer Institute, Non-small cell lung cancer treatment, http://www.cancer.gov/cancertopics/pdq/treatment/non-small-cell-lung/healthprofessional, accessed: 2014-11-15.

[2] National Cancer Institute, Surveilance, epidermiology and end results, Tech. rep. (2013).

[3] M. Machtay, K. Bae, B. Movsas, et al., Higher biologically effective dose of radiotherapy is associated with improved outcomes for locally advanced non-small cell lung carcinoma treated with chemoradiation: An analysis of the radiation therapy oncology group, International Journal of Radiation Oncology*Biology*Physics 82 (1) (2012) 425 – 434.

[4] P. Lambin, R. G. P. M. van Stiphout, M. H. W. Starmans, et al., Predicting outcomes in radiation oncology[mdash]multifactorial decision support systems, Nat Rev Clin Oncol 10 (1) (2013) 27–40.

[5] L. B. Marks, S. M. Bentzen, J. O. Deasy, et al., Radiation dose-volume effects in the lung, International Journal of Radiation Oncology*Biology*Physics 76 (3, Supplement 1) (2010) S70 – S76.

[6] I. Turesson, J. Nyman, E. Holmberg, A. Odn, Prognostic factors for acute and late skin reactions in radiotheraphy patients, International Journal of Radiation Oncology*Biology*Physics 36 (5) (1996) 1065 – 1075.

[7] K. Fleckenstein, B. Gauter-Fleckenstein, I. L. Jackson, Z. Rabbani, M. Anscher, Z. Vujaskovic, Using biological markers to predict risk of radiation injury, Seminars in Radiation Oncology 17 (2) (2007) 89 – 98.

[8] M. S. Anscher, L. B. Marks, T. D. Shafman, et al., Using plasma transforming growth factor beta-1 during radiotherapy to select patients for dose escalation, Journal of Clinical Oncology 19 (17) (2001) 3758–3765.

[9] L. Zhao, K. Sheldon, M. Chen, et al., The predictive role of plasma tgf-β during radiation therapy for radiation-induced lung toxicity deserves further study in patients with non-small cell lung cancer, Lung Cancer 59 (2) (2008) 232 – 239.

[10] Y. Chen, P. Rubin, J. Williams, E. Hernady, T. Smudzin, P. Okunieff, Circulating IL-6 as a predictor of radiation pneumonitis, International Journal of Radiation Oncology*Biology*Physics 49 (3) (2001) 641 – 648.

[11] J. H. Oh, J. M. Craft, R. Townsend, J. O. Deasy, J. D. Bradley, I. El Naqa, A bioinformatics approach for biomarker identification in radiation-induced lung inflammation from limited proteomics data, Journal of Proteome Research 10 (3) (2011) 1406–1415.

[12] L. Zhao, L. Wang, W. Ji, et al., Association between plasma angiotensin-converting enzyme level and radiation pneumonitis, Cytokine 37 (1) (2007) 71 – 75.

[13] S. N. Ghosh, R. Zhang, B. L. Fish, et al., Renin-angiotensin system suppression mitigates experimental radiation pneumonitis, International Journal of Radiation Oncology*Biology*Physics 75 (5) (2009) 1528–1536.

[14] I. E. Naqa, J. Bradley, A. I. Blanco, et al., Multivariable modeling of radiotherapy outcomes, including dose volume and clinical factors, International Journal of Radiation Oncology*Biology*Physics 64 (4) (2006) 1275 – 1286.

[15] S. J. Spencer, D. A. Bonnin, J. O. Deasy, J. D. Bradley, I. El Naqa, Bioinformatics methods for learning radiation-induced lung inflammation from heterogeneous retrospective and prospective data., Journal of Biomedicine and Biotechnology (2009) 1 – 14.

[16] N. Issam El, P. Piotr, S. Jan, Monte carlo role in radiobiological modelling of radiotherapy outcomes, Physics in Medicine and Biology 57 (11) (2012) R75.

[17] D. M. Chickering, D. Heckerman, C. Meek, Large-sample learning of bayesian networks is np-hard, J. Mach. Learn. Res. 5 (2004) 1287–1330.

[18] N. Friedman, D. Geiger, M. Goldszmidt, Bayesian network classifiers, Mach. Learn. 29 (2-3) (1997) 131–163.

[19] J. A. Hoeting, D. Madigan, A. E. Raftery, C. T. Volinsky, Bayesian model averaging: a tutorial, Statistical Science 14 (1999) 382–417.

[20] D. Madigan, J. York, D. Allard, Bayesian Graphical Models for Discrete Data, International Statistical Review 63 (2) (1995) 215–232.

[21] S. K. Das, S. Chen, J. O. Deasy, S. Zhou, F.-F. Yin, L. B. Marks, Combining multiple models to generate consensus: application to radiation-induced pneumonitis prediction, Medical Physics 35 (11) (2008) 5098–109.

[22] W. P. Smith, J. Doctor, J. Meyer, I. J. Kalet, M. H. Phillips, A decision aid for intensity-modulated radiation-therapy plan selection in prostate cancer based on a prognostic bayesian network and a markov model, Artificial Intelligence in

Medicine 46 (2) (2009) 119 – 130.

[23] K. Jayasurya, G. Fung, S. Yu, et al., Comparison of bayesian network and support vector machine models for two-year survival prediction in lung cancer patients treated with radiotherapy, Medical Physics 37 (4) (2010) 1401–1407.

[24] J. H. Oh, J. Craft, R. A. Lozi, et al., A bayesian network approach for modeling local failure in lung cancer, Physics in Medicine and Biology 56 (6) (2011) 1635.

[25] A. J. Hope, P. E. Lindsay, I. E. Naqa, et al., Modeling radiation pneumonitis risk with clinical, dosimetric, and spatial parameters, International Journal of Radiation Oncology*Biology*Physics 65 (1) (2006) 112 – 124.

[26] G. Gan, C. Ma, J. Wu, Data clustering - theory, algorithms, and applications., SIAM, 2007.

[27] D. Koller, M. Sahami, Toward optimal feature selection, in: In 13th International Conference on Machine Learning, 1995, pp. 284–292.

[28] K. P. Murphy, The bayes net toolbox for matlab, Computing Science and Statistics 33 (2001) 2001.

[29] N. Friedman, M. Goldszmidt, A. Wyner, Data analysis with bayesian networks: A bootstrap approach, in: Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence, UAI'99, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1999, pp. 196–205.

[30] D. Heckerman, C. Meek, G. Cooper, A bayesian approach to causal discovery, Tech. rep., Microsoft Research (1997).

[31] B. Efron, R. Tibshirani, Improvements on cross-validation: The 632+ bootstrap method, Journal of the American Statistical Association 92 (438) (1997) 548–560.

[32] S. L. Kwa, J. V. Lebesque, J. C. Theuws, et al., Radiation pneumonitis as a function of mean lung dose: an analysis of pooled data of 540 patients, International Journal of Radiation Oncology*Biology*Physics 42 (1) (1998) 1 – 9.

[33] E. X. Huang, A. J. Hope, P. E. Lindsay, et al., Heart irradiation as a risk factor for radiation pneumonitis, Acta Oncologica 50 (1) (2011) 51–60.

[34] F.-M. Kong, J. A. Hayman, K. A. Griffith, et al., Final toxicity results of a radiation-dose escalation study in patients with non-small-cell lung cancer (NSCLC): Predictors for radiation pneumonitis and fibrosis, International Journal of Radiation Oncology*Biology*Physics 65 (4) (2006) 1075 – 1086.

[35] B. Parashar, A. Edwards, R. Mehta, et al., Chemotherapy significantly increases the risk of radiation pneumonitis in radiation therapy of advanced lung cancer, Am J Clin Oncol 34 (2) (2011) 160–4.

[36] M. Scutari, On the prior and posterior distributions used in graphical modelling, Bayesian Analysis 8 (3) (2013) 505–532.

[37] G. J. Kutcher, C. Burman, Calculation of complication probability factors for non-uniform normal tissue irradiation: the effective volume method, International Journal of Radiation Oncology*Biology*Physics 16 (6) (1989) 1623–1630.

[38] J. D. Bradley, A. Hope, I. E. Naqa, et al., A nomogram to predict radiation pneumonitis, derived from a combined analysis of RTOG 9311 and institutional data,

International Journal of Radiation Oncology*Biology*Physics 69 (4) (2007) 985 – 992.

[39] A. Niemierko, Reporting and analyzing dose distributions: A concept of equivalent uniform dose, Medical Physics 24 (1) (1997) 103–110.

[40] D. Koller, N. Friedman, Probabilistic Graphical Models: Principles and Techniques - Adaptive Computation and Machine Learning, The MIT Press, 2009.

[41] B. Sahiner, H. P. Chan, L. Hadjiiski, Classifier performance prediction for computer-aided diagnosis using a limited dataset, Med Phys 35 (4) (2008) 1559–70.

[42] N. Friedman, D. Koller, Being bayesian about network structure – bayesian approach to structure discovery in bayesian networks, in: Machine Learning, 2000, pp. 201–210.

[43] A. Y. Ng, M. I. Jordan, On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes, in: T. G. Dietterich, S. Becker, Z. Ghahramani (Eds.), Advances in Neural Information Processing Systems 14, MIT Press, 2002, pp. 841–848.

CHAPTER 8
Modeling of radiation pneumonitis after lung stereotactic body radiotherapy: a Bayesian network approach

Contents

8.1 Introduction

Chapter 7 elaborated on the methodology of applying the Bayesian network framework to the problem of radiation pneumonitis. The study was however limited to conventionally fractionated radiotherapy. It is expected that dosimetric and biological relationships with toxicity are different for SBRT regimen (see section 1.4, 2.3.1). It would be of great clinical interest to observe how the Bayesian network descriptions of RP depend on the fractionation schemes. This chapter intends to identify important dosimetric, clinical and biological RP risk factors after lung SBRT and test the feasibility of the BN method as a prediction tool that integrates the identified factors.

The study presented in this chapter was submitted in March 2016 as the following paper: "Modeling of radiation pneumonitis after lung stereotactic body radiotherapy: a Bayesian network approach" authored by: Lee, S., Ybarra, N., Jeyaseelan, K., Faria, S., Kopek, N., Brisebois, P., Vu, T., Filion, E., Campeau, M.-P., Lambert, L., Del Vecchio, P., Trudel, D., El-Sokhn, N., Roach, M., Robinson, C., and El Naqa, I.

