MANCHESTER 1824
The University of Manchester

Voxel-Based Causal Inference in Radiotherapy: A Simulation Study

A thesis submitted to the University of Manchester for the degree of
Master of Philosophy
in the Faculty of Biology, Medicine and Health

2021

Alexander Llywelyn Jenkins
School of Medical Sciences
Department of Radiotherapy Related Research

Contents

Word Count: 14443

List of Figures

List of Tables

List of Publications

1. Jenkins, A., Vásquez Osorio, E., Merchant, M., Green, A., van Herk, M, Sperrin, M. & McWilliam, A. Average Causal Effect Estimation for High-Dimensional Spatial Treatments: Radiotherapy Simulation Study. *Royal Statistical Society 2021 International Conference*. `https://app.oxfordabstracts.com/events/1880/program-app/submission/252573` (Sep. 2021).

Acronyms

AI Artificial Intelligence. 61

ATE Average Treatment Effect. 5, 11, 17, 33–35, 38, 41, 42, 44, 47, 51–56, 58, 60, 63, 64, 66, 67

AUC Area Under Curve. 7, 27, 31

CATE Conditional Average Treatment Effect. 33–35

CCT Concurrent Chemotherapy. 7, 27

CNN Convolutional Neural Network. 61

CPHM Cox Proportional Hazards Model. 28, 30

CT Computed Tomography. 16, 20, 42

CTV Clinical Target Volume. 4, 20, 22

DAG Direct Acyclic Graph. 4, 5, 33–38, 46–48, 51, 54, 61–64, 66

DIR Deformable Image Registration. 7, 27, 28, 31

DVH Dose-Volume Histogram. 4, 7, 16, 20, 22, 23, 25–28, 42, 43, 45, 61

FSU Functional Subunit. 24, 25, 62

GTV Gross Tumour Volume. 19, 20

HN Head and Neck. 7, 27

ITE Individualised Treatment Effect. 32, 33

LKB Lyman-Kutcher-Burman. 7, 23–25, 31

LR Logistic Regression. 7, 31

MRI Magnetic Resonance Imaging. 16, 20, 26, 30, 42, 43

MSE Mean Squared Error. 6, 11, 42, 53, 54, 59

NLP Natural Language Processing. 61, 62

NTCP Normal Tissue Complication Probability. 4, 7, 16, 17, 21–31, 44

OAR Organ At Risk. 4, 11, 16, 17, 20, 22–26, 41, 42, 44, 45, 60–62, 64, 66, 67

PCM Pharyngeal Constrictor Muscle. 7, 27

PET Positron Emission Tomography. 16, 20, 42

PT Permutation Test. 7, 31

RCT Randomised Controlled Trial. 17, 44, 45, 66

RILD Radiation-Induced Lung Disease. 7, 31

RIR Rigid Image Registration. 7, 31

RT Radiotherapy. 7, 27

SCM Structural Causal Model. 29, 33–35, 38, 48, 49, 51, 58, 60, 63, 64, 66, 67

TCP Tumour Control Probability. 4, 21

TFCE Threshold Free Cluster Enhancement. 7, 31

VMAT Volumetric Modulated Arc Therapy. 46, 49, 60, 66

Abstract

Radiotherapy is the most common treatment for cancer, delivering 3D, personalised radiation dose to the tumour. Radiotherapy planning requires considering a high-dimensional continuous optimisation space to achieve tumour control while limiting the probability of treatment complications. However, estimating the Average Treatment Effect (ATE) of radiation dose on complications across the anatomy is difficult; mainly because of confounding in observational data. Under certain assumptions, developing a causal framework provides methods to adjust for confounding.

The aim of this work is to use simulated data to investigate if unbiased and consistent voxel-based causal inference is possible, how, under what circumstances, and with what accuracy. I simulate radiotherapy treatment plans from a simplified, yet realistic, data generating process. Patients have a single tumour (random location) where dose is maximal and a single Organ At Risk (OAR) (fixed location) where dose is minimal. Variables control fall-off of dose around the tumour, fall-off and magnitude of dose at the organ, and a covariate that confounds the treatment plan. I simulate realistic treatment uncertainties: random shifts in x- and y-directions of the entire planned dose distribution, spatially correlated noise sampled from a Gaussian process prior, and independent noise at each pixel. A continuous complication is generated via a linear function of the delivered dose to a spatially inhomogeneous set of pixels (ATE estimand), a covariate that also affects the delivered dose distribution, and a spatially inhomogeneous interaction between delivered dose and the covariate. Three methods based in causal inference are used to estimate the ATE at each pixel: 1) pixel-wise sparse causal regression, 2) sparse causal regression and 3) a causal regression. The sparse estimator used is the Adaptive Lasso. These are compared to methods currently used in radiotherapy.

I found that all methods based in causal inference performed with lower total Mean Squared Error (MSE), MSE_{tot}, across all parameterisations tested in the simulation compared to the currently used voxel-based statistical methods in radiotherapy. Exploiting the oracle property of the Adaptive Lasso to simultaneously identify important pixels with dose-response and estimate ATE, was in general a successful technique over all parameterisations of the simulation tested. The only method capable of unbiased estimation was the causal regression, however, multicollinearity hinders accurate ATE estimation at specific regions of parameter space and at a high resolution. The estimation method that scored a consistently low MSE_{tot} over all parameterisations was the sparse causal regression. This method was able to assign a near zero effect to unimportant pixels, and whilst estimates elsewhere were biased, they were accurate and efficient; especially at lower resolutions.

Lay Abstract

Radiotherapy is the most common treatment for cancer, delivering high-energy radiation in 3D to destroy tumour cells. Key research questions now focus on reducing treatment complications and improving the quality of life for patients living with and beyond cancer. It is believed that complications of radiotherapy treatment are caused by radiation dose to healthy tissue. Whilst there exists strong evidence linking certain treatment complications to radiation dose to specific organs, it remains uncertain how other treatment complications are caused.

Analytical methods have been developed to identify, from past radiotherapy patients, regions of the anatomy where dose could be generating the treatment complications. However, between studies of the same treatment complication, there exists uncertainty in the exact regions identified and how sensitive the complication is to increasing dose in the region, i.e. the effect of dose. I hypothesise this is due to confounding present in the analysis of past data, i.e. the presence of factors that influence the treatment plan and the complication, such as the patients' diagnosis. Establishing the causal effect of dose on complication, i.e. the measured effect when confounding and other biases are removed, across the anatomy, could reduce current uncertainty and lead to new and impactful insights on the causes of treatment complications. To do this, the field of mathematics, known as causal inference, needs to be used.

The aim of this work is to use simulated data to investigate if unbiased and consistent voxel-based causal inference is possible, how, under what circumstances, and with what accuracy. I simulated simplified, yet realistic, radiotherapy treatment-outcome data and sought to establish methodologies that, for the first time, can estimate the causal effect of dose on complication across the anatomy. I found that all methods based in causal inference performed better in the simulation compared to current voxel-based statistical methods in radiotherapy. In particular, one method I tested, the sparse causal regression, was able to accurately identify and ignore regions with zero causal effect. By ignoring the values of dose in these unimportant regions, accurate causal effect estimates were possible in the important regions; where it was needed most. This method showed potential to work for lower sample sizes and also at higher resolutions. Whilst our work makes the first steps, further work is needed to scale up our methodologies for the analysis of clinical data.

Declaration of Originality

I hereby confirm that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

i The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made *only* in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see `http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=24420`), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see `http://www.library.manchester.ac.uk/about/regulations/`) and in The University's policy on Presentation of Theses.

Acknowledgements

This MPhil thesis would not have been possible without the kind help and support of my colleagues, friends and family.

I would firstly like to thank all my supervisors: Dr Eliana Vásquez Osorio, Dr Michael Merchant, Dr Andrew Green, Dr Matthew Sperrin and Dr Alan McWilliam. Your patience, guidance and support, has been truly invaluable. I would like to especially thank my main supervisors, Dr Alan McWilliam and Dr Matthew Sperrin, for their encouragement to keep experimenting and pursuing the ideas presented in this thesis, despite uncertainty at times. I would also like to thank Professor Marcel van Herk for welcoming me back into the Radiotherapy Related Research (RRR) group at The Christie NHS Foundation Trust, and for your time, guidance, kindness, full support and constant inspiration during my time there. I would like to extend this thank you to the PhD students and additional researchers in the RRR group. It has been uniquely challenging pursuing this MPhil during the COVID-19 pandemic, but it would have been impossible without these colleagues. It has been an absolute pleasure to work with you.

Secondly, I would like to thank my parents, sister and beautiful girlfriend Alice, for supporting me and keeping me motivated. I could not have done this without you.

Finally, I would like to thank Cancer Research UK and the Manchester Cancer Research Centre for funding this research.

Chapter 1

Introduction

Radiotherapy is the most affordable and accessible form of cancer treatment [1]. It is, therefore, the most common form of cancer treatment, with approximately 50% of cancer patients receiving it during their course of treatment [2]. Advancements in imaging, e.g. 3D Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), alongside advancements in treatment delivery and technology, e.g. image-guided radiotherapy, mean that uncertainties in radiotherapy treatment are reducing. As a result of technological developments in imaging and treatment delivery allowing for a more conformed dose to the tumour, 5-year survival rates have been increasing over the past 25 years for most tumour sites [3]. Therefore, key research questions now focus on reducing treatment complications and improving the quality of life for patients living with and beyond cancer.

In order to maximise the quality of life for patients, the probability of treatment complications must be minimised. Many complications of radiotherapy are currently believed to be caused by radiation dose to Organs At Risk (OARs). Therefore, current radiotherapy treatment plans are constructed based on this evidence, recommended OAR dose constraints, and complication predictions from Normal Tissue Complication Probability (NTCP) models [4]. NTCP models are a method to predict treatment complications from the proposed treatment plan. Ultimately, NTCP models could be used in clinical practice to select the optimal treatment plan or treatment modality (e.g. photons or protons) that minimises the risk of complications for an individual.

NTCP models that are currently used in clinical practice are built from dose statistics extracted from the Dose-Volume Histogram (DVH) of the whole OAR considered. However, following the introduction of the 'voxel-based analysis' of radiotherapy treatment-outcome studies, there has been growing evidence to support the existence of sub-regions of OARs that are particularly radiosensitive for a given complication; as shown for the bladder [5], the parotid glands [6], white-matter structures [7], the lung [8] and the heart [9]. The voxel-based analysis is able to overcome the assumption of DVH methods that the whole organ equally drives complications; which is known to be false [10, 11]. Therefore, modelling NTCP using dose to these sub-regions, rather than the entire OAR, is suggested to be better predictive of complications.