8.2 Background

In recent years, stereotactic body radiotherapy (SBRT) has become a standard choice of a radiotherapy technique for non-operable early stage non small cell lung cancer (NSCLC), demonstrating a local control rate close to 90% [1]. Incidence of

pulmonary toxicity, usually defined as symptomatic radiation pneumonitis (RP), is reported to be less than 10% [2] due to focused radiation to a small target which spares large volume of healthy lung [3]. However, several non-dosimetric factors reportedly increase or decrease the RP risk, such as central tumor location [4], baseline interstitial pneumonitis [5] and chronic obstructive pulmonary disease (COPD) [6]. Ignoring these factors could underestimate RP risk for certain patients. Thus, there is a clinical need to augment dosimetric RP models with patient-specific biological and clinical risk modifiers for more patient-specific prediction.

We propose Bayesian network (BN) as a multivariate modeling platform to accommodate such high-dimensional data. BN can be characterized as graphical representation of relationships between input variables called a directed acyclic graph (DAG). Variables in a DAG are connected along the direction of influence. Compared to conventional multivariate modeling methods such as logistic regression, BN has a number of advantages; First, BN can be designed by combining domain knowledge and observed data. This concept was applied to radiotherapy outcome prediction in prostate cancer [7] and radiation pneumonitis [8]. Second, its probabilistic characteristics allows prediction with incomplete data without relying on imputation to fill them in. Predictive benefit of this feature was demonstrated by Jarasurya et al. [9] in prediction of lung cancer outcome. Finally, from users' perspective, its graphical feature provides more transparency in its inference paths.

Radiation pneumonitis was modeled using the BN concept for conventional fractionation [8] where finding a consensus of prediction results from several BN models

(ensemble approach) was shown to improve RP prediction. However, the BN approach hasn't been applied to SBRT cases where dose-volumetric relationships and biological damage mechanisms haven't been as well established. The aim of this study is to develop a Bayesian Network RP model from an SBRT cohort. While a primary objective is to assess its predictive potential, we will also address its ability to generate new hypotheses.

8.3 Methods

8.3.1 Patient cohort

Forty three stage I and II NSCLC patients were prospectively recruited for this study from the following 3 institutions upon approval of respective institutional review boards: McGill University Health Centre (MUHC), Centre Hospitalier de l'Université de Montréal (CHUM), and Washington University in St. Louis (WashU), 32 patients from MUHC and CHUM formed a training cohort for BN modeling. 11 patients from WashU was reserved for model validation. Every patient met the following eligibility criteria: 1) received SBRT of equal or less than 5 fractions with curative intent, 2) no history of previous lung irradiation, and 3) baseline Karnofsky performance status ≥ 70. Detailed cohort characteristics are shown in table 8.1. The patients were treated with radiotherapy (RT) without any adjuvant therapy. Depending on institutions, three different delivery techniques were used: 3-dimensional (3D) conformal radiotherapy, RapidArc^{TM}(Varian Medical Systems, Palo Alto, CA) Volumetric Arc Therapy (VMAT), and CyberKnife (Accuray Inc, Sunnyvale, CA). Detailed RT procedures are summarized in the table 8.2 and 8.3.

Table 8.1: Characteristics of the training and validation cohorts. *Calculated for whole lung subtracted by planning target volume and converted to equivalent dose in 2 Gy fraction.

	Patient count (%)	
	Training	Validation
Cohort size	32	11
Tumor stage		
I	32 (100)	9 (81)
II	0 (0)	2 (19)
RP grades		
0	17 (53)	0 (0)
1	11 (34)	10 (91)
2	2 (6)	0 (0)
3	2 (6)	1 (9)
≥ 4	0 (0)	0 (0)
≥ 2	4 (13)	1 (9)
Mean lung dose*		
median	4.9	6.3
range	2.4-10.9	1.2-9.9
RT modality		
3D conformal	19 (59)	11 (100)
VMAT	5 (16)	0 (0)
CyberKnife	8 (25)	0 (0)
RT prescription		
60 Gy in 3 fractions	8 (25)	0 (0)
60 Gy in 5 fractions	5 (16)	1 (9)
50 Gy in 5 fractions	4 (13)	4 (36)
48 Gy in 3 fractions	12 (38)	0 (0)
34 Gy in 1 fractions	3 (9)	0 (0)
54 Gy in 3 fractions	0 (0)	5 (45)
55 Gy in 5 fractions	0 (0)	1 (9)

Table 8.2: Detailed radiotherapy procedures used for the training cohort. GTV: gross tumour volume, ITV: internal target volume, IGTV: internal gross tumour volume, PTV: planning target volume, Tx: prescription dose, 4DCT: 4-dimensional computed tomography, IGRT: image-guided radiotherapy, fx: fraction, MU: monitoring unit.

Institution	MUHC	CHUM	
Technique	3D-CRT	VMAT	CyberKnife
Dose prescription	Dose normalized to 100 % at Tx, 95% of PTV receives Tx or higher (D95% $\geq$ Tx)	Dose normalized to 100 % at Tx which covers 95% or more of the PTV	
Dose planning procedure/ calculation algorithm	Forward planning using Eclipse (Varian, USA)/ superposition-convolution algorithm with heterogeneity correction	Inverse planning with RapidArc (Varian, USA)/ superposition-convolution algorithm with heterogeneity correction	Inverse planning with Multiplan (Accuray, USA) / Monte Carlo calculation
Beam type	6 MV photon	6 MV photon	6 MV photon
Target volume definition	ITV: drawn from 4DCT using maximum intensity projection PTV: ITV + 5 mm margin	IGTV: drawn on extreme phases of 4DCT to represent its full extent PTV: IGTV + 5 mm margin	GTV: drawn on breath hold, corrected if needed for deformation/rotation using extreme phases PTV: GTV + 5 mm margin
Dose fractionation	50 Gy in 5 fx: tumor at central location and/or close to critical organs (chest wall/large vessels/spinal cord) 34 Gy in 1 fx: otherwise, upon patients' request for shorter treatment 48 Gy in 3 fx: otherwise	50 Gy in 5 fx: tumour at central location 60 Gy in 5 fx: peripheral tumour close to OARs 60 Gy in 3 fx: otherwise	
Dose constraints to OARs	50 Gy in 5 fx: RTOG 0915 48 Gy in 3 fx: RTOG 0915 34 Gy in 1 fx: RTOG 0813	Timmerman et al. [10]	
Immobilization	BodyFix (Elekta Oncology, Norcross, GA)	BodyFix (Elekta Oncology)	Vac-Lok (Civco Medical Solutions, Orange City, IA)
IGRT	CBCT at every fraction	Pre- and mid-treatment CBCT at every fraction	Real-time target tracking
Plan verification	Independent MU check	Independent MU check, daily dynalog verification	Independent MU check

Table 8.3: Detailed radiotherapy procedures used for the validation SBRT cohort.

Institution	WashU
Technique	3D-CRT
Dose prescription	Dose generally prescribed to 80% isodose line (range 60-90%) and covers ¿95% of PTV
Dose planning procedure/ calculation algorithm	Forward planning with 7-11 non-coplanar beams using Pinnacle (Philips, Netherlands)/ superposition-convolution algorithm with heterogeneity corrections
Beam type	6 MV photons
Target volume definition	ITV: drawn from 4DCT using maximum intensity projection PTV: ITV + 5 mm margin
Dose fractionation	50-60 Gy in 5 fx: central location or close to critical organs 54 Gy in 3 fx: all others
Dose constraints to OARs	50-60 Gy in 5 fx: RTOG 0813 54 Gy in 3 fx: RTOG 0618
Immobilization	Abdominal compression (CDR systems, Canada)
IGRT	CBCT at every fraction with KV fluoroscopy
Plan verification	Independent MU check

8.3.2 Data collection

Blood samples from the patients were first acquired on the CT simulation day as a baseline and 6 weeks post-treatment. Enzyme-linked immunosorbent assay (ELISA) was used for measuring biomarker concentrations in the samples. Incidence rate of symptomatic RP, classified as Common Toxicity Criteria for adverse events (CTCAE) toxicity (version 4) grade 2 or higher, was 13% (4/32) from the training and 9% (1/11) from the validation cohorts. Median follow-up was 12 months for training and 34 months for the validation cohort.

8.3.3 Candidate variables

Candidate variables for the BN pneumonitis model were chosen from 3 main categories: biological, dosimetric and clinical variables. Candidate biological variables consisted of serum concentration of interleukin(IL)-6, IL-8, angiotensin converting enzyme (ACE), alpha-2-macroglobulin (α2M) , and transforming growth factor (TGF)-β1 and plasma concentration of osteopontin (OPN). As summarized in [11], these markers represent different biological processes involved in pathogenesis of radiation-induced lung injury, such as pro-(IL-6 [12], OPN [13]) and anti-(IL-8 [14]) inflammatory reactions, fibrogenesis (TGFβ [15]), vascular damage (ACE [16]) and modulation of inflammatory reactions (α2M [17]). 12 features in total were extracted (6 markers x 2 time points) from the biomarker data. The biomarker features at 6-weeks were taken as a percentage difference from the respective baseline level. The following 7 clinical RP risk factors were chosen by literature survey: superoinferior PTV location (PTVCOMSI) [18], age [19], smoking status [20], COPD [6], ACE inhibitor [21], baseline interstitial lung disease [5], and centrally located tumours [4]. Dosimetric factors were derived from planned dose converted to equivalent dose in 2 Gy fraction (EQD2) using an alpha-beta ratio of 4 Gy for lung [22] and 2 Gy for heart [23]. For lung dose calculation, PTV was subtracted from contoured lung. Mean lung dose (MLD) and various Vx values (lung volume receiving $>$ x Gy) for ipsilateral and whole lung were considered. Due to high correlation between these parameters [2], exploratory analysis was performed to find smaller number of features that capture dose heterogeneity relevant to RP. In this analysis, Vx was computed at various threshold dose x in three different ways: 1) x as an absolute dose or relative

to a prescription dose, 2) Vx normalized to lung volume or as a absolute volume, and 3) ipsilateral or whole lung. In addition to lung dose, we also considered mean heart dose (MHD) [24], fraction size [25], and PTV volume [26].

8.3.4 Bayesian Network training

A Bayesian network ensemble model was trained from the candidate variables following the methods in [8]. Computation was performed using the MATLAB-based code package released in: `https://github.com/meson200/MCMCBNE`. In brief, the training was done in 4 steps:

1. Data discretization: every continuous variable was discretized into 2 bins at a boundary that maximizes mutual information with respect to RP, as shown in table 8.4.
2. Feature selection with the Koller-Sahami (KS) filter: The number of candidate variables were reduced to the smallest subset that maximized explanatory power measured by cross-entropy with respect to RP. This particular filtering technique was chosen due to its information theoretic approach, which is independent of prediction schemes [27].
3. DAG training: Posterior distribution of Bayesian network graphs was obtained by Markov-Chain Monte Carlo (MCMC) sampling under causality constraints between variables.
4. Parameter learning: Every variable in a BN is treated as a probabilistic distribution which is conditioned upon its upstream variables ("parents"). BN parameters, referred to as conditional probability values for every pair of a node

and its parents, were learned from data using the expectation-maximization algorithm.

5. RP prediction: Probability of RP is computed by each fully trained BN graph using known input variables. Then, a composite (ensemble) prediction was made P(RP) by averaging the results from BN graphs ($P_i(RP)$) with N highest posterior probability from the MCMC simulation (equation 8.1). Prediction from individual BN models in the ensemble was weighted by the posterior probability of the BN graph ($P(i|D)$). Ensemble size (N) was varied from 1 to 200.

$$P(RP) = \frac{\sum_{i=1}^{N} P_i(RP)P(i|D)}{\sum_{i=1}^{N} P(i|D)} \tag{8.1}$$

8.3.5 Logistic regression training

As an alternative prediction model, multivariate logistic regression function was trained with the same variable set used for Bayesian network. Probability of pneumonitis was derived using the logistic models from variable values $\{x_i\}$ which were standardized prior to fitting using means (m_i) and standard deviations (σ_i)in the training set (equation 8.2, 8.3)

$$p(RP) = \frac{1}{1 + exp(-a)} \tag{8.2}$$

$$a = w_0 + \sum_{i}^{N} w_i \frac{(x_i - m_i)}{\sigma_i} \tag{8.3}$$

The model coefficients $\{w_0, ..., w_N\}$ were trained under the L_1 regularization method which effectively decreases the magnitude of the coefficients to mitigate overfitting. Missing data in the dataset, taking 0% in the training and 6% in the validation set, was imputed using the K nearest neighbor method.

8.3.6 Evaluation of predictive performance

Classification of RP events was made by thresholding on the RP probabilities from BN and logistic regression models. Classification performance was measured using three receiver operating characteristics (ROC) metrics: area under the curve (AUC), sensitivity, and specificity at the optimal operating threshold maximizing a sum of sensitivity and specificity. Model testing was carried out in two ways:

- Internal validation using the .632+ bootstrap method [28]: the training was repeated in 200 replicates which was resampled from the original data with replacement, and the instances that were not sampled into the replicates were used for testing. At this stage, confidence levels on the connections between variables in the BN graphs were defined as the frequency of its appearance throughout the bootstrapping.
- External validation: The BN structure and parameters were trained on the original training dataset and applied to the external validation dataset. We also investigated the impact of uncertainty in the BN parameter values on validation performance. This was implemented by repeating parameter learning in the 200 bootstrap replicates under the same structure (see supplementary material) and testing a bag of BN models with varying parameters on the validation set. The logistic regression model was validated in the same fashion.

95% confidence intervals for all the performance metrics were evaluated by taking 97.5% and 2.5% percentile of the bootstrap samples. Comparisons of performance between different models were made using the paired t-test on the bootstrap results at the 95% confidence level.

8.4 Results

8.4.1 Exploratory analysis on lung DVH parameters

Correlation between lung Vx and RP was examined by the change in odds ratios, computed by a logistic fit on raw data, at various threshold dose (x) (figure 8.1). When x was used as absolute dose, highest odds ratio (5.685) was marked at 5 Gy for ipsilateral lung. When the percentage of a prescription dose was used as x, increase in correlation was observed in a high dose region beyond 50% of the prescription dose. Guided by this analysis, we chose two Vx parameters that represent low dose and high dose spillage respectively: percentage of ipsilateral lung volume receiving 5 Gy or more (V5) and absolute lung volume receiving more than 105% of prescription dose (V105%). In a similar fashion, ipsilateral MLD (odds ratio: 2.400) was preferred over MLD for the whole lung (odds ratio: 2.365).

8.4.2 Variable selection and the Bayesian Network ensemble model

The KS variable filter was applied to 25 candidate variables from 3 categories (dosimetric, biological, and clinical) which was reduced to the following 6: 1) pre-treatment OPN, 2) 6 weeks ACE, 3) pre-treatment TGFβ, 4) ipsilateral V5, 5) V105%, and 6) PTVCOMSI. The KS filtering results are summarized in table 8.4. Causal connections between these variables and RP were established in BN graphs. Bootstrap test on BN graph learning detected 11 significant links out of possible 19

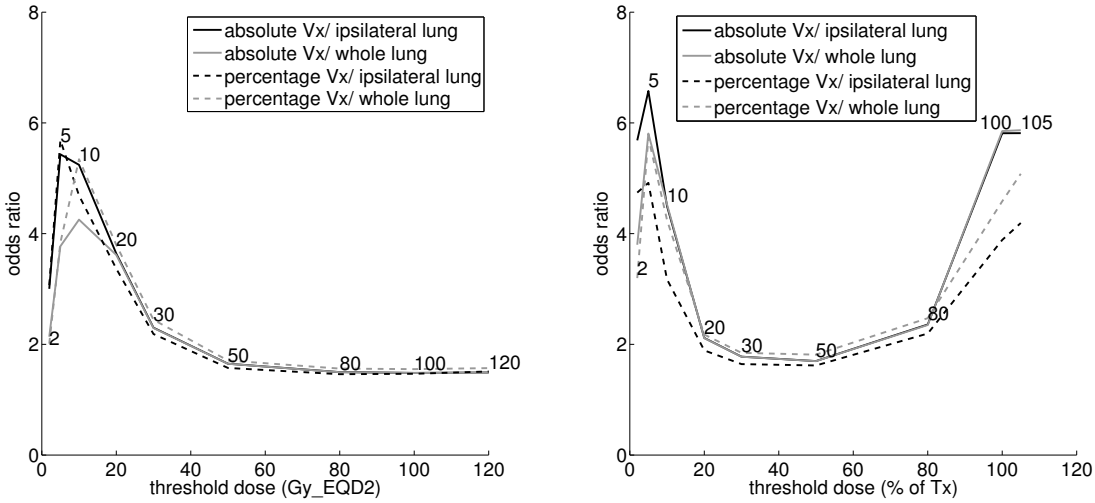

Figure 8.1: Odds ratios of lung Vx measured at various threshold dose values (x), normalization schemes, and lung volume definition.

from an ensemble of 50 graphs where bootstrapped RP prediction performance was optimal (figure 8.3). A mean confidence level of the significant links was 0.57, while an upper bound of random variation was 0.29 [29].

8.4.3 Prediction performance of the BN model

When bootstrap validation was used, RP prediction improved upon increasing number of graphs in an ensemble (figure 8.3). AUC, sensitivity and specificity increased respectively from 0.96, 0.93, and 0.92 at size 1 to 0.99, 1, and 0.98 at size 50 where the performance reached optimum. At the optimal classification threshold, sensitivity was consistently higher than specificity. In external validation, AUC increased from 0.65 at size 1 to 0.8 at ensemble sizes 5-30 beyond which a slight decrease to 0.75 was observed. The BN model was subsequently tested using only the information available at baseline i.e. without ACE at 6 weeks. As a result, AUC

Table 8.4: Odds ratios of candidate variables, bin boundary used for discretization, and frequency of selection obtained by bootstrapping the KS variable filtering. P-values were adjusted for multiple comparison using a method by Benjamini and Hochberg [30]. *variables selected for the BN modeling stage. † taken as a percentage change from baseline.

	Odds ratio (p-value)	Bin boundary	Selection frequency
Biological variables			
OPN (baseline)*	0.887 (0.886)	54.2 ng/ml	0.394
OPN (6 weeks†)	1.150 (0.886)	80.9 %	0.133
IL8 (baseline)	2.862 (0.210)	31.0 pg/ml	0.228
IL8 (6 weeks)	0.404 (0.637)	-60.4 %	0.264
ACE (baseline)	1.999 (0.529)	141.1 ng/ml	0.308
ACE (6 weeks)*	0.002 (0.010)	-15.8 %	0.782
IL6 (baseline)	0.070 (0.657)	7.0 pg/ml	0.2
IL6 (6 weeks)	1.106 (0.886)	-7.0 (%)	0.058
a2M (baseline)	0.553 (0.638)	5.3 mg/ml	0.328
a2M (6 weeks)	0.848 (0.886)	-7.6 %	0.142
TGFb (baseline)*	1.866 (0.540)	42.2 ng/ml	0.504
TGFb (6 weeks)	0.493 (0.610)	1.4 %	0.053
Dosimetric variables			
MLD (ipsilateral)	2.400 (0.391)	13.8 Gy	0.107
V5 (ipsilateral)*	5.685 (0.060)	42.4 %	0.454
V105%*	5.848 (0.023)	1.4 cc	0.668
Fraction size	0.752 (0.886)	20 Gy per fraction	0.142
PTV volume	1.932 (0.518)	20.5 cc	0.064
MHD	1.945 (0.529)	9.0 Gy	0.153
Clinical variables			
PTVCOMSI*	0.379 (0.391)	0.5	0.448
Age	1.172 (0.886)	69	0.121
Smoking	1.077 (0.945)		0.146
IP	1.300 (0.768)		0.061
Central tumour	1.800 (0.854)		0.068
COPD	0.750 (0.886)		0.120
ACE inhibitor	0.800 (0.638)		0.054

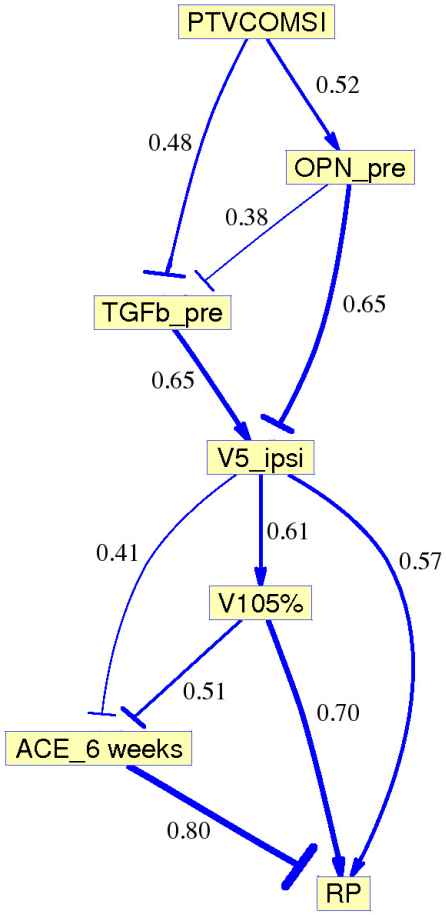

Figure 8.2: Variables connected by significant associations detected in an ensemble of 50 graphs. Edge thickness was drawn proportionally to bootstrap estimated confidence level. Arrow-headed and bar-headed edges are assigned to positive and negative correlation, respectively. ipsi: ipsilateral lung, _pre: baseline biomarker levels

Table 8.5: Parameters for the trained logistic regression model as used in equation 8.3. bl.: baseline, ipsi.: ipsilateral.