Building better NTCP models is not the only goal of voxel-based methods. They could re-

duce complications by allowing for the identification of new and better interventions that the clinician and treatment planning staff can make on the radiotherapy treatment plan. For example, to avoid a newly identified highly sensitive region completely rather than the whole OAR. In addition, the voxel-based method is also being used to create 3D maps of the spatial effect of radiotherapy dose, and other clinical variables, on complications [12, 13]. This is highly desirable as it gives an insight into the mechanism that is potentially generating the complications, and could allow us to formulate new spatial dose-constraints for complications. In summary, the aims of the voxel-based analysis in radiotherapy are (1) to identify sub-regions of organs with a given dose-response and (2) estimate the effect of radiotherapy voxel dose on outcome.

In radiotherapy treatment-outcome studies, the main source of bias is confounding. A confounder is a variable that influences the treatment plan and also influences the likelihood of complication. However, current voxel-based methods do not adequately adjust for confounding. This could explain the variability in the location of radiosensitive sub-regions identified for OARs, for example, McWilliam et al. [9] and Bogaard et al. [14] identified unique sub-regions of the heart that are particularly radiosensitive. It also could explain why models of the same complication differ in their identified significant variables and their effect, and hence, why the external validation of current NTCP models is so important [15, 16]. Under certain assumptions, developing a causal framework provides methods to adjust for confounding.

The gold standard causal framework is typically the Randomised Controlled Trial (RCT). In an RCT, treatment and control groups are formed by random assignment. As a result, the two groups are seen as balanced in terms of patients' characteristics. In an RCT, one can infer the Average Treatment Effect (ATE) of treatment A versus treatment B on an outcome of interest with respect to the RCT population, where treatment A and B could be exposure versus no exposure, two competing treatments, or the same treatment modality but at different doses. However, using an RCT to identify the effect of radiotherapy treatment at different anatomical regions is not ethical. The planned dose distribution should always be tailored to the patient characteristics in order to give the best chances of curative treatment whilst reducing probability of complications. Therefore, under this ethical constraint, a causal framework must be developed using observational data.

The aim of this work is to use simulated data to investigate if unbiased and consistent voxel-based causal inference is possible, how, under what circumstances, and with what accuracy. To do this, I simulate data from a simplified, yet realistic, data generating process for radiotherapy treatment-outcome data.

Chapter 2

Radiotherapy

2.1 Radiotherapy physics

2.1.1 X-ray therapy

The X-ray was discovered in 1895 by Rontgen. Just 3 months after their discovery, X-rays were used for diagnostics, and within 1 year X-rays were used to treat the first cancer patient [17]. Due to the properties of X-rays, X-ray therapy remains a key treatment modality in cancer treatment today. X-rays are high-energy photons: massless and neutrally charged force carrying particles of the electromagnetic force. X-rays interact only with the electromagnetic force, hence they can interact with charged particles. Most commonly, photons interact with electrons when passing through a medium such as the human body. Photons can interact with a medium of charged particles via the photoelectric effect, Compton scattering or pair production. X-ray radiotherapy typically uses photons of energy 6-20 MeV; at this energy Compton scattering is the most common form of interaction [17].

For an X-ray beam of initial intensity I_0, incident on a medium with linear attenuation coefficient μ, the intensity as a function of depth in the medium x can be written as

$$I(x) = I_0 \exp(-\mu x), \tag{2.1}$$

which is plotted in figure 2.1 [18]. It should be noted that the X-ray depth-dose curve is greatest upon entry into the medium, and then decays exponentially as a function of depth. It should be noted that $I(x) \rightarrow 0$ as $x \rightarrow \infty$, i.e. X-rays do not stop unless involved in an interaction. In cutting-edge X-ray radiotherapy today, X-ray beams are delivered via a gantry that can accumulate a conformal dose to the tumour from multiple beam directions. Therefore, despite the high entry dose of X-rays, and low exit dose, using multiple beam directions allows for the prescribed dose to be accumulated at and conformed to the target, whilst minimising the dose to normal tissue.

2.1.2 Proton therapy

Protons, unlike photons, have both charge and mass. A proton has a charge of -1e (positive) and mass 911 MeV/c^2. When a proton is passing through a medium at speed, it will interact

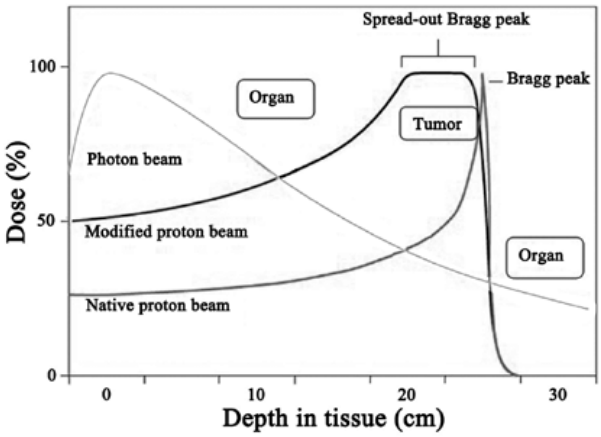


Figure 2.1. A depth-dose curve for photons and protons as they transit tissue. Adapted from [18].

with other charged particles through the Coulomb force, for a given transit time (attracted to oppositely charged electrons, and repelled by like charged atomic nuclei). With each interaction, momentum is transferred from the proton to the medium, in an amount proportional to the time of transit. As energy is proportional to the square of momentum transferred, the energy transferred in each interaction is proportional to the square of the transit time. As a result, the proton transfers most of its energy as it is about to stop. The depth-dose curve for protons is illustrated in figure 2.1, where the properties described above for protons are observed as a *Bragg peak* on the plot [18].

The Bethe-Bloch equation accurately describes the *stopping power* of ionising particles, i.e. the energy transferred from an ionising particle per unit distance of a medium. For a proton at clinical energies (3-300 MeV), the stopping power can be written as

$$-\frac{dE}{dx} = 0.3072\frac{Z}{A}\frac{\rho}{\beta^2}\left[\ln\frac{W_m}{I} - \beta^2\right]\frac{\text{MeV}}{\text{g/cm}^2}, \tag{2.2}$$

where $W_m = \frac{2m_e c^2\beta^2}{1-\beta^2}$ is the largest energy a proton can transfer in a collision with a free electron, Z is the atomic number of the medium, A is the mass number of the medium, ρ is the density of the medium, $\beta = v/c$, where v is the speed of the proton and c is the speed of light, I is the mean excitation energy of the medium, and m_e is the mass of an electron [19]. The Bethe-Bloch equation describes the underlying physics that gives rise to the Bragg peak phenomenon illustrated in figure 2.1 [18]. Proton beam therapy was first proposed by Wilson [20] in 1946 after theorising the depth-dose properties of protons could be exploited to accurately treat cancers. In the 1950's the first patients received proton beam therapy [21].

2.2 The high-level radiotherapy workflow

The key to the success of radiotherapy is the clinician and treatment planning staff. In current radiotherapy clinical practice, the dose and the number of fractions are set depending on the best evidence from past clinical trials. The clinician will segment the Gross Tu-

mour Volume (GTV) or Clinical Target Volume (CTV) (depending on cancer site), and the treatment planning staff will segment the OARs of radiation damage on a 3D CT/MRI/PET scan of the patient, depending on what is useful and available. With segmentations complete, a dose distribution (dose accumulated over a set of beam directions) is then found by minimising the dose to the OARs whilst delivering the planned dose conformed to the tumour. Today, the dose to the structures is quantified using a DVH, where each OAR has a guideline for absorbed dose limits and the expected risk of complications, as stated in the QUANTEC guidelines [4]. However, the clinician could preferentially adapt these OAR dose guidelines, based on their clinical experience, clinical trial data, etc., of what may be beneficial for the patient given their diagnosis, demographic or co-morbidities. Even in this scenario, the priority of treatment is tumour control, and therefore it may be inevitable that dose to specific OARs must be increased in order to maintain tumour coverage.

Chapter 3

Dose-Response Modelling

3.1 Fundamentals of dose-response modelling

As discussed in section 2.2, the clinician and treatment planning staff are key to success in radiotherapy. It is their job to maximise dose to the tumour, i.e. maximising Tumour Control Probability (TCP), whilst minimising the dose to the surrounding normal tissue, i.e. minimising NTCP. This trade-off is highlighted in figure 3.1, where the optimal dose for the patient is found in the therapeutic window, at the position of maximum TCP for minimum NTCP [22].

With radiotherapy advancement, a key focus for specific tumour sites is to reduce treatment complications. NTCP modelling is a method to relate dose to normal tissue with complication outcomes. According to the Encyclopedia of Radiation Oncology [23], NTCP is defined as:

> *The probability that a given dose of radiation will cause an organ or structure to experience complications considering the specific biological cells of the organ or structure.*

However, how can one determine the dose-response relationship of NTCP? Many different NTCP models exist, developed from a range of sources: *in vitro*/*in vivo* measurements for different tissue types, bio-statistical models, and data-driven models (considering observed

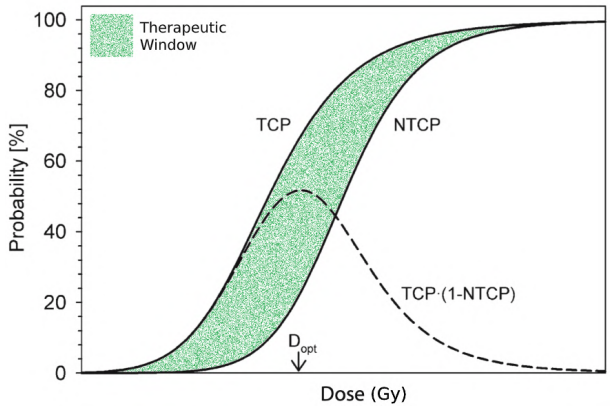


Figure 3.1. The dose-response curve for Tumour Control Probability (TCP) and Normal Tissue Complication Probability (NTCP). Adapted from [22].