	Variables						
	OPN (bl.)	ACE (6 weeks)	TGF (bl.)	V5 (ipsi.)	V105%	COMSI	w_0
mean (m_i)	110 ng/ml	3.23 %	41.2 ng/ml	27.4%	1.34 cc	0.59	
standard deviation (σ_i)	55.9 ng/ml	29.5 %	11.7 ng/ml	12.1 %	2.00 cc	0.20	
coefficients (w_i)	0	-0.044	0.014	0.004	0.148	-0.040	0.125

Table 8.6: Comparison of the Bayesian network model with size 5 (BN5) and logistic regression in predictive performance. Numbers in parentheses are 95% confidence intervals from bootstrapping.

	632+ bootstrap			External validation		
	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity
BN5	0.99 (0.90,1)	1 (1,1)	0.97 (0.76,1)	0.8 (0.35,0.85)	1 (1,1)	0.7 (0.3,0.8)
logistic	0.93 (0.62,1)	0.99 (0.75,1)	0.87 (0.59,1)	0.7 (0,0.9)	1 (0,1)	0.7 (0.1,1)

and sensitivity from the 632+ bootstrap test decreased significantly at all ensemble sizes (figure 8.3). Reduction in AUC ranged from 0.02 at largest ensemble size to 0.06 at ensemble size 1. In the validation cohort, however, better performance was observed with only baseline information.

The ensemble size 5 BN model (see supplementary material for specification), the smallest model that performed best in external validation, was compared with the multivariate logistic model. The best-fit parameters for the logistic model are shown in table 8.5. The logistic model was less effective in predicting RP than Bayesian network, recording significantly lower AUC both in internal 632+ and external validation (table 8.6).

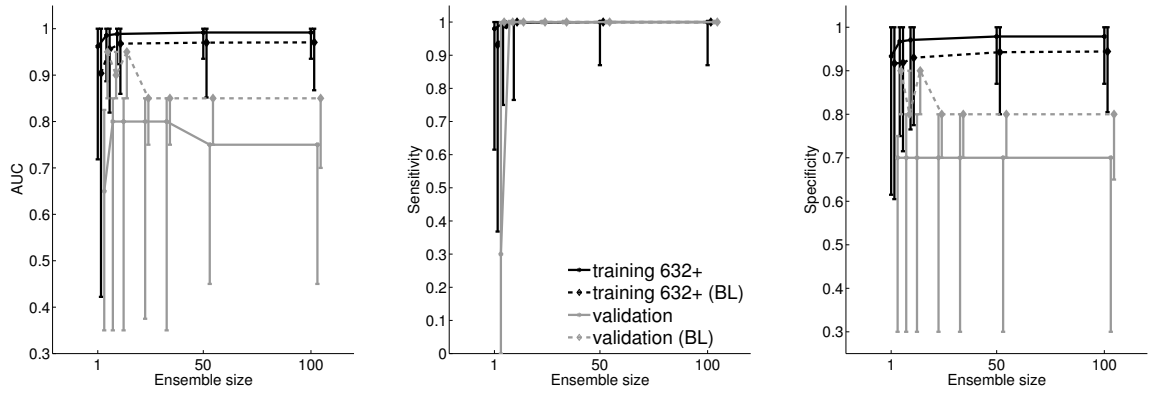

Figure 8.3: Classification performance of the Bayesian network model in two cohorts with varying ensemble size. Error bars: bootstrap-estimated 95% confidence intervals.

8.5 Discussion

Events of RP from lung SBRT are rare and identifying the susceptible patients before radiotherapy remains a difficult task, with conflicting results between studies. This study intended to objectively select and combine RP risk factors into a Bayesian network and test its predictive potentials. Two factors account for good bootstrap performance of the resulting model in the training cohort. First, the main driving force was strong individual predictive power of the key variables in the model. Univariate AUC values of ACE at 6 weeks, V5, and V105%, 3 variables connected to RP with high confidence, was respectively 0.94, 0.85 and 0.96 in the training cohort. Another factor was the use of an ensemble instead of a single model; Learning a Bayesian network structure from limited data involves high degree of uncertainty due to large number of permutations in connecting the nodes [31]. Bayesian model averaging helps overcoming the model selection uncertainty and thus improves out-of-sample performance [32]. In our results, better performance of an ensemble model

with a size larger than 1 was seen both in training and validation cohorts. Predictive benefit of the ensemble approach was already shown by other outcome studies [8] [33].

Amongst the biomarkers we studied, we found lower concentration of ACE at 6 weeks was the most strongly associated with RP events. This result is in line with investigation by Zhao et al. [15] who reported lower ACE level at baseline and mid-treatment for patients with RP grade ≥ 2. In the Bayesian network ensemble the ACE was connected to dosimetric variables with high confidence. Causality of this relationship can be justified by the knowledge that main production site of ACE is in lung epithelium and external stress to vasculature such as ionizing radiation or bleomycin decreases serum ACE [16]. We also gathered information on use of ACE inhibitors at baseline but it was not a significant predictor of RP ($p = 0.64$), suggesting direct measurement of ACE as a more sensitive way to predict RP.

Choice of 6 weeks as a time point to gauge post-treatment biomarker response was adequate to predict late toxicity before it happened, as the earliest occurrence of RP was 94 days post-RT. However, biomarker response to RT would not be observable at a treatment planning stage where RP prediction is most relevant in clinical decision making. We tested this scenario by attempting prediction without 6-weeks ACE. BN is capable of handling missing information by marginalizing probability distribution over unknown variables. Prediction with less information was less accurate in the 632+ bootstrap test. However, this was not clearly reproduced in our external validation cohort where the ACE decrease at 6 weeks was not as specific to RP as in the training set, although one patient with RP showed 10% decrease in

ACE. Nevertheless, our BN model is still capable of making baseline prediction due to its probabilistic property, which is absent in the multivariate logistic model. Furthermore, the BN model with post-RT response biomarkers might still be valuable as a semi-mechanistic model that elucidates mechanisms to RP.

We also observed that the size of high dose spillage, represented by lung volume outside PTV receiving dose $\geq 105\%$, was predictive of RP in univariate analysis and also one of the key variables in the BN model. This "high dose effect" on RP has been previously reported by a number of studies [18] [34]. Our results on exploratory analysis on Vx point out that both low-dose (V5) and high dose components might be relevant to RP. Previous lung SBRT protocols including RTOG 0236 and 0813 stipulate this volume as one of the quality assurance metrics to be regulated, setting its upper limit on 15% of the PTV volume. Further studies are needed to clarify the mechanism of a smaller volume of high dose irradiation to lung causing RP.

Multivariate logistic regression is a popular method in predictive modeling and has been applied to prediction of radiation pneumonitis [18] [35]. In our cohorts, it was less successful in predicting RP than the BN model given the same set of variables. In addition, logistic regression recorded higher variance (sensitivity of prediction performance to variation in training example); The variance was measured by perturbing training data by bootstrapping, repeating the model training in the perturbed training data, and taking the distribution of the resulting AUC values in the external validation set. Figure 8.4 shows that although both models showed considerable variance, contributing to large confidence intervals in figure 8.3, the Bayesian network approach managed to reduce the sensitivity. This result advocates

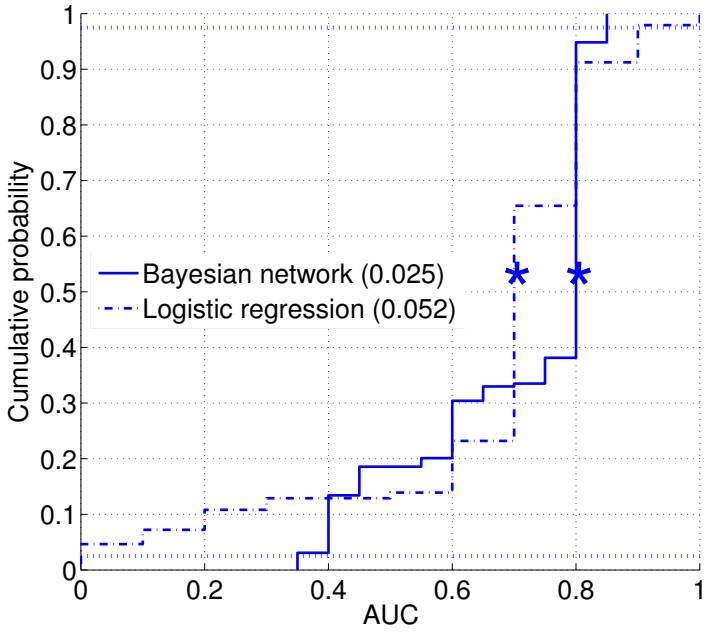

Figure 8.4: Variance of the BN and logistic models represented by cumulative distribution of the areas under the curve for external validation resulting from random training dataset change. Numbers in parentheses are variance of a model defined as mean squared deviation from the unperturbed AUC (marked as asterisks). Two horizontal lines represent 97.5% and 2.5% quantiles.

BN as a more robust choice for a multi-dimensional data model, while more aggressive feature selection might be needed prior to training logistic regression.

The main limitation of this study is the low number of toxicity events in the cohorts, which led to relatively low specificity of the optimized model. Nevertheless, our computational approach reduced the data dimensionality to the key variables in order to mitigate the impact of a low event rate on overfitting. In the future, the methodology could be extended to better handle class imbalance. For example, a model can be trained with different weights to misclassification of RP or non-RP

events (also known as cost-sensitive learning [36]). Also, the links that we discovered in the BN graphs do not always imply direct causality although non-causal relationships were ruled out during graph searching. Ultimately, further validation on larger datasets is required to confirm the observations learned from the developed BN model.

8.6 Conclusion

We developed a Bayesian Network ensemble model for radiation pneumonitis after lung SBRT. The process of building the model and the resulting model structure identified potential key players in predicting RP from SBRT, such as high dose spillage in lung and dose response of post-treatment ACE level. The model can be used as a prediction tool that can operate under varying availability of information.

8.7 Acknowledgements

We thank Dr. Jean-François Carrièr and Dr. Robert Doucet for providing us a part of clinical data and Dr. Jan Seuntjens for useful advices. The computational work was enabled in part by computer resources provided by WestGrid (www.westgrid.ca). This research was funded by the Canadian Institute of Health Research (CIHR) grant MOP-114910 and the CREATE Medical Physics Research Training Network grant (NSERC, Canada).

8.8 Supplementary materials

This supplementary material provides specifications on the structures and parameters of the best-performing BN model for any interested readers who wish to apply it to their cohorts. This ensemble model was chosen based on its performance

in external validation. The model consists of 5 graphs $\{g_1, ..., g_5\}$ from which probability of RP of grade 2 or higher is computed using the following Bayesian model averaging formula:

$$P(RP) = \frac{0.0033P_{g1}(RP) + 0.0027P_{g2}(RP) + 0.0027P_{g3}(RP) + 0.0023P_{g4}(RP) + 0.0023P_{g5}(RP)}{0.0133} \tag{8.4}$$

where $P_{g1}(RP), ..., P_{g5}(RP)$ are RP probabilities predicted by each individual graphs. Figure 1-5 shows the structures and parameters for the models $\{g_1, ..., g_5\}$. The parameters consist of prior probabilities and conditional probability tables for each node. Note that variables have to be discretized using the boundaries shown in the table 4 in the main text.

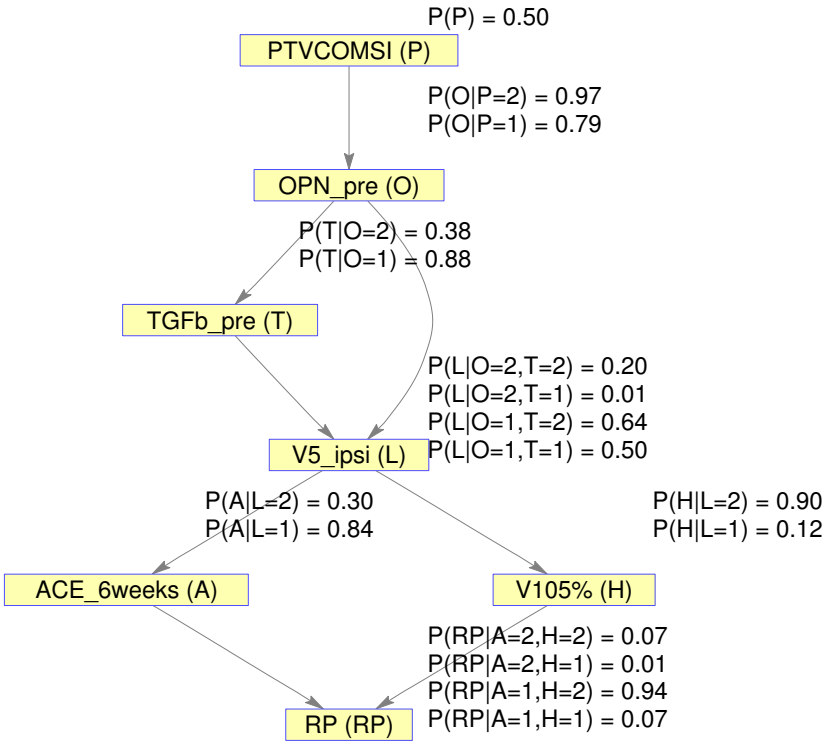

Figure 8.5: A structure and parameters for the model $g1$.

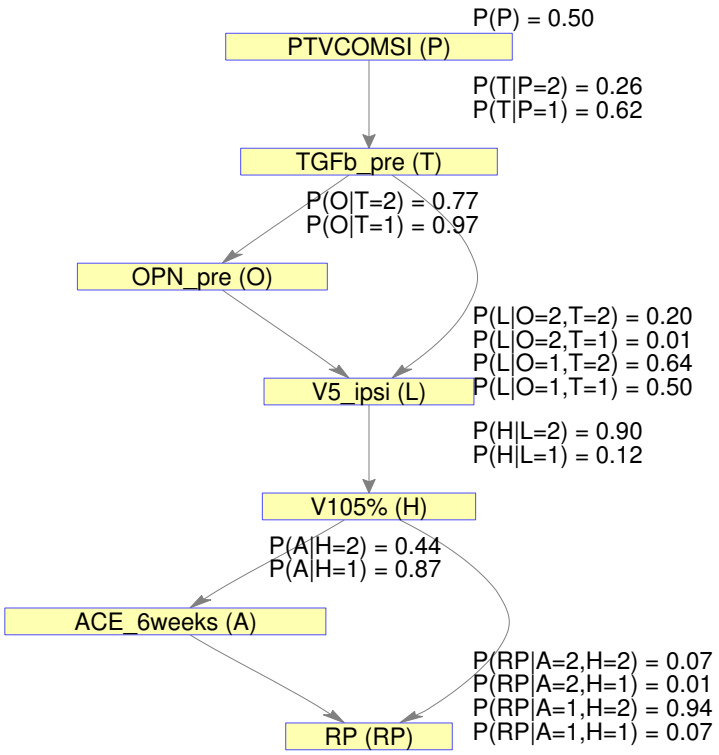

Figure 8.6: A structure and parameters for the model $g2$.

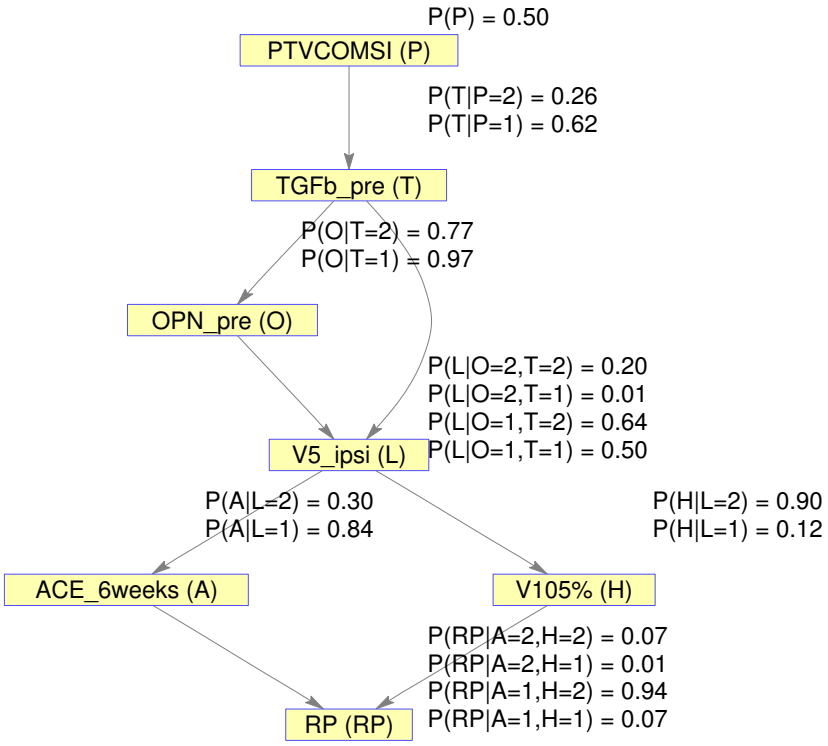

Figure 8.7: A structure and parameters for the model $g3$.

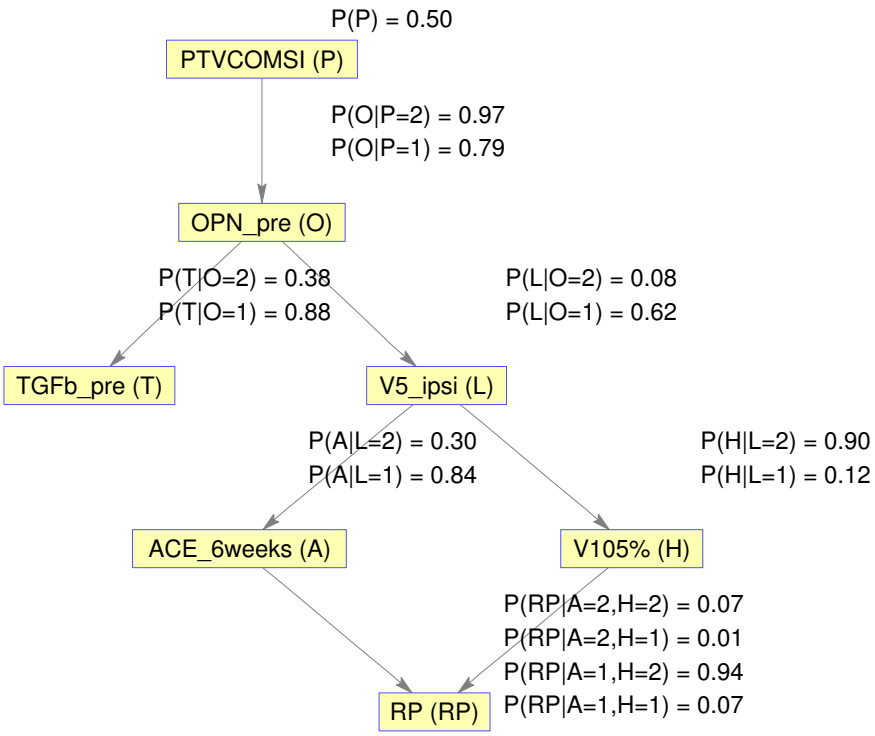

Figure 8.8: A structure and parameters for the model $g4$.

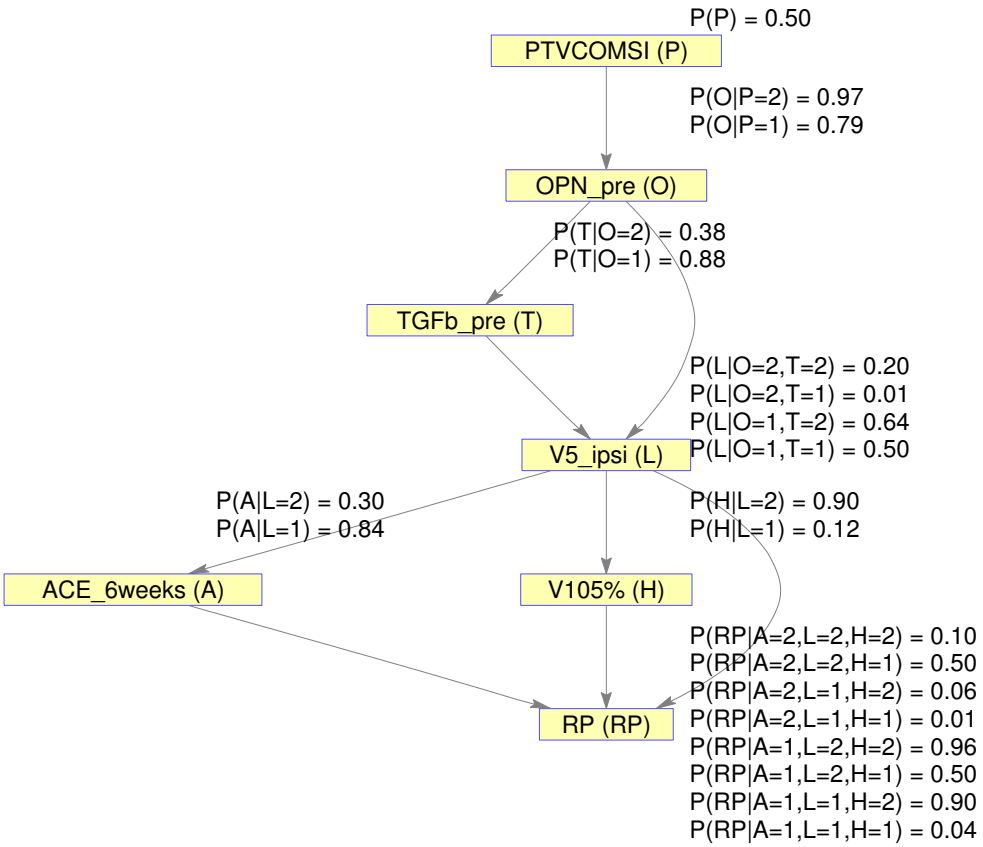

Figure 8.9: A structure and parameters for the model $g5$.