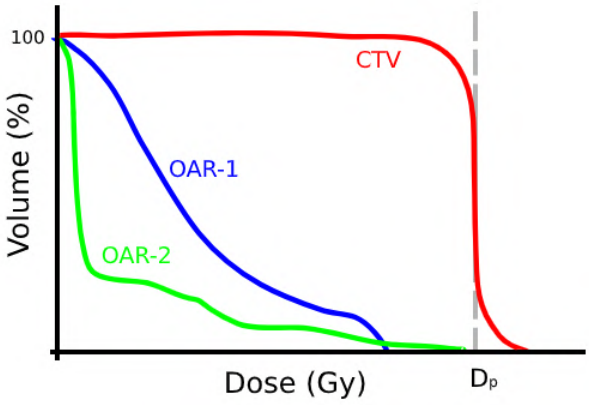


Figure 3.2. An example Dose-Volume Histogram (DVH) evaluating a treatment plan for a fictional patient with two different Organ At Risk (OARs). Note that a large proportion of the Clinical Target Volume (CTV) receives the high prescribed dose, D_p, whilst a large proportion of the OAR volumes receives a low dose.

probability rates from the clinic). However, regardless of model choice, NTCP modelling is not deterministic given the unknowns in the complex system describing the patient, the biology and the radiation response. Therefore, NTCP is better described probabilistically. In the next section, I discuss the history of NTCP and critique current methodologies.

3.2 Early work on NTCP

3.2.1 Volume-based models

Early evidence originating in the 1940's highlighted that radiation response of normal tissue depends on the volume of tissue irradiated, a result since confirmed in animal experiments [24, 25, 26]. Therefore, a graphical tool, namely the DVH, has been developed to summarise the dose distribution and assess the dose-volume dependency of segmented volumes [27]. As an example, figure 3.2 shows a DVH for a CTV and two different OARs, and from this you can see that the OAR volumes typically receive a lower (yet non-zero) dose compared to the CTV.

Lyman-Kutcher-Burman model

In 1985, Lyman [28] proposed a 4-parameter model of NTCP that is based on the DVH. In this model, the probability of complication, c, given the mean dose to volume, D, volume irradiated uniformly, V, and parameter vector, $\mathbf{w}$, is defined as

$$P(c|D,V,\mathbf{w}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{x^2}{2}} dx, \tag{3.1}$$

where

$$t = \frac{1}{m}\left(\frac{D}{D_{50}(\nu)} - 1\right), \tag{3.2}$$

$$D_{50}(\nu) = D_{50}(1)\nu^{-n}, \tag{3.3}$$

and

$$\nu = \frac{V}{V_{ref}}. \tag{3.4}$$

The parameters of the model are $\mathbf{w} = \{D_{50}(1), V_{ref}, m, n\}$, where $D_{50}(1)$ is the tolerance dose for a whole volume that gives rise to a 50% complication rate, V_{ref} is a reference volume, n relates partial and whole volume effects, and m parameterises the dose-response curve [28, 22]. It should be noted that Lyman's model assumes the following:

1. The volume, V, is radiated uniformly with dose, D.
2. The radiosensitivity within the volume, V, is homogeneous.
3. The dose, D, to volume, V, uniquely drives complications.

The parameters of the Lyman model are determined by fitting data, and can be determined for different tissue types, and different biological and clinical endpoints [28]. However, in 1985, the parameter estimates were highly uncertain, given the fact that the best available data at the time was from observations, often incomplete, of patients treated before the days of sophisticated treatment planning (there were significant errors in position of OAR irradiated and uniformity of irradiation) [29, 30].

The uncertainty in model parameters, and hence NTCP, limited the utility of the Lyman model. But in 1991, Emami et al. [31] summarised the existing literature on normal tissue response, and added expert clinical insight to the problem. For each OAR, Emami et al. [31] established the most clinically severe complication, and highlighted what doses to different sub-volumes ($\frac{1}{3}$, $\frac{2}{3}$ and whole) of these organs will give rise to D_5 and D_{50}, the tolerance doses that give rise to a 5% and 50% complication rate, respectively.

In the same year, Burman et al. [32] took Lyman [28]'s model for NTCP, and interpolated Emami et al. [31]'s guidance to estimate NTCP for any combination of dose and irradiated volume. Kutcher et al. [33] then tackled the assumptions posed by the Lyman model, specifically the uniform irradiation assumption, to provide a method of calculating NTCP under non-uniform irradiation of organs. This method involves fitting the Lyman model to Emami et al. [31]'s data and estimating the four parameters, and then transforming a non-uniform DVH into biologically iso-effective uniform DVH using a DVH-reduction algorithm. The Lyman-Kutcher-Burman (LKB) model was formed by taking the original Lyman model and fusing Kutcher's and Burman's improvements, to calculate ν using Kutcher et al. [33]'s DVH-reduction algorithm and to use Emami et al. [31]'s and Burman et al. [32]'s parameter values, respectively. The LKB model soon became the most popular method to estimate NTCP in radiotherapy. However, some limitations remain:

1. The radiosensitivity within the volume, V, is assumed to be homogeneous. Yet, evidence exists to prove otherwise [11].

2. The parameters D_5 and D_{50} at partial and whole volumes, have been estimated by Emami et al. [31] *by-eye* from noisy, old, data.

3. Assumes that radiation dose single-handedly drives complications. In reality, NTCP is a complex interplay of radiation, interaction with variables such as concurrent chemotherapy, and confounded by variables such as patient genetics, demographics, medical history, etc.

3.2.2 Biophysical models

Early work by Schultheiss et al. [34] and Withers et al. [35] introduced the concepts of *serial* and *parallel* organs, and the Functional Subunit (FSU), alongside their probabilistic description. Serial and parallel organs are assumed to be composed of FSUs; a sub-volume of the organ that contributes to its function [35]. For serial organs, disabling one FSU can cause complications, for example, this is thought to describe organs such as the spinal cord, brain and bowel [35, 22]. For parallel organs, many/all FSUs must be disabled to cause complications, for example, this is thought to describe organs such as the kidneys, liver and lungs [35, 22]. In FSU models, the conditional probability for complication of the subunit, c, can be written as,

$$P(c|D, D_{50}, k) = \frac{1}{1 + \left(\frac{D_{50}}{D}\right)^k}, \tag{3.5}$$

where D is the dose to the subunit, D_{50} is the subunit's tolerance dose for an observed 50% complication probability, and k parameterises the subunit's dose-response curve [34, 35]. Therefore, FSU models assume that each subunit has a specific dose-response; alleviating the assumption of homogeneous dose-response in the LKB model.

Critical element model

The critical element model was introduced by Niemierko and Goitein [36] in 1991. The model assumes that each OAR consists of numerous identical FSUs [35], and that a complication will occur if a single FSU is disabled. Therefore, the critical element model is thought to describe how complications arise in serial organs.

For an OAR composed of N identical FSUs under non-uniform irradiation of dose D, the probability of whole organ complication can be written as

$$P(c|D, 1) = \prod_i^N P(c|D_i, \nu), \tag{3.6}$$

where $P(c|D_i, \nu)$ is the probability of complication in each FSU, $\nu = \frac{1}{N}$ is the volume of each FSU and D_i is the dose to each FSU [34]. Therefore, the magnitude of the product is

driven by the FSUs that receive the highest doses. As a result, in NTCP models for serial organs, the maximum dose to the volume is often considered as a predictor [35, 36].

Critical volume model

The critical volume model was introduced by Jackson et al. [37] and Niemierko and Goitein [38] in 1993. The model assumes that each OAR is composed of numerous identical FSUs, and that a complication occurs when a critical number, L, FSUs are disabled. Therefore, the critical volume model describes how complications arise in parallel organs.

In this case, the probability of whole organ complication can be written as

$$P(c|D,1) = \sum_{i=L}^{N} P(i|D), \tag{3.7}$$

where $P(i|D)$ is the conditional probability that exactly i subunits are destroyed given dose D [37, 38]. As a result, in NTCP models for parallel organs, the volume that receives a specific dose, and the mean dose in the volume, are often considered as a predictors [37, 38].

3.3 What is the current state-of-the-art in NTCP?

3.3.1 DVH-based NTCP models

DVH-based NTCP models are the first category of models that will be critiqued here. Here, the hypothesis is that dose to a specific OARs drives specific complications. For each patient, the OARs are segmented during treatment planning by the clinician and treatment planning staff. For analysis, the dose distribution enclosed by each OAR is summarised using a DVH, then multiple summary statistics are extracted from each DVH. For example, minimum dose, maximum dose, mean dose, and the volume that receives X Gy [39]. From the list of statistics summarising the dose to each region, and alongside clinical variables, features are selected to model a given complication. This process is illustrated by figure 3.3.

The advantage of DVH-based NTCP models is in their simplicity. They compress a complex dose distribution into a few numbers, and the parameters of the statistical model can be interpreted. In addition, the data-driven approach overcomes one of the assumptions of the LKB model, that dose uniquely drives complications, as additional clinical variables can easily feature here. However, this process involves some critical assumptions:

1. Assumes that one specific complication is driven only by a single segmented organ. However, there is growing evidence to support that the onset of complications can be driven by dose to multiple regions [10]. On top of this, the definition of the organ is dependent on segmentation by the clinician and treatment planning staff, a process

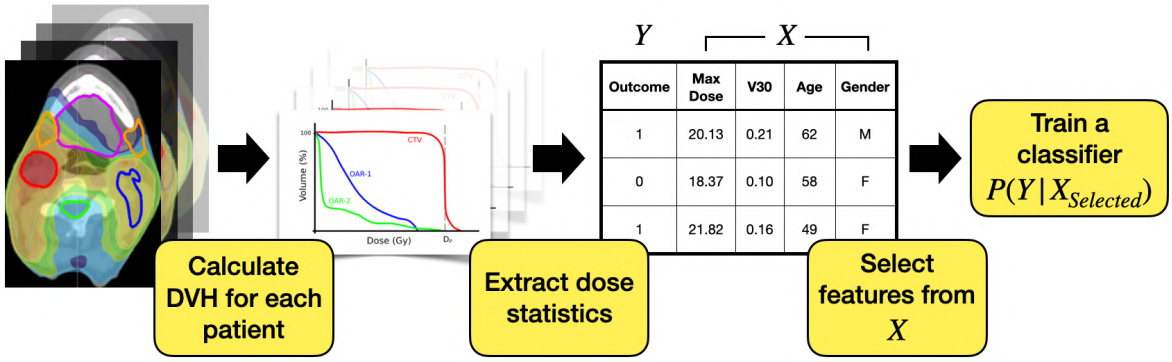


Figure 3.3. The typical process used in literature to create a multivariate Normal Tissue Complication Probability (NTCP) model from the planned dose distributions (left) and the Dose-Volume Histogram (DVH) of an organ relevant to the complication outcome.

known to be highly subjective and a large source of error in radiotherapy treatment planning [40]. This segmentation uncertainty will manifest as uncertainty in the statistics drawn from the DVH.