REFERENCES

[1] T. Xia, H. Li, Q. Sun, et al., Promising clinical outcome of stereotactic body radiation therapy for patients with inoperable stage I/II non-small-cell lung cancer, International Journal of Radiation Oncology*Biology*Physics 66 (1) (2006) 117 – 125.

[2] L. B. Marks, S. M. Bentzen, J. O. Deasy, et al., Radiation dose-volume effects in the lung, International Journal of Radiation Oncology*Biology*Physics 76 (3, Supplement 1) (2010) S70 – S76.

[3] H. Yamashita, W. Takahashi, A. Haga, K. Nakagawa, Radiation pneumonitis after stereotactic radiation therapy for lung cancer, World J Radiol 6 (9) (2014) 708–15.

[4] R. Timmerman, J. Galvin, J. Michalski, et al., Accreditation and quality assurance for radiation therapy oncology group: Multicenter clinical trials using stereotactic body radiation therapy in lung cancer, Acta Oncologica 45 (7) (2006) 779–786.

[5] H. Yamashita, S. Kobayashi-Shibata, A. Terahara, et al., Prescreening based on the presence of CT-scan abnormalities and biomarkers (KL-6 and SP-D) may reduce severe radiation pneumonitis after stereotactic radiotherapy, Radiat Oncol 5 (2010) 32.

[6] A. Takeda, E. Kunieda, T. Ohashi, et al., Severe copd is correlated with mild radiation pneumonitis following stereotactic body radiotherapy, Chest 141 (4) (2012) 858–66.

[7] W. P. Smith, J. Doctor, J. Meyer, I. J. Kalet, M. H. Phillips, A decision aid for intensity-modulated radiation-therapy plan selection in prostate cancer based on a prognostic bayesian network and a markov model, Artificial Intelligence in Medicine 46 (2) (2009) 119 – 130.

[8] S. Lee, I. El Naqa, Machine learning methodology, in: I. El Naqa, R. Li, M. J. Murphy (Eds.), Machine Learning in Radiation Oncology, Springer International Publishing, 2015, pp. 21–39.

[9] K. Jayasurya, G. Fung, S. Yu, et al., Comparison of bayesian network and support vector machine models for two-year survival prediction in lung cancer patients treated with radiotherapy, Medical Physics 37 (4) (2010) 1401–1407.

[10] R. D. Timmerman, An overview of hypofractionation and introduction to this issue of seminars in radiation oncology, Seminars in Radiation Oncology 18 (4) (2008) 215 – 222.

[11] K. Fleckenstein, B. Gauter-Fleckenstein, I. L. Jackson, Z. Rabbani, M. Anscher, Z. Vujaskovic, Using biological markers to predict risk of radiation injury, Seminars in Radiation Oncology 17 (2) (2007) 89 – 98.

[12] Y. Chen, P. Rubin, J. Williams, E. Hernady, T. Smudzin, P. Okunieff, Circulating IL-6 as a predictor of radiation pneumonitis, International Journal of Radiation Oncology*Biology*Physics 49 (3) (2001) 641 – 648.

[13] Anthony O'Regan, The role of osteopontin in lung disease, Cytokine and Growth Factor Reviews 14 (6) (2003) 479 – 488, cytokines and Lung Disease.

[14] J. P. Hart, G. Broadwater, Z. Rabbani, et al., Cytokine profiling for prediction of symptomatic radiation-induced lung injury, International Journal of Radiation Oncology*Biology*Physics 63 (5) (2005) 1448 – 1454.

[15] L. Zhao, L. Wang, W. Ji, et al., Association between plasma angiotensin-converting enzyme level and radiation pneumonitis, Cytokine 37 (1) (2007) 71 – 75.

[16] B. Bénéteau-Burnat, B. Baudin, Angiotensin-converting enzyme: Clinical applications and laboratory investigations on serum and other biological fluids, Critical Reviews in Clinical Laboratory Sciences 28 (5-6) (1991) 337–356.

[17] J. H. Oh, J. M. Craft, R. Townsend, J. O. Deasy, J. D. Bradley, I. El Naqa, A bioinformatics approach for biomarker identification in radiation-induced lung inflammation from limited proteomics data, Journal of Proteome Research 10 (3) (2011) 1406–1415.

[18] A. J. Hope, P. E. Lindsay, I. E. Naqa, et al., Modeling radiation pneumonitis risk with clinical, dosimetric, and spatial parameters, International Journal of Radiation Oncology*Biology*Physics 65 (1) (2006) 112 – 124.

[19] I. R. Vogelius, S. M. Bentzen, A literature-based meta-analysis of clinical risk factors for development of radiation induced pneumonitis, Acta Oncologica 51 (8) (2012) 975–983.

[20] H. Jin, S. L. Tucker, H. H. Liu, et al., Dose-volume thresholds and smoking status for the risk of treatment-related pneumonitis in inoperable non-small cell

lung cancer treated with definitive radiotherapy, Radiotherapy and Oncology 91 (3) (2009) 427 – 432.

[21] S. Bracci, M. Valeriani, L. Agolli, V. D. Sanctis, R. M. Enrici, M. F. Osti, Renin-angiotensin system inhibitors might help to reduce the development of symptomatic radiation pneumonitis after stereotactic body radiotherapy for lung cancer, Clinical Lung Cancer.

[22] J. B. S. M. Bentzen, J. Z. Skoczylas, Quantitative clinical radiobiology of early and late lung reactions, International Journal of Radiation Biology 76 (4) (2000) 453–462.

[23] S. Schultz-Hector, K.-R. Trott, Radiation-induced cardiovascular diseases: Is the epidemiologic evidence compatible with the radiobiologic data?, International Journal of Radiation Oncology*Biology*Physics 67 (1) (2007) 10 – 18.

[24] E. X. Huang, A. J. Hope, P. E. Lindsay, et al., Heart irradiation as a risk factor for radiation pneumonitis, Acta Oncologica 50 (1) (2011) 51–60.

[25] M. Roach, D. R. Gandara, H. S. Yuo, et al., Radiation pneumonitis following combined modality therapy for lung cancer: analysis of prognostic factors., Journal of Clinical Oncology 13 (10) (1995) 2606–12.

[26] Z. Allibhai, M. Taremi, A. Bezjak, et al., The impact of tumor size on outcomes after stereotactic body radiation therapy for medically inoperable early-stage non-small cell lung cancer, International Journal of Radiation Oncology*Biology*Physics 87 (5) (2013) 1064 – 1070.

[27] D. Koller, M. Sahami, Toward optimal feature selection, in: In 13th International Conference on Machine Learning, 1995, pp. 284–292.

[28] B. Efron, R. Tibshirani, Improvements on cross-validation: The 632+ bootstrap method, Journal of the American Statistical Association 92 (438) (1997) 548–560.

[29] M. Scutari, On the prior and posterior distributions used in graphical modelling, Bayesian Analysis 8 (3) (2013) 505–532.

[30] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing, Journal of the Royal Statistical Society. Series B (Methodological) 57 (1) (1995) 289–300.

[31] D. M. Chickering, D. Heckerman, C. Meek, Large-sample learning of bayesian networks is np-hard, J. Mach. Learn. Res. 5 (2004) 1287–1330.

[32] J. A. Hoeting, D. Madigan, A. E. Raftery, C. T. Volinsky, Bayesian model averaging: a tutorial, Statistical Science 14 (1999) 382–417.

[33] S. K. Das, S. Chen, J. O. Deasy, S. Zhou, F.-F. Yin, L. B. Marks, Combining multiple models to generate consensus: application to radiation-induced pneumonitis prediction, Medical Physics 35 (11) (2008) 5098–109.

[34] H. Yamashita, K. Nakagawa, N. Nakamura, et al., Exceptionally high incidence of symptomatic grade 2-5 radiation pneumonitis after stereotactic radiation therapy for lung tumors, Radiation Oncology 2 (1) (2007) 21.

[35] J. D. Bradley, A. Hope, I. E. Naqa, et al., A nomogram to predict radiation pneumonitis, derived from a combined analysis of RTOG 9311 and institutional data, International Journal of Radiation Oncology*Biology*Physics 69 (4) (2007) 985 – 992.

[36] C. Elkan, The foundations of cost-sensitive learning, in: In Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence, 2001, pp. 973–978.

CHAPTER 9
Conclusions and Outlook

Contents

9.1 Summary and conclusions

Prediction and understanding of radiation induced lung disease (RILD) is an important element in lung cancer treatment with radiotherapy. The patients who are less vulnerable to RILD can be salvaged by an intensified radiotherapy plan, thus contributing to better overall survival. As well, expanded understanding of radiobiological mechanisms for RILD can be applied to other radiation induced toxicities for better prediction and management. However, the clinical RILD models are static and population-based, assuming the same expected toxicity for any given time if dose were the same across all the patients. These models are not adequate for explaining heterogeneous response despite its convenience as a clinical guideline. This thesis was intended to extend the RILD models to its dynamic behaviour and patient specificity. The extended RILD models were aimed at giving more accurate disease prediction and also contributing to biophysical domain knowledge for RILD using the fitted model parameters.

We first examined the dynamic behaviour of radiation fibrosis from longitudinal analysis of computed tomography (CT) images after radiotherapy. Regional CT density changes were modelled as a function of dose distribution and time. The sigmoidal Lyman-Kutcher-Burman (LKB) model, a widely-used analytical model for radiation-induced toxicity, was fit to the dose-CT change relationships. By observing the fitted LKB parameters, we noticed the dose response was sensitive to post-treatment time and tumour position in superior-inferior direction, but not to dose calculation algorithms. We also observed transition from linear-non-threshold to a nonlinear threshold dose response at 3 months post-RT, suggesting the biological influence that triggers the fibrotic process.

Inspired by these results, we investigated the impacts of non-dosimetric - biological and clinical - factors on RILD risk by exploring a large amount of patient data. After comprehensive literature review (chapter 3), 6 proteins were chosen as candidate RILD biomarkers so that they represent different biological processes involved in RILD pathogenesis. We first attempted to relate these biomarkers to heterogeneity in radiologic response that we previously observed, and capture the onset of such heterogeneity before RT completion. Dose response in CT change during radiotherapy was characterized using a linear mixed model. Although we detected significant patient heterogeneity in model parameters, no significant correlation was found between the biomarkers and CT change in the studied cohort. This study is in progress and we will study more deeply into image texture changes other than HU histogram statistics.

The latter half of the thesis studied how to combine variables from biological, dosimetric and clinical domains to build a prediction model for radiation pneumonitis (RP), an early phase of RILD where response heterogeneity is high. These non-dosimetric variables, along with mean lung dose and irradiated volume at different threshold dose, were first studied individually for their univariate correlations with the disease. Then, High dimensionality of data required a careful model design that regulates model complexity and incorporates uncertainty in model selection. The candidate variables were boiled down to a smaller subset using a information theory based filtering method. The selected variables were connected to each other via a Bayesian network (BN), a probabilistic model that can make inference about probability of RP based on observed variables. In order to account for uncertainty in a single BN graph, Bayesian-style RP prediction was made by averaging the results from an ensemble of BN graphs.

The Bayesian network modelling approach was applied separately to two treatment cohorts - conventionally fractionated (CFRT) and stereotactic body radiotherapy (SBRT) - which are expected to show different toxicity patterns due to fraction size difference. For both cohorts, post-RT change in angiotensin converting enzyme (ACE) was a significant univariate predictor of RP. Mean dose to heart was significant for the CFRT patients, while SBRT cohort revealed volume of high dose ($\geq 105\%$ of prescription dose) spillage outside a target volume as the most important dosimetric predictor. The BN ensemble model predicted the incidence of RP using the connections between the filtered variables, first after observing biomarker response to RP

and again at the treatment planning stage with unknown biomarker response. Although prediction accuracy decreased with less information (missing mid-treatment biomarkers) for both cohorts, the BN model was flexible enough to make inference under varying information availability. The BN model achieved better predictive performance than conventional risk models such as multivariate logistic regression and generalized equivalent uniform dose (gEUD).

9.2 Limitations

Conclusions from this thesis have to be taken with precaution due to a few limitations in data and studies' methodology.

First of all, statistical power of all the presented studies is limited by a small sample size which was never larger than 100. Moreover, each patient group was heterogeneous in treatment modalities, fractionations, or comorbidities. This made it difficult to isolate the effect of one risk factor due to possible confounders. One of the purposes of Bayesian network was to explicitly model this confounder effect. With modest sample size, however, it was difficult to isolate a small number of "concensus" networks, as we observed a large spread of graph posterior.

Scoring outcomes is an important source of uncertainty in RILD modelling (section 2.5). In this work, therapeutic intervention was used as a cutoff severity for defining RP events. Some cases with the presence of intervention were not classified as RP based on suspicion of different etiology such as viral and bacterial infection. This ambiguity in differential diagnosis might translate into uncertainties in our modelling results. If this issue can be resolved, the proposed BN model could be used for assisting differential diagnosis using the predicted RP probability.

Another source of uncertainty is possible discrepancy from planned to delivered dose distribution. SBRT dose with fewer fractions are more likely to suffer from this uncertainty due to higher impact of patient misalignment to summed dose. In BN modelling, however, dosimetric values are discretized and thus small deviation is less likely to affect model integrity than other models that uses continuous values. In the future, accuracy of the dosimetric variables can be improved by correcting for displacement and deformation of patient anatomy using daily cone-beam CT.

9.3 Future outlook

Before translating the presented RILD models to clinical practice, a few obstacles need to be overcome.

First of all, both of the fibrosis or RP models gave a continuous risk score for the tested patients. Prediction performance was measured based on these scores by applying different decision thresholds to the scores and then either: 1) averaging the thresholds (AUC) or 2) taking the classification accuracy at the optimal threshold. However, determination of a decision threshold is should not be based solely on prediction of toxicity. Rather, it should lie in a equation that balances costs (toxicity) and benefits (tumour control) of increasing dose. Adapting decision thresholds to patient preferences for outcomes could be one possible future direction [1].

Secondly, although the model performance and network features were validated using bootstrapping which simulates data distributions different than the training, large-scale external validation is nevertheless the most direct and ultimate test of the utility of the presented RILD models. As seen in Kwa et al.'s DVH analysis from the pooled datasets from 5 institutions [2], studies involving external datasets can help

isolate predictive patterns that are robust to institutional variations in patient and treatment characteristics. However, consistent ways of collecting data and reporting outcomes need to be implemented across centres, as well as adopting a better data sharing or pooling culture [3].

Moreover, Bayesian network graph searching can be assisted by introducing more expert knowledge to the search algorithm. Currently, constraints are used for blocking non-causal links and also for maintaining hierarchical arrangement of variables (e.g. baseline biomarkers cannot directly influence RP risk). In the future, allowed links could be assigned to different prior probabilities according to the degree of confidence that we have on the interactions. For example, a link between biomarkers α2M and TGFβ could be assigned to higher prior probability because α2M is known to bind to TGFβ to regulate its activity [4]. The Bayesian approach with Monte-Carlo sampling provides the opportunity to allow this fusion between prior knowledge with data [5].

REFERENCES

[1] W. P. Smith, J. Doctor, J. Meyer, I. J. Kalet, M. H. Phillips, A decision aid for intensity-modulated radiation-therapy plan selection in prostate cancer based on a prognostic bayesian network and a markov model, Artificial Intelligence in Medicine 46 (2) (2009) 119 – 130.

[2] S. L. Kwa, J. V. Lebesque, J. C. Theuws, et al., Radiation pneumonitis as a function of mean lung dose: an analysis of pooled data of 540 patients, International Journal of Radiation Oncology*Biology*Physics 42 (1) (1998) 1 – 9.

[3] J. O. Deasy, S. M. Bentzen, A. Jackson, et al., Improving normal tissue complication probability models: The need to adopt a data-pooling culture, International Journal of Radiation Oncology*Biology*Physics 76 (3, Supplement) (2010) S151 – S154, quantitative Analyses of Normal Tissue Effects in the Clinic.

[4] X. Chen, X. Kong, Z. Zhang, et al., Alpha-2-macroglobulin as a radioprotective agent: a review, Chin J Cancer Res 26 (5) (2014) 611–21.

[5] A. Werhli, D. Husmeier, Reconstructing Gene Regulatory Networks with Bayesian Networks by Combining Expression Data with Multiple Sources of Prior Knowledge, Statistical Applications in Genetics and Molecular Biology 6 (1).

APPENDIX A
Texture analysis

A.1 Overview

This chapter of the appendix provides introduction and mathematical details on computed tomography (CT) features used in chapter 5.

A.2 First order features

First order features [1], the most basic types of textures, pertains to distribution of grey levels without consideration of their spatial distribution. To obtain such features, distribution of HUs in a region of interest is evaluated as a histogram ($P(b)$) of equally spaced 16 ($= L$) bins from minimum to maximum HU. The following metrics are calculated from the histogram:

- Mean HU : $\sum_{b=1}^{L} bP(b)$
- Quantile metrics: HUs at 70%, 50% (median), 30%, and minimum HU.
- Binned entropy:

$$S = -\sum_{b=1}^{L} P(b) log_2 P(b) \tag{A.1}$$

- Unbinned entropy: For this metric, histogram was recreated with the smallest bin width (1 HU) and the number of bins (L) equal to difference between maximum and minimum HU. Then, the same formula as A.1 was applied to recalculate the entropy in the new histogram.

A.3 Grey level co-ocurrence matrix

Grey level co-occurrence matrix (GLCM) features [2] concern how grey levels are spatially correlated. Specifically, they identify co-ocurrence of a pair of grey levels in neighbouring pixels. In order to calculate GLCM features, a joint probability distribution of a pair of grey levels $1 \leq a, b \leq L$ is calculated from every pair of pixels $\{p_1, p_2\}$ separated by a distance d and an angle θ:

$$P(a, b; d, \theta) = P(p_1 = a, p_2 = b) \tag{A.2}$$

This 2D joint distribution or histogram is called a grey level co-ocurrence matrix. Characteristics of this matrix are known to be related to a type of textures. For example, an image with fine textures tends to have a uniform GLCM, while a course texture increases the skewness of the matrix towards a diagonal [1].

In this study, a single GLCM matrix was evaluated for a separation of $d = 1$ marginalized over the 4 angles (0°, 45°, 90°, and 135°). These four angles cover all possible directions of neighbourhood searching under the assumption of angular symmetry. From the resulting GLCM matrix $P(a, b)$, five metrics were derived: sum average, sum of square variance, sum entropy, difference entropy, and entropy. Mathematical definitions of these features can be found in Haralick et al. [2].

A.4 Fractal features

Fractal refers to self-similar patterns in curves or surfaces. A fractal object can be created by superposition of its downscaled version. This degree of self-similarity

can be quantified by a fractal dimension which was first conceptualized by Mandelbrot [3]. He claimed that any geometrical shapes (curve,surface,...) can be approximated as superposition of N segments of a size r. From this, he defined a fractal dimension as:

$$D = -\frac{\log N}{\log r(N)} \tag{A.3}$$

Measurement of the fractal dimension involves finding the number of small segments (N) that reconstructs the original image while varying its size r. This study used three methods: the Brownian motion method [4], fine and coarse box counting methods [5].

A.5 Laws' filter features

The Laws' filter [6] emphasizes a certain patterns in images into a form of as "energy map". First, an image is convoluted by 4 different 5 x 5 2 dimensional filters, and each pixel of the filtered image is averaged with its neighbours to create a energy map. Entropy of the 4 resulting energy maps from each filter (equation A.1) is taken as a feature value.

The 2D filters are created by multiplying a pair of 1D masks. There are 5 kinds of 1D masks which are designed to emphasize different patterns:

- L5 (level) = [1 4 6 4 1]
- E5 (edge) = [-1 -2 0 2 1]
- S5 (spot) = [-1 0 2 0 -1]
- R5 (ripple) = [1 -4 6 -4 1]
- W5(wave) = [-1 2 0 -2 -1]

For example, a filter E5L5 is created by:

$$\begin{bmatrix} -1 \\ -2 \\ 0 \\ 2 \\ 1 \end{bmatrix} \times \begin{bmatrix} 1 & 4 & 6 & 4 & 1 \end{bmatrix} = \begin{bmatrix} -1 & -4 & -6 & -4 & -1 \\ -2 & -8 & -12 & -8 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 2 & 8 & 12 & 8 & 2 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix} \tag{A.4}$$

This filter, by construction, emphasizes a horizontal edged pattern line. Following the results by Cunliffe et al. [7], four types of 2D filters - R5L5, S5L5, E5L5, W5L5 - were used by this study.

APPENDIX B
Dose calculation algorithms

B.1 Overview

This chapter of the appendix will introduce the two main dose calculation algorithms used in the literatures on lung RILD models.