2. Assumes the whole organ equally drives complication. This is known to be an over-simplification of the radiobiology within the organ [10]. Whilst a volume effect on complication is often acknowledged in literature [24, 25, 26], radiosensitivity within organs is known to be inhomogeneous [11]. In addition, there is growing evidence to support the existence of radiosensitive sub-regions of organs that may drive complications [5, 6, 7, 8, 9].

3. Ignores the spatial dose distribution within the volume, i.e. many different spatial dose distributions can give rise to the same DVH. Although histogram reduction techniques can be applied to describe inhomogeneous dose-volume radiation, as discussed in section 3.2, the spatial dose information remains lost in the simplification of the dose distribution to a histogram.

The literature on DVH-based statistical NTCP models is extensive, as it dates back to the 1990's and is easily adapted to a range of clinical outcomes. To highlight the variety of DVH-based models, a review of the literature regarding a range of clinical endpoints relevant in head and neck cancer are summarised in table 3.2. In summary of table 3.2, 11/16 of the studies model NTCP at a single organ using a simple linear/logistic regression. The advantage of these models is that they are interpretable, yet despite this, only 25% of the studies in the table have been externally validated. As a result, studies of the same complication often identify alternate significant predictors. See the review of Sharabiani et al. [41] for an extensive literature search on this subject.

3.3.2 Voxel-based NTCP models

Voxel-based methods seek to identify regions of an image which are significantly associated with an outcome of interest. Where a voxel is a point / tiny volume in 3D space. Voxel-based methods are not new, the method has been well established in the field of neuroimaging since the 1990's [62]. For example, Whitwell [63] used the methodology to find regions of MRI images associated with Alzheimer's disease. The potential of the methodology was fast to adapt to the field of radiation oncology, with the first voxel-based analysis emerging

Date	Authors	Patients	Outcomes	Model	Significant predictors	Variable selection	External validation
				Xerostomia			
1999	Eisbruch et al. [42]	88 (HN)	Xerostomia at each parotid (25% reduction in salivary flow at 12 months)	Logistic regression	Mean dose, baseline saliva output, high/low dose indicator variable	Stepwise model selection. Spearman's rank correlation coefficient for performance	NA
2006	Naqa et al. [43]	55 (HN)	Xerostomia (25% flow reduction at 6 months)	Logistic regression	mean dose to both parotids, gender	Bootstrapping and Bayesian information criterion. Spearman's rank correlation coefficient for performance	NA
2012	Beetz et al. [44]	165 (xerostomia) and 167 (sticky saliva)	4-point scale graded by patients at 6 months	logistic regression	Mean dose to parotid, age, baseline score (xerostomia); mean dose to submandibular and sublingual glands, age, baseline score	Bootstrapping and forward variable selection Likelihood measures	NA
2014	Lee et al. [45]	206 (HN)	Xerostomia graded by patients (grade 3+ at 3 and 12 months)	Logistic regression with Lasso	Mean dose to contralateral (D_{mean-c}) and ipsilateral (D_{mean-i}) parotids, age (3 months); D_{mean-c}, D_{mean-i}, education and financial status, smoking and T-stage (12 months)	Forward model selection with bootstrapping. Performance Brier score, AUC, chi-squared, omnibus, hosmer-lemeshow test	NA
				Hearing loss			
2015	Lee et al. [46]	422 (HN)	Tinnitus (grade 2+ at any point during follow-up)	Logistic regression	Mean dose to cochlea >32 Gy	Performance measured by Brier score and AUC	NA
2017	Cheraghi et al. [47]	35 (HN)	Pure tone audiometry assessment each ear at 12 months	Logistic regression	Cochlea volume	Model selection using Akaike's information criterion	NA
				Swallowing dysfunction, feeding tube and dysphagia			
2012	Christianen et al. [48]	354 (HN)	Grade 2+ swallowing dysfunction by clinician at 6 months	Logistic regression	Mean dose to PCM and mean dose to supraglottic larynx	Bootstrapping and forward variable selection. Likelihood measures and AUC	Externally validated by Christianen et al. [49] (same country). NTCP overestimated by 5% on average, but good fit
2014	Wopken et al. [50]	335 (HN)	Feeding tube dependence at 6 months	Logistic regression with Lasso	Baseline weight loss, treatment modality (CCT, accelerated RT or RT + cetuximab), T-stage, PCM superior and inferior mean dose, contralateral parotid dose, cricopharyngeal muscle mean dose (Gy)	Bayesian information criterion for selection of Lasso hyperparameter. 10-fold cross-validation for performance, measured using AUC	External validation of this model by Kanayama et al. [51], predicted twice as many as observed feeding tube dependencies. Adjusting only intercept (baseline risk) was needed for good fit
2014	Wopken et al. [52]	427 (HN, prospective)	Feeding tube dependence at 6 months	Logistic regression with Lasso	T and N-stage, weight loss and treatment modality (bilateral, accelerated RT and CCT)	Bayesian information criterion for selection of Lasso hyperparameter	Not externally validated. Test set used (183 patients) with good performance via AUC
2015	Otter et al. [53]	253 (HN)	Grade 3+ pharyngeal dysphagia by clinician at 2 months	Logistic regression	Inferior PCM mean dose, CCT, gender	Model fitted using bootstrap. Performance measured using AUC	NA
				Neurological performance			
2013	Redmond et al. [54]	19 paediatric cases (Brain). Prospective	Neuropsychological performance (Memory, vocabulary, visual perception, motor speed)	Linear mixed effects regression	Mean dose to hippocampus and temporal lobes	Intuition	NA
2018	Zureick et al. [55]	70 paediatric cases (Brain)	Memory (visual and verbal)	Linear regression	V_{20} for the left hoppocampal region, gender	Bayesian information criterion	NA
				Hypothyroidism			
2012	Cella et al. [56]	53 (HL)	Thresholded Thyroid Stimulating Hormone	Logistic regression	V_{30} to thyroid, gender, volume of thyroid	Bootstrapping and forward variable selection. AUC	NA
2013	Rønjom et al. [57]	203 (HN)	Thresholded Thyroid Stimulating Hormone	Logistic regression, mixture model (for censored data)	Mean dose to thyroid, volume of thyroid, latency	Model selection using Akaike's information criterion	Externally validated by Rønjom et al. [58] (same findings)
				Esophagitis			
2012	Huang et al. [59]	374 (Lung)	Grade 2+ acute esophagitis	logistic regression	Mean dose and V_{30} to oesophagus	Bootstrapping and forward variable selection. AUC	NA
2015	Wijsman et al. [60]	53 (Lung)	Grade 2+ acute esophagitis	Logistic regression	Mean dose to oesophagus, stage, grade, CCT	Bootstrapping and forward variable selection. AUC	Externally validated by Dankers et al. [61] (same findings)

Table 3.2. A review of the existing literature for head and neck based Normal Tissue Complication Probability (NTCP) models built from Dose-Volume Histogram (DVH) statistics and patient covariates. Acronyms: Head and Neck (HN), Pharyngeal Constrictor Muscle (PCM), Area Under Curve (AUC), Radiotherapy (RT), Concurrent Chemotherapy (CCT), volume of organ that receives dose X (V_X).

freedom from the DVH of the segmented organ eliminates the uncertainty with respect to segmentations.

Typically, the voxel-based NTCP model follows this methodology:

1. Spatial normalisation of patient images. In this step, one patient is chosen from the population to be the common frame of reference, and all other patients are *registered* to this coordinate frame. Registration is the process of normalising/overlapping patient images/structures, and typically this is done using elastic Deformable Image Registration (DIR). The transformation required for each patient is saved.

2. The dose distributions for each patient are then transformed to the common frame of

reference. This is done by applying the transformation found using DIR, to the dose distribution of each patient.

3. Statistical analysis at each voxel. There are two main categories of voxel-based statistical analysis: two-sample hypothesis testing and regression analysis.

   - In hypothesis testing, the first step is to group each patients' voxel-dose data by their binary outcome label, then define a null hypothesis. The hypothesis is tested using a statistical hypothesis test, e.g. a two-sample t-test (parametric) or a two-sample Mann-Whitney U test (non-parametric), choosing the former if the data is normally distributed [65]. The result is a test statistic at each voxel, e.g. t-statistic if t-test was used. This is a univariate analysis, where only dose and outcome are investigated.

   - In a regression analysis, the voxel-dose data and other features are correlated with ordinal or continuous outcome labels, and voxel regression coefficients are calculated. The outcome labels can be defined at a specific time point, or even a Cox Proportional Hazards Model (CPHM) [12] can be fitted per-voxel if time-to-event data is available. The result is a set of regression coefficients and their standard errors at each voxel. This is a multivariate analysis, and the regression coefficients are adjusted for the other included features.

4. Infer a region where the dose is significantly correlated with outcome. However, to do this, one cannot interpret the significance of the test statistic/regression coefficient at each voxel due to the 'multiple comparisons problem'. That is, the more inferences that are made, the more likely erroneous inferences are to occur. To account for this, one can control for the family-wise error rate or by controlling for the false discovery rate [65, 66].

5. Having identified the significant region, the region is commonly propagated back to each patients' dose distribution, and dose summary statistics from a DVH such as the mean or maximum dose are computed. NTCP is then typically modelled using a logistic regression, using the dose summary dose statistics and other patient features. However, each regression performed at each voxel could also be combined in a statistically and biologically intuitive way, to model overall NTCP [67].

[65]. This procedure is illustrated in figure 3.4.

The limitations of this procedure are as follows:

- The normalisation of patients to a common frame of reference is a source of uncertainty. As a result of applying the uncertain transformation to each patients dose distribution, the transformed dose will be uncertain. Methods exist in literature to address this uncertainty, for example, McWilliam et al. [68] smoothed each transformed dose distribution using a 3D Gaussian kernel with a constant standard deviation, set to the

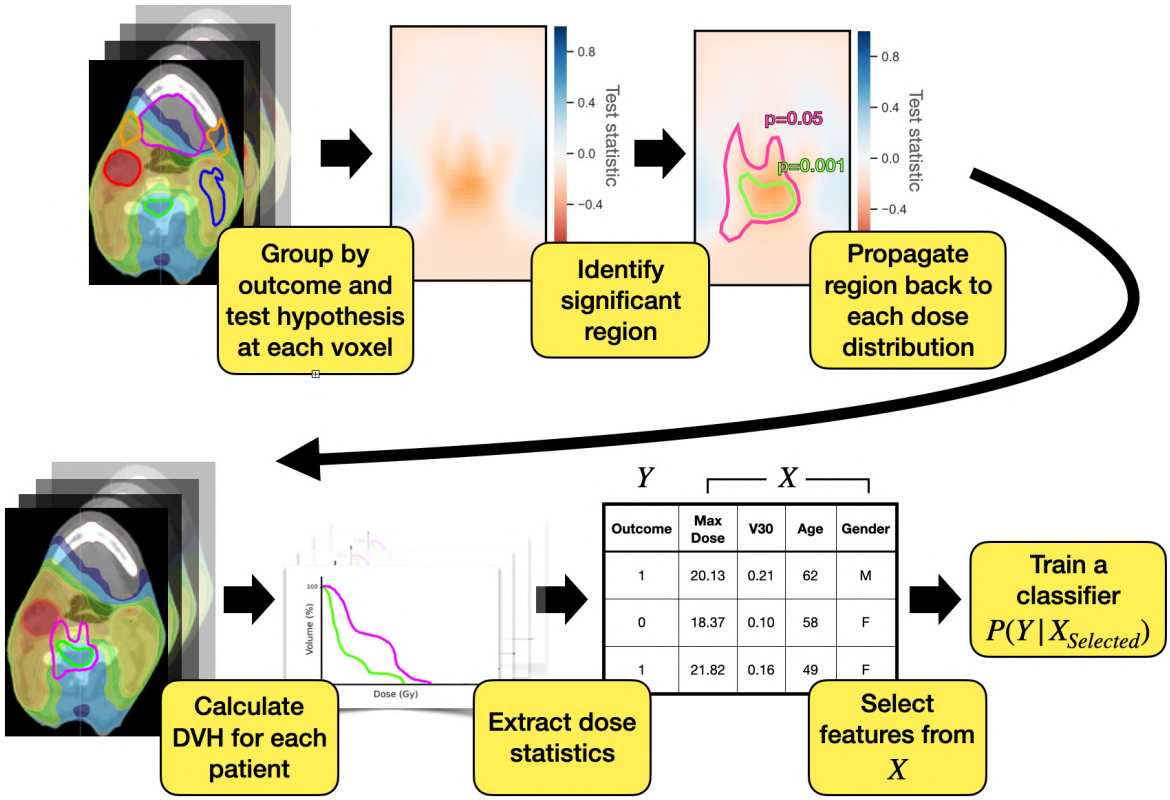


Figure 3.4. The typical process used in literature to create a multivariate Normal Tissue Complication Probability (NTCP) model from a voxel-based analysis of the dose distributions. Once the dose distributions for each patient are normalised to a common frame of reference, the dose distributions are typically grouped by their outcome. Next, the (defined) null hypothesis is tested at each voxel. A permutation test is typically conducted to control for multiple testing (controlling for the family wise error rate), to identify a significant region if present. This region is then propagated back to the normalised dose distributions, and dose statistics can be extracted for each patient.

standard deviation of the distance between centre of mass coordinates of a select structure of interest in the patient and in the common frame of reference.

- As for any multivariate model, care must be taken when defining the 'adjustment set', i.e. the features to include. Depending on what variables are included, or missing, could lead to meaningless and misleading adjusted coefficient estimates. This limitation is known as the 'Table 2 Fallacy' [69], and is a problem for inference and not prediction. For instance, if a multivariate regression is performed at each voxel, then our estimates of each adjusted regression coefficient and standard error is dependent on our choice of model, and if this choice is unrealistic, then the model can mislead our interpretation. To alleviate this, the data generating process can be modelled using a Structural Causal Model (SCM), and reliable and interpretable causal inferences can be made by defining an appropriate adjustment set [69, 70]. However, to the best of my knowledge, these techniques have not been adapted to NTCP modelling. To reinstate, voxel-based NTCP models are deemed to extract a high amount of interpretable information from the dose distributions, however one must take appropriate care when defining the adjustment set if they take a multivariate approach, else the method is arguably no more interpretable than black-box models.

- As each voxel is modelled independently, a voxel-based analysis is limited to identify-

ing highly localised linear group differences [71]. The technique generally struggles when the outcome of interest is spatially complex, multivariate and subtle. As a result, whilst the method may perform well on a population, it may struggle to perform well at the patient-level [72]. In addition, the region of interest identified in a voxel-based analysis has potential to overfit to the training data, and therefore it is crucial that studies are validated. To alleviate the above limitations, one model could be used to map the dose distribution to a prediction. For example, in neuroimaging, Carroll et al. [73] used each voxel in functional-MRI images as features, and used a linear Elastic Net (combination of Ridge and Lasso regularisation) to predict the mental state of patients. This regularisation method controls for the number of voxels in the resulting model, and the extent to which each voxel is included. As a result, the model proved highly predictive and able to detect complex spatial patterns driving prediction [73]. This method has since been applied to NTCP modelling by Jiang et al. [74], who found regions of the parotid glands correlated with xerostomia.

Table 3.4 highlights the existing literature for voxel-based NTCP models. The literature on this topic is becoming more common in recent years. In summary of table 3.4, 8/10 studies used univariate hypothesis testing to compare differences in dose at each voxel between each group, and of these 8, 7 corrected for the family-wise error-rate to control for multiple testing. The aim of 6/10 of the studies was to identify a region linked to the outcome, the remaining 4 studies modelled NTCP by extracting dose statistics from the significant region. Of these 4 studies, 2 identified significant dosimetric variables using a CPHM, whilst, Guo et al. [75] and Palma et al. [67] sought to leverage and combine the dose from each voxel into one NTCP model. Unfortunately, none of the studies were found to be externally validated yet.

Date	Authors	Patients	Outcomes	Registration	Method	Results
				Head and neck		
2017	Monti et al. [76]	42	Dysphagia	Log-diffeomorphic demons [77]	TFCE of dose differences between groups	Regions identified. Cricopharyngeus muscle and cervical esophagus
2018	Beasley et al. [78]	86	Trismus (incisor-to-incisor opening distance at 6 months)	RIR followed by DIR (B-splines) [79]	PT of Spearman rank correlation coefficient per-voxel	Region identified. Ipsilateral masseter
2020	Guo et al. [75]	146	Xerostomia (grade 2+ at 18 months)	Coherent Point Drift [80]	PT of dose differences between groups. LR with Ridge per-voxel in region for NTCP	Good AUC. Mapped radiosensitivity in region. Contralateral sides of parotid
				Prostate		
2010	Witte et al. [64]	352	Recurrence at 48 months	NA. Mapped points as distance and direction from prostate	PT of t-test per-voxel. Cox-regression of dose statistics in region and covariates	Regions identified. Obturatorial and presacral of prostate
2013	Acosta et al. [81]	105	Rectal bleeding (grade 1+ at 24 months)	Demons [82]	T-test per-voxel. No correction for multiple testing	Regions identified. Anterior rectal wall
2019	Mylona et al. [83]	272	Urinary complication (before and after 3 months)	Laplacian-based DIR [84]	PT of Mann-Whitney U test statistic per voxel	Regions identified. Bladder and the urethra
				Lung		
2016	Palma et al. [85]	98	RILD	Log-diffeomorphic demons [77]	TFCE of dose differences between groups	Regions identified. Peripheral medial-basal lung
2017	McWilliam et al. [9]	1101	Survival	DIR (B-splines) [79]	PT of t-test per-voxel. Cox-regression of dose statistics in region and covariates	Mean dose to base of the heart was significant
2019	Palma et al. [67]	98	Lung fibrosis	Log-diffeomorphic demons [77]	Logistic regression per voxel. Combine models with similar method to LKB for NTCP	Good AUC scores
2019	Palma et al. [86]	178	Radiation pneumonitis	DIR (B-splines) [79]	PT of t-test per-voxel	Regions identified. Lower lungs and in heart

Table 3.4. A review of the existing literature for voxel-based Normal Tissue Complication Probability (NTCP) models and data mining results. Acronyms: Threshold Free Cluster Enhancement (TFCE), Rigid Image Registration (RIR), Deformable Image Registration (DIR), Permutation Test (PT), Area Under Curve (AUC), Logistic Regression (LR), Radiation-Induced Lung Disease (RILD), Lyman-Kutcher-Burman (LKB).

Chapter 4

Causal Inference and Sparsity

Sparsity is useful tool for efficient estimation in the presence of noise. Sparsity is an assumption applied to systems with a greater number of features, p, than observations, n, where many of the input features have no important effect on the outcome of interest. The assumption of sparsity is that the true model has k important features, where $k < n$. Sparse statistical models aim to identify the correct sub-set of k important features from p, during optimisation of the model for predictive performance. If I attempt to model a high-dimensional system ($p >> n$) using a regular dense model, where each feature has an influence on the outcome variable, then the number of observations is too small to accurately estimate each parameter in the model. Therefore, sparsity, if a valid assumption, enables efficient and accurate parameter estimation in the presence of noise.

One of the main assumptions in causal inference is conditional ignorability, i.e. no hidden confounders [70]. However, for causal inference in high-dimensional systems the underlying mechanism is often complex, and can sometimes be unknown. Therefore, identification of confounders could be prevented. In this setting, sparse models and their variable selection properties could be put to use. In theory, and under specific conditions, it may be possible for sparse models to identify the correct sparse representation of variables and satisfy conditional ignorability. By reducing the dimensionality and satisfying conditional ignorability, sparse methods could enable efficient causal effect estimation in high-dimensional systems. However, it should be noted that for observational studies one can never be certain that conditional ignorability is satisfied; in this case, one can check for the presence and extent of hidden confounding using a sensitivity analysis.

4.1 Introduction to causal inference

A treatment, A, has a causal effect on an outcome, Y, if interventions on the treatment variable change the value of Y. The ultimate aim of causal inference is to estimate the unbiased effect of an intervention for an individual. To do this, the potential outcomes theory is commonly used [87]. There are many different potential outcomes that depend on the finite values of the treatment. I write this as $Y^{A=a}$, which reads as 'the outcome when I intervene to set $A = a$'. For example, in the case of a binary treatment, $A = 1$ for treatment and $A = 0$ for no treatment, with potential outcomes $Y^{A=1}$ and $Y^{A=0}$, respectively. The Individualised Treatment Effect (ITE) is formally defined as: the treatment A has a causal effect

on an individual's outcome Y if $Y^{A=a_0} \neq Y^{A=a_1}$, i.e. the potential outcomes of competing treatments differ [87].

There exists a fundamental problem in ITE estimation: we only ever observe a single potential outcome for an individual, that is, for the treatment they were assigned. Therefore, instead of ITE estimation, the ATE is typically estimated, i.e. the average effect of a treatment on a population. A treatment A is said to have a finite ATE on Y if $\mathbb{E}[Y^{A=a_0}] \neq \mathbb{E}[Y^{A=a_1}]$, where expectation is over the specific population under consideration. If the effect to be estimated is heterogeneous with respect to what we are conditioning on, then the Conditional Average Treatment Effect (CATE) estimator is used.

Whilst the potential outcomes framework holds for finite treatments, e.g. binary treatments, the theory gets complicated when the treatment is continuous valued, e.g. a dosage; as in this situation there are infinite potential outcomes. To estimate CATE for continuous treatments, outcome regressions and SCMs are typically used. Whilst Direct Acyclic Graphs (DAGs) represent the conditional independence properties between variables, SCMs are the equations that relate the variables. In contrast to potential outcomes theory, ATE estimation is possible using SCMs by leveraging specific adjustment criterions and *do-calculus*, as invented by Pearl [88].

4.1.1 Structural Causal Models

To show how SCMs can be adapted to continuous treatments, consider the following example as illustrated in the DAG in figure 4.1. Here, A represents the treatment variable, C represents a confounding variable, M represents a mediating variable and Y represents the outcome variable. In an SCM of the DAG in figure 4.1, the data is assumed to be described using the following set of non-parametric functions,

$$\begin{aligned} C &= f_C(U_C), \\ A &= f_A(C, U_A), \\ M &= f_M(A, U_M), \\ Y &= f_Y(M, C, U_Y), \end{aligned} \tag{4.1}$$

where each function is written as a function of its parents (variables that influence the current variable, symbolised by an arrow pointing into the current variable) in the DAG and independent random noise. U_C, U_A, U_M and U_Y are random variables that affect the covariate, treatment, mediator and outcome, respectively; and are assumed to be jointly independent but arbitrarily distributed [89]. The system represented by these functions is known as *structural* if each function is invariant to changes in the form of the other functions [89]. Pearl's do-calculus represents interventions, e.g. if I want to intervene on the system by setting $A = a_0$, which is represented by the do operator, $do(A = a_0)$. In this example, the

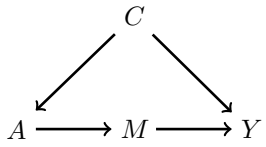


Figure 4.1. A Direct Acyclic Graph (DAG) illustrating the directed relationships between variables A, representing a treatment variable, C representing a confounding variable, M representing a mediating variable and Y representing an outcome variable.

expected outcome under this intervention can be written as $\mathbb{E}[Y|do(A = a_0)]$. This intervention would change the equations in 4.1 to the following,

$$\begin{aligned} C &= f_C(U_C), \\ A &= a_0, \\ M &= f_M(A, U_M), \\ Y &= f_Y(M, C, U_Y). \end{aligned} \tag{4.2}$$

Note that making this intervention changes the joint distribution of the data to what is known as the *interventional distribution*.

To illustrate how the ATE can be inferred using a SCM, let us look again at the functions in equation 4.1 and the DAG in figure 4.1. But now, let us assume the outcomes can be described using a simple linear model,

$$Y = f_Y(M, C, U_Y) = \alpha M + \beta C + U_Y. \tag{4.3}$$

Assuming I am interested in determining the ATE of M on Y, I can write the CATE as

$$\begin{aligned} \text{CATE}(C) :&= \mathbb{E}[Y|do(M = m_0), C] - \mathbb{E}[Y|do(M = m_1), C] \\ &= \mathbb{E}[(\alpha m_0 + \beta C + U_0) - (\alpha m_1 + \beta C + U_1)] \\ &= \mathbb{E}[\alpha(m_0 - m_1) + U_0 + U_1] \\ &= \alpha(m_0 - m_1). \end{aligned} \tag{4.4}$$

From this, I can then calculate the ATE by taking the expectation over C,

$$\begin{aligned} \text{ATE} :&= \mathbb{E}_{\mathbb{C}}[\text{CATE}(C)] \\ &= \mathbb{E}_{\mathbb{C}}[\alpha(m_0 - m_1)] \\ &= \alpha(m_0 - m_1). \end{aligned} \tag{4.5}$$

In the case that M is binary ($m_0 = 1$ and $m_1 = 0$), $ATE = \alpha$. If M is continuous, and $m_0 = m_1 + 1$, then $ATE = \alpha$; where α represents the average effect of increasing M by 1 unit. Therefore, in the special case that the outcome function in the SCM can be described using a simple linear model, the regression coefficients can be interpreted causally. If I describe a system which is non-linear using linearity assumptions, then this can significantly

impact our estimate of the ATE. For example, if I use equation 4.3 in our SCM, but in fact, the outcomes are generated via the function

$$Y = f_Y(M, C, U_Y) = \alpha M + \beta C + \gamma C^2 + U_Y, \tag{4.6}$$

where the outcome is now a non-linear function of C. Then, our estimate of the ATE of M on Y, $\hat{\alpha}$, under our hypothesis of equation 4.3, is given by

$$\hat{\alpha} = \alpha + \gamma \frac{\mathbb{E}[CM]\mathbb{E}[C^2] - \mathbb{E}[M^2]\mathbb{E}[C^2 M]}{\mathbb{E}[CM]^2 - \mathbb{E}[C^2]\mathbb{E}[M^2]}. \tag{4.7}$$

That is, our estimate of the true ATE, α, can be made arbitrarily larger or smaller depending on γ [90]. Therefore, the parametric form of the models used is of great importance. To extend causal inference to the non-linear regime, a plethora of non-parametric methods exist, including: Hill [91] introduced Bayesian Additive Regression Trees (BART) which can be used for ATE inference from discrete and continuous treatment variables; Athey and Imbens [92] and Wager and Athey [93] use random forests; Hoyer et al. [94] and Zigler et al. [95] use Gaussian Processes; and Johansson et al. [96], Shalit et al. [97], and Lopez-Paz et al. [98] use neural networks.

4.2 Criterions for identifiability

A causal relationship is said to be identifiable if it can be estimated consistently from observational data with an arbitrarily large sample size. In the previous section, 4.1.1, I discussed how CATE and ATE can be estimated from data using SCMs, with parametric, semi-parametric and non-parametric methods. However, in order for the estimation methods detailed in the previous section to work correctly when confounding is present, I must define a set of variables to include in the model and adjust for, i.e. an adjustment set. Pearl [99] and Sewall Wright [100] introduced three different criterions to identify the adjustment set from a DAG: the Back-Door criterion, the Front-Door criterion, and instrumental variables.

4.2.1 The Back-Door criterion

Pearl [99] defines a Back-Door path as any path from X to Y that starts with an arrow pointing into X. X and Y are said to be de-confounded if we block every Back-Door path between them. To illustrate this criterion, I look again at our example DAG in figure 4.1. If I am interested in the effect of A on Y, I notice from the DAG in figure 4.1 that the path $A \leftarrow C \rightarrow Y$ is a Back-Door path. Information can flow from A to Y through this Back-Door path, and this is responsible for the spurious component of the relationship between A and Y. Therefore, I can adjust for C in order to de-confound A and Y.

In general, an adjustment set Z is said to satisfy the Back-Door criterion when:

1. Z blocks every Back-Door path between X and Y;

2. No variable in Z is a descendant of X.

The latter condition is important because adjusting for a descendant of X could partially or completely remove the effect we are trying to measure. Mathematically, the Back-Door criterion is expressed as

$$P(Y|do(X)) = \sum_{z} P(Y|X = x, Z = z)P(Z = z), \tag{4.8}$$

where I confirm the causal effect is identifiable due to our ability to express the causal estimand (left-hand side) in terms of observational distributions (right-hand side) [99]. If the Back-Door criterion is satisfied, conditional ignorability in causal inference is also satisfied; they are mirrors of each other [70].

Before continuing, it is important to define collider bias. Consider the scenario where I am interested in the causal effect of X on Y, but the following non-causal path exists $X \rightarrow C \leftarrow Y$. Notice the directionality of the connecting arrows, C is caused by both X and Y. Therefore, C contains information from X and Y, but if a causal path between X and Y does not exist and this is the only non-causal path, then X and Y are actually independent. This path is known as a collider, and if I conditioned on C I would induce a spurious correlation between X and Y. This type of spurious correlation is known as collider bias [101]. Now consider the scenario where I remain interested in the causal effect of X on Y, but a new non-causal path exists, $X \leftarrow Q \rightarrow W \leftarrow E \rightarrow Y$. Following our definition above, this non-causal path is a Back-Door path between X and Y, but how do I block it? The variables Q, W and E lay on this Back-Door path, but which should I include in the adjustment set? From our definition above, W is a collider. However, if I added W to the adjustment set, I induce collider bias; as by adjusting for W I allow information to flow between X and Y. To effectively block the Back-Door path in this example, I could either (1) do nothing, (2) add Q, E or both to the adjustment set, or (3) add W and either Q or E (or both) to the adjustment set. Visualising the assumed true causal structure using a DAG can help identify Back-Door paths, confounders and colliders; which can be leveraged to define the correct adjustment set.

4.2.2 The Front-Door criterion

The front-door criterion can be used when the effect of the treatment, X, on the outcome, Y, is mediated by another observed variable, M, and possible confounders of treatment and outcome are latent/unobserved. Because the confounders are latent, I cannot control for them using the Back-Door criterion as described above. Despite this seemingly tricky situation, the front-door criterion makes the causal estimand identifiable.

In general, an adjustment set Z is said to satisfy the front-door criterion when:

1. Z intercepts all directed paths from X to Y;

2. There is no Back-Door path from X to Z;

3. All Back-Door paths from Z to Y are blocked by X.

[99]. Mathematically, I express the front-door criterion as

$$P(Y|do(X)) = \sum_z P(Z = z|X = x) \sum_{x'} P(Y|X = x', Z = z)P(X = x'), \tag{4.9}$$

where the causal effect of X on Y is identifiable using the adjustment set Z that satisfies the front-door criterion [99]. In the example above, the adjustment set would consist of only the variable M.

4.2.3 Instrumental variables

A variable I is said to be an instrumental variable relative to a pair of variables (X, Y) if X and Y are generated from the DAG shown in figure 4.2, where U is a latent common cause of both X and Y [102]. In this scenario, the influence of I on Y is completely mediated by X. Therefore, if X is held constant, I and Y are independent. Instrumental variables are useful because by making Back-Door adjustments for U, $P(Y|do(I))$ and $P(X|do(I))$ become identifiable. In turn, $P(Y|do(X))$ is identifiable when the following equation is solved,

$$P(Y|do(I)) = \sum_x P(Y|do(X = x))P(X = x|do(I)). \tag{4.10}$$

Essentially, when this equation is solved, the instrumental variable is used to account for the effect of the latent variable U on X and Y [102]. Therefore, instrumental variables, if present, can act as a powerful method for causal inference with latent variables.

4.3 Sparse statistical models

So far, I have discussed causal inference for discrete and continuous treatments, however, the treatment could be high-dimensional. For example, radiotherapy treatment is defined as a 3D array of continuous dose values, $\mathbf{D}$, which has of order 10^6 features (voxels). In this example, $p = \mathcal{O}(10^6)$, which is magnitudes greater than the typical number of samples in voxel-based studies, $n = \mathcal{O}(10^2)$ (calculated from table 3.4); a high-dimensional system ($p >> n$). However, I discussed in section 3.3 that there is growing evidence to support the existence of radiosensitive sub-regions of organs that may drive outcomes [5, 6, 7, 8, 9], i.e.

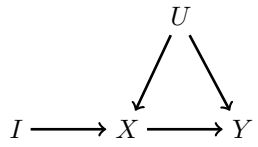


Figure 4.2. A Direct Acyclic Graph (DAG) illustrating the process that defines I as an instrumental variable. X is the treatment variable, Y is the outcome variable, and U is a latent common cause of X and Y.

only a subset of $\mathbf{D}$ is important. Therefore, radiotherapy treatment can be considered to be sparse with respect to specific outcomes.

Sparsity is a powerful tool for efficient causal inference in high-dimensional systems, enabling us to reduce the problem to a sub-set of the high-dimensional system that has an important influence on variables. So far, I have emphasised the importance of defining the correct adjustment set and form of the estimating equations for ATE inference using SCMs. However, for high-dimensional systems, it might not be possible to draw a DAG and precisely define the adjustment set. To illustrate this difficulty, consider the DAG in figure 4.3a, where I have a high-dimensional treatment, $\mathbf{A}$, a confounding variable, C, and an outcome Y. It is straight forward to define this high-level DAG. However, if $\mathbf{A} = \{A_0, A_1, A_2\}$, I could re-draw the DAG in figure 4.3a as that shown in figure 4.3b, where the dashed arrows represent possible causal relationships between the variables. However, for particular systems, we may not have the knowledge required to define with certainty which arrows exist. For example, in radiotherapy, I have the high-level knowledge that treatment affects outcome, but I am unsure what parts of the treatment are responsible. In this case, one could use a non-parametric approach by assuming that all possible connections exist, and then use the DAG and the criterions for identifiability as discussed in section 4.2 to define an initial adjustment set. A sparse model could then be used as the estimating equation for ATE inference in the SCM, which would perform simultaneous variable selection and estimation.

However, for a sparse model to be successful for causal inference it must be consistent in both variable selection and parameter estimation. This is known as the *oracle* property; an important property for unbiased estimation using sparse models. Mathematically, these conditions are written as follows:

1. Consistency in variable selection: $\{j : \hat{\beta}_j \neq 0\} = \mathcal{A}$;
2. Consistency in parameter estimation: $\sqrt{n}(\hat{\boldsymbol{\beta}}_{\mathcal{A}} - \boldsymbol{\beta}_{\mathcal{A}}) \rightarrow_d N(\mathbf{0}, \boldsymbol{\Sigma}^*)$.

Condition (1) states that the set of estimated non-zero parameters, $\{j : \hat{\beta}_j \neq 0\}$, is equal to the true set of non-zero parameters, $\mathcal{A}$; and condition (2), otherwise known as *asymptotic normality*, states that the difference in estimated and true parameters tends to a mean-zero normal distribution as $n \rightarrow \infty$, where $\boldsymbol{\Sigma}^*$ is the covariance matrix under the true model [103]. Note that, if the true model is linear, the true non-zero parameters, $\mathcal{A}$, can

Figure 4.3. (a) A high-level Direct Acyclic Graph (DAG) consisting of a confounding variable, C, high-dimensional treatment, $\mathbf{A}$, and outcome variable, Y. (b) A low-level DAG of the same process as (a) where $\mathbf{A} = \{A_0, A_1, A_2\}$. The dashed arrows represent possible causal relationships between the variables C, Y and the features of $\mathbf{A}$.

correspond to the identified causal effects. However, if the true model is non-linear, I could identify $\mathcal{A}$, but I would need a non-parametric method to identify the causal effects; as explained in section 4.1.1.

4.3.1 The Lasso

A simple example of a sparse model is the Lasso regression. The Lasso, otherwise known as l_1 regularisation, is an additional term in the loss function that constrains the magnitude of the parameter estimates [104]. In contrast, the Ridge, otherwise known as l_2 regularisation, is an additional term in the loss function which differs from the Lasso by penalising the square of the estimated parameters [105]. In both cases parameter estimation is a convex problem and depends on both the data and the regularisation, in fact, the estimates with lowest loss occur at equilibrium between the two. However, because the Lasso penalises parameter magnitudes, rather than penalising square magnitudes like the Ridge, the shape of the loss function allows shrinkage of parameter estimates to exactly zero. As a result, the Lasso is a popular technique for simultaneous variable selection and estimation [104]. In the simple case of a linear model, the parameters of a Lasso regression can be estimated from the following loss function,

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \left(\sum_{i=1}^{n} (y_i - \sum_{j=1}^{p} X_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right), \tag{4.11}$$

where $\hat{\boldsymbol{\beta}} \in \mathbb{R}^p$ is a vector of estimated model parameters, $\mathbf{y} \in \mathbb{R}^n$ is the outcome vector, $\mathbf{X} \in \mathbb{R}^{n \times p}$ is the feature matrix, and the additional term $\lambda \sum_{j=1}^{p} |\beta_j|$ corresponds to the Lasso penalty, where λ is a tuneable hyper-parameter [104, 106].

The Lasso regression encourages sparsity, however, Zou [103] showed that the Lasso has the oracle property only under very specific conditions. In addition, Meinshausen and Bühlmann [107] showed that the Lasso is inconsistent in variable selection when the hyper-parameter, λ, is optimised for predictive performance; and Fan and Li [108] showed that the Lasso is inconsistent in parameter estimation, as it leads to biased estimation of large parameters. Therefore, the Lasso is generally considered to not have the oracle property.

4.3.2 The Adaptive Lasso

The Adaptive Lasso was developed by Zou [103] to give the Lasso the oracle property. The difference between the Adaptive Lasso and the original Lasso, is the idea to penalise each parameter differently. The parameters of the Adaptive Lasso can be estimated from the following loss function,

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \left(\sum_{i=1}^{n} (y_i - \sum_{j=1}^{p} X_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} w_j |\beta_j| \right), \tag{4.12}$$

where in comparison to equation 4.11, the adaptive weight vector $\mathbf{w} \in \mathbb{R}^p$ has been added to the Lasso penalty term. The adaptive weights are estimated from data as,

$$\hat{\mathbf{w}} = \frac{1}{|\hat{\boldsymbol{\beta}}^*|^\gamma}, \tag{4.13}$$

where $\gamma > 0$ is an additional hyper-parameter, and $\hat{\boldsymbol{\beta}}^*$ are the parameter estimates from an additional regression that is consistent in parameter estimation. Under these conditions, Zou [103] showed that the Adaptive Lasso has the oracle property.

In general, any consistent estimator can be used to calculate $\hat{\boldsymbol{\beta}}^*$, however, Zou [103] recommend using an ordinary sum of squares regression, or the best performing Ridge regression if multicollinearity between features is present. The Ridge regression differs from the Lasso by penalising the square of the parameter estimates. This will heavily penalise exploding parameter estimates, which multicollinearity is known to cause. Whilst the Ridge regression is inconsistent in parameter estimation, i.e. it is a biased estimator, it does reduce the variance of the model [105].

Chapter 5

Average Treatment Effect Estimation for High-Dimensional Spatial Treatments: Radiotherapy Simulation Study

This chapter is the bases of a publication and it is structured as such.

Authors

Mr Alexander L. Jenkins[1], Dr Eliana Vásquez Osorio[1 2], Dr Michael Merchant[1 2], Dr Andrew Green[1 2], Professor Marcel van Herk[1 2], Dr Matthew Sperrin[1 3], Dr Alan McWilliam[1 2 3].

5.1 Abstract

Radiotherapy is the most common treatment for cancer, delivering 3D, personalised radiation dose to the tumour. Radiotherapy planning requires considering a high-dimensional continuous optimisation space to achieve tumour control while limiting the probability of treatment complications. However, estimating the Average Treatment Effect (ATE) of radiation dose on complications across the anatomy is difficult; mainly because of confounding in observational data. Under certain assumptions, developing a causal framework provides methods to adjust for confounding.

The aim of this work is to use simulated data to investigate if unbiased and consistent voxel-based causal inference is possible, how, under what circumstances, and with what accuracy. I simulate radiotherapy treatment plans from a simplified, yet realistic, data generating process. Patients have a single tumour (random location) where dose is maximal and a single Organ At Risk (OAR) (fixed location) where dose is minimal. Variables control fall-off of dose around the tumour, fall-off and magnitude of dose at the organ, and a covariate that confounds the treatment plan. I simulate realistic treatment uncertainties: random shifts in x- and y-directions of the entire planned dose distribution, spatially correlated

[1]The University of Manchester, United Kingdom.
[2]The Christie NHS Foundation Trust, United Kingdom.
[3]Joint senior authors.

noise sampled from a Gaussian process prior, and independent noise at each pixel. A continuous complication is generated via a linear function of the delivered dose to a spatially inhomogeneous set of pixels (ATE estimand), a covariate that also affects the delivered dose distribution, and a spatially inhomogeneous interaction between delivered dose and the covariate. Three methods based in causal inference are used to estimate the ATE at each pixel: 1) pixel-wise sparse causal regression, 2) sparse causal regression and 3) a causal regression. The sparse estimator used is the Adaptive Lasso. These are compared to methods currently used in radiotherapy.

I found that all methods based in causal inference performed with lower total Mean Squared Error (MSE), MSE_{tot}, across all parameterisations tested in the simulation compared to the currently used voxel-based statistical methods in radiotherapy. Exploiting the oracle property of the Adaptive Lasso to simultaneously identify important pixels with dose-response and estimate ATE, was in general a successful technique over all parameterisations of the simulation tested. The only method capable of unbiased estimation was the causal regression, however, multicollinearity hinders accurate ATE estimation at specific regions of parameter space and at a high resolution. The estimation method that scored a consistently low MSE_{tot} over all parameterisations was the sparse causal regression. This method was able to assign a near zero effect to unimportant pixels, and whilst estimates elsewhere were biased, they were accurate and efficient; especially at lower resolutions.

5.2 Introduction

Radiotherapy is the most affordable and accessible form of cancer treatment [1]. It is, therefore, the most common form of cancer treatment, with approximately 50% of cancer patients receiving it during their course of treatment [2]. Advancements in imaging, e.g. 3D Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), alongside advancements in treatment delivery and technology, e.g. image-guided radiotherapy, mean that uncertainties in radiotherapy treatment are reducing. As a result of technological developments in imaging and treatment delivery allowing for a more conformed dose to the tumour, 5-year survival rates have been increasing over the past 25 years for most tumour sites [3]. Therefore, key research questions now focus on reducing treatment complications and improving the quality of life for patients living with and beyond cancer.

Many complications of radiotherapy treatment are considered to be caused by the dose to OARs. In current clinical practice, dose to OARs is typically quantified using a Dose-Volume Histogram (DVH), where each OAR has a guideline for absorbed dose limits and the expected risk of complications, as stated in the QUANTEC guidelines [4]. However, because the QUANTEC guidelines leverage the DVH, the following assumptions apply:

1. A single treatment complication is driven only by a single organ. Yet, there is growing evidence to support that the onset of complications can be driven by dose to multiple

regions [10].

2. The whole organ equally drives the complication. Yet, this is known to be an oversimplification of the radiobiology within the organ [10].
3. The spatial dose distribution within the organ volume is ignored. Yet, many different spatial dose distributions can give rise to the same DVH.

Voxel-based methods seek to identify regions of an image which are significantly associated with a continuous / binary label of interest. Where a voxel is a point / tiny volume in 3D space. Voxel-based methods have been well established in the field of neuroimaging since the 1990's [62]. For example, Whitwell [63] used the methodology to find regions of MRI images associated with Alzheimer's disease. The potential of the methodology was fast to adapt to the field of radiation oncology, with the first voxel-based analysis emerging in 2010 by Witte et al. [64], where the authors related regions of the dose distribution to recurrence of prostate cancer. Since then, voxel-based methods have been applied to multiple treatment sites, and there is growing evidence to support the existence of sub-regions of organs that radiation dose is linked to complications, as shown for the bladder [5], the parotid glands [6], white-matter structures [7], the lung [8] and the heart [9]. Recently, a 'cook book' for voxel-based methods in radiotherapy has been published [65], which illustrates the importance of this technique to radiation oncology.

However, as each voxel is modelled independently, a voxel-based analysis is limited to identifying highly localised linear group differences [71]. The technique generally struggles when the outcome of interest is spatially complex, multivariate and subtle. To overcome this, a single model could be used to map the dose distribution to a prediction. For example, in neuroimaging, Carroll et al. [73] used each voxel in functional-MRI images as features, and used a linear Elastic Net (combination of Ridge and Lasso regularisation) to predict the mental state of patients. This regularisation method controls for the number of voxels in the resulting model, and the extent to which each voxel is included. As a result, the model proved highly predictive and able to detect complex spatial patterns driving prediction [73]. This method has since been applied to radiotherapy by Jiang et al. [74], who found regions of the parotid glands linked with xerostomia.

Alongside using the voxel-based methods for identifying sub-regions of organs, the regression coefficients at each voxel are often interpreted [12, 13]. This is because a 3D map of the regression coefficients is highly desirable, as it gives an insight into the spatial effect of radiotherapy treatment on a given outcome; which could be used in the future to optimise the treatment plan for reduction of complications. Despite the success of voxel-based methods, and even Carroll et al. [73]'s Elastic Net, the predictions nor the parameters of algorithms optimised for predictive performance have a causal interpretation. In fact, many sources of bias can exist in observational data, for example, confounding bias, collider bias and selection bias [109]. A predictive algorithm is unable to adjust for these biases outside of a causal framework, in fact, it could even leverage them for prediction. Therefore, in order to use prediction algorithms for unbiased estimation of the causal effect of radiotherapy

treatment on the outcome of interest at each voxel or at the identified region of the dose distribution, a causal framework must be developed. In radiotherapy, the main source of bias is confounding, i.e. variables that affect the treatment and also affect the treatment response. Current voxel-based methods based on statistical inference do not adequately adjust for confounding. This could explain the variability in the location of radiosensitive sub-regions identified for OARs, for example, McWilliam et al. [9] and Bogaard et al. [14] identified unique sub-regions of the heart that is particularly radiosensitive. It also could explain why models of the same outcome differ in their identified significant variables and their effect, and hence, why the external validation of current NTCP models is so important [15, 16]. Therefore, it is of great importance to try to estimate the spatial causal effect of radiotherapy treatment on a given outcome.

The gold standard causal framework is typically the Randomised Controlled Trial (RCT). In an RCT, treatment and control groups are formed by random assignment. As a result, the two groups are seen as balanced in terms of potential outcomes. In an RCT, one can infer the ATE of treatment A versus treatment B on an outcome of interest with respect to the RCT population, where treatment A and B could be exposure versus no exposure, two competing treatments, or the same treatment modality but at different doses. However, using an RCT to identify the effect of radiotherapy treatment at different anatomical regions is not ethical. The planned dose distribution should always be tailored to the patient characteristics, in order to give the best chances of curative treatment whilst reducing probability of complications. Therefore, under this ethical constraint, a causal framework must be developed using observational data.

5.3 Simulation study

5.3.1 Aims

The aim of this work is to use simulated data to investigate if unbiased and consistent voxel-based causal inference is possible, how, under what circumstances, and with what accuracy. To do this, I simulate data from a simplified, yet realistic, data generating process for radiotherapy treatment-outcome data. However, assumptions must be made that reduce the complexity of the data, whilst retaining its defining properties.

From an analytical perspective, the defining properties of the radiotherapy dose distribution is that there are multiple voxels, that dose at each voxel is continuous valued, and that dose at adjacent voxels is correlated [110]. It is crucial that the simulated dose distributions retain these properties. The major challenge from the causal inference perspective is constructing an estimation method to work under the properties of this complicated data.

From a clinical perspective, the properties to consider are firstly that the dose distribution has high and low dose regions. The tumour volume/s have the highest prescribed dose, and the OARs have the lowest prescribed dose. However, because dose is accumulated to achieve

the prescription: if I reduce dose in one region, it must increase in another. Therefore, this conservation property of the dose distribution must also be represented in the simulated data.

It is also important to consider the difference between the *planned dose distribution* and the *delivered dose distribution*. The planned dose distribution is designed by the clinician and treatment planning staff, and it defines the pre-treatment prescription. The delivered dose distribution is the dose that the patient actually received at the end of their course of treatment. The delivered dose distribution differs to the planned dose distribution through numerous dosimetric and geometric uncertainties. Whilst the planned dose distribution is designed to limit the effect these uncertainties have on changing the prescription for a population [111], the delivered dose distribution is known to be a better predictor of outcomes [112]. In fact, the delivered dose is the cause of radiation-induced outcomes, as it is the dose that the patient actually receives. In addition, radiation-induced outcomes are not just generated by the delivered dose, they could be generated by a complex function of covariates and corresponding interaction terms with the delivered dose [110].

It is also important to consider the mechanism for how clinical variables affect the dose distribution. Planned dose distributions are created by the clinician and treatment planning staff, and therefore, the individual acts as a mediator between the clinical variables and the resultant planned dose distribution. These individuals can intervene and affect the dose distribution in the following ways:

- Segmentations of tumour and OARs. Segmented regions could be extended based on the patients diagnosis, for example, the volume of the neck irradiated for head and neck cancer patients depends on the stage of the patients' cancer [113]. It is also known that segmentations differ between individuals and even between the same individual on two different occasions, this is known as inter- and intra-observer variation, and it has been hypothesised that this can depend on factors such as the imaging modality [114], a variable knowledge / interpretation of volume definitions [115], and segmentation protocol [116].
- Planned dose distribution parameters. For example, the individual can adjust the DVH statistics of each OAR to find the plan that gives the best tumour coverage and lowest dose to the organs whilst following the QUANTEC guidelines and considering the clinical knowledge of the patient, e.g. frailty, comorbidities, etc. However, the dose prescribed to the tumour, and the fractionation schedule, are typically set through latest RCT evidence.

5.3.2 Data generating process

To create a realistic, yet simplified, data generating process for radiotherapy treatment-outcome data that retains the defining properties discussed in the previous section, I make the following assumptions:

1. I will work with 2D, rather than 3D dose distributions. By working with pixels rather than voxels, the number of features in the dose distribution is reduced exponentially. This allows for faster simulation and estimation times, whilst being able to incorporate the defining complexities of the problem discussed in section 5.3.1.
2. Each pixel will represent the same region of the anatomy for all patients. This is the same assumption that is used in current voxel-based methods, where patients are placed into a common reference frame using techniques in image registration [65].
3. Each patient will have one tumour centred at a random pixel in the image (sampled independently with uniform probability).
4. All patients will have one organ at risk centred at a common pixel in the image. If the tumour satisfies the same location as the organ for a patient, the tumour has priority and the organ will be ignored.
5. Tumour and organ are of constant size and shape: a single pixel. For simplicity, I ignore differences in the shapes and sizes of the tumour and organ for all patients, alongside segmentation uncertainty.
6. The prescribed dose to the tumour is constant across the simulated population, with value $T_D = 4$ [arb. units]. In addition, I will assume this dose is delivered in a single fraction.
7. The planned dose distribution is created in the style of Volumetric Modulated Arc Therapy (VMAT) treatment, where radiation is accumulated at the tumour by varying dose in a continual arc around the patient. With this in mind, the planned dose at each pixel will be decided as a smooth function of the distance from the tumour and the distance from the *organ ray*. An illustration and mathematical definition will follow in section 5.3.4.
8. The outcome variable will be continuous random variable.
9. The outcome variable will be generated by the delivered dose to sensitive regions of the anatomy (constant in both location and effect during the simulation), an interaction between delivered dose and a covariate, alongside the covariate. Therefore, the system will be described as *sparse*, i.e. only a sub-set of the features (each pixel in the delivered dose distribution) have an important effect on the outcome.
10. The