B.2 Convolution-superposition algorithm

Convolution-superposition algorithm calculates 3D dose deposition by convolving released energy with an energy deposition kernel [8]. The released energy element, total energy released in the mass (TERMA), or $T(\vec{r})$, represents energy released by a primary beam at a point $\vec{r}$ in a medium, and can be written as following:

$$T(\vec{r}) = \frac{\mu}{\rho}\Phi(\vec{r}) \tag{B.1}$$

where μ/ρ is mass attenuation coefficient and $\Phi(\vec{r})$ is a primary fluence at point $\vec{r}$.

The kernel is a 3D exponential function which models energy deposition by scattered beams, and is pre-calculated using the Monte-Carlo method. The final dose D at point $\vec{r}$ is computed as:

$$\begin{aligned} D(\vec{r}) &= \int T(\vec{r})A(\vec{r}-\vec{r'})d^3\vec{r'} \\ &= (T * A)(\vec{r}) \end{aligned}$$

As seen from the use of integral, dose deposition is modelled as linear superposition of scattered dose originating from points over patient body. Heterogeneity correction is possible by scaling the vectors $\vec{r}$ and $\vec{r} - \vec{r'}$ by the density along the paths (radiological path length).

The pencil-beam kernel is one of the earliest implementation of the convolution-superposition in clinic. However, the pencil beam kernel is forward-peaked, which does not adequately model lateral electron transport [9]. Moreover, assumption of charged particle equilibrium results in large underestimation of dose in presence of low density materials where the equilibrium is lost by electrons travelling further (e.g. lung) [9].

The analytical anisotropic algorithm (AAA) has been implemented for the Eclipse^{TM} TPS to enhance the accuracy of pencil beam convolution in heterogeneous media while maintaining reasonable computation time (around a few seconds per beam). AAA improved an accuracy of the pencil beam kernel by separate modelling for primary photons, scattered extra-focal photons, and contaminant electrons [10]. Heterogeneity handling was also improved by correcting the kernel for density in multiple directions [10]. Substantial improvement in accuracy was shown in heterogeneous materials as well as low dose and penumbra region [11].

B.3 Monte-Carlo method

The Monte-Carlo (MC) method refers to a use of repeated random sampling to simulate a process. When applied to dose calculation, it explicitly simulates transport of electrons and photons of a energy range kV∼MV and resulting dose deposition in media. Particle transport in a medium is modelled as a random process;

A random number generator in the MC code samples an interaction cross-section for a given medium to determine the type of interaction and the distance to the next interaction point. Cross-section data is obtained from measurement and important in determining the accuracy of MC code systems. Particle transport is repeated many times (called *histories*) to acquire sufficient statistics, which can be parallelized to improve computation speed.

Unlike the previous two model-based methods, the MC method enables accurate scatter dose and heterogeneity handling, and thus considered as a gold standard [12]. There are multiple Monte Carlo code systems in medical physics with differences in the used language, geometry definition, and a use of approximation. This section will introduce one of the systems that was chosen for this study - the electron gamma shower code system (EGSnrc) developed in the National Research Council in Canada [13]. The name "shower" stems from its unique way of generating particle tracks: while an initial "primary") particle interacts with matters in a medium, it spawns secondary, tertiary, and etc. particle tracks which are appended to the primary particle's history. The tracks are terminated when they reach a certain predefined minimum energy (ECUT for electrons, PCUT for photons) or cross the boundaries of a defined geometry. Figure B.1 summarizes how photon transport is simulated by the EGSnrc.

MC dose calculation with the EGSnrc system consists of two parts: simulation of particle transport in a linear accelerator head, and transport/dose deposition in a patient body.

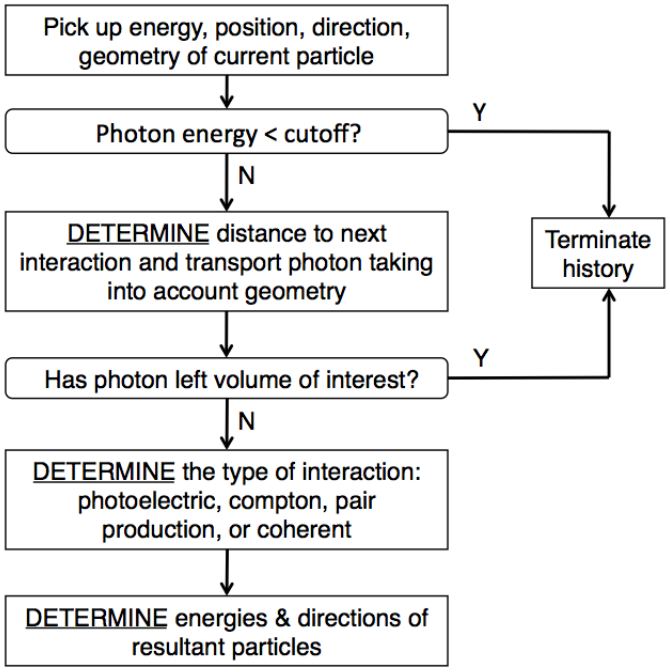

Figure B.1: Flowchart of a MC simulation of photon transport. DETERMINE means that the parameter of the event is found by sampling from an relevant probability distribution. Reproduced with permission from [13].

Linac head simulation. BEAMnrc [14] is a part of the EGSnrc code system that is dedicated to modelling radiation sources and simulating particle transport in a linac head. BEAMnrc requires a geometric model or a "beam model" of the accelerator. A beam model consists of *component modules* (CMs) that are ordered along beam direction. Each CM specifies the geometric shape of an actual accelerator component (e.g. target, monitoring unit chamber, primary and secondary collimators). Although the CMs are designed based on the actual schematic for a linac, some CM parameters are to be determined via "commissioning": the matching of MC-simulated dose distribution to actual measurement. The complete simulation generates a *phase-space file* which contains information about particle position, direction, charge, etc. for every particle crossing a plane in the air gap between the linac head and a patient body. In this work, the beam model for the Varian Clinac^{TM}21X machine was commissioned by testing the accuracy of calculated dose profiles and central-axis percent depth dose (PDD) curves in a water phantom. The simulation was run using a cutoff energy for electrons (ECUT) of 0.700 MeV and photons (PCUT) of 0.010 MeV cross-section data corresponding to the cutoff energies from the photon cross section database (XCOM) from the National Institute of Standards and Technology (NIST). Typical simulation time for creating one phase-space file for 50 million particles was between 2 to 3 hours with 8 x 2.26 GHz processors.

Patient dose calculation. DOSXYZnrc is an EGSnrc component that simulates particle interactions in patient body and scores dose using information on incident particles from the BEAMnrc simulation. Patient geometry is represented in a voxelised phantom derived from an CT image. The phantom is created by

segmenting the CT image into several materials (e.g. air, lung, tissue, bone) based on CT voxel values. Each material is assigned to its unique electron density and cross-section data, which explicitly handles heterogeneity. A dose is scored in the DOSXYZ voxel if a particle entering the voxel has an energy smaller than a threshold (ECUT for electrons, PCUT for photons) . In this work, a free-breathing planning CT was converted to a DOSXYZ phantom with 4 materials (air, lung, tissue, bone) and a voxel size of 5 mm x 5 mm x 5 mm. ECUT of 0.521 MeV and PCUT of 0.010 MeV were used. Dose at each voxel was scored first in units of Gy/particle and later scaled to Gy by multiplying a calibration factor acquired from standard open-beam simulation [1] . Calculation was carried out with the sufficient number of particle histories to achieve clinically acceptable dose uncertainty (under 2% within the target). The required simulation time per beam was 2 to 3 hours on average.

[1] Standard open-beam setup refers to a 10 cm $\times$ 10 cm square field to a water phantom at 100 cm distance from a radiation source to a patient (phantom) surface. The MC dose was sampled at 5 cm depth in phantom and scaled to the dose at d_{max}, depth at dose maximum, using a known PDD. A calibration factor was obtained as a ratio of a known beam output value to the scored dose at d_{max}.

REFERENCES

[1] W. K. Pratt, Digital Image Processing, Wiley, New York, 2001, Ch. Image feature extraction, pp. 509–550.

[2] R. M. Haralick, K. Shanmugam, I. Dinstein, Textural features for image classification, IEEE Transactions on Systems, Man, and Cybernetics SMC-3 (6) (1973) 610–621.

[3] B. Mandelbrot, How long is the coast of britain? statistical self-similarity and fractional dimension, Science 156 (3775) (1967) 636–638.

[4] C. C. Chen, J. S. DaPonte, M. D. Fox, Fractal feature analysis and classification in medical imaging, IEEE Transactions on Medical Imaging 8 (2) (1989) 133–142.

[5] R. Creutzburg, E. Ivanov, Fast algorithm for computing fractal dimensions of image segments, in: Recent Issues in Pattern Analysis and Recognition, Springer-Verlag, London, UK, UK, 1989, pp. 42–51.

[6] K. Laws, Textured image segmentation, Report, University of Southern California (1980).

[7] A. R. Cunliffe, r. Armato, S. G., X. M. Fei, R. E. Tuohy, H. A. Al-Hallaq, Lung texture in serial thoracic ct scans: registration-based methods to compare anatomically matched regions, Med Phys 40 (6) (2013) 061906.

[8] A. Ahnesjo, M. Saxner, A. Trepp, A pencil beam model for photon dose calculation, Med Phys 19 (2) (1992) 263–73.

[9] T. Knoos, A. Ahnesjo, P. Nilsson, L. Weber, Limitations of a pencil beam approach to photon dose calculations in lung tissue, Physics in Medicine and Biology 40 (9) (1995) 1411.

[10] I. Gagne, S. Zavgorodni, Evaluation of the analytical anisotropic algorithm (AAA) in an extreme water-lung interface phantom using monte carlo dose calculations, Journal of Applied Clinical Medical Physics 8 (1).

[11] A. Van Esch, L. Tillikainen, J. Pyykkonen, et al., Testing of the analytical anisotropic algorithm for photon dose calculation, Vol. 33, AAPM, 2006, 11.

[12] I. J. Chetty, B. Curran, J. E. Cygler, et al., Report of the aapm task group no. 105: Issues associated with clinical implementation of monte carlo-based photon and electron external beam treatment planning, Medical Physics 34 (12) (2007) 4818–53.

[13] D. W. O. Rogers, A. F. Bielajew, Monte carlo techniques of electron and photon transport for radiation dosimetry, in: The Dosimetry of Ionizing Radiation, Academic Pr., 1990.

[14] D. W. O. Rogers, B. A. Faddegon, G. X. Ding, C.-M. Ma, J. We, T. R. Mackie, BEAM: A monte carlo code to simulate radiotherapy treatment units, Medical Physics 22 (5) (1995) 503–524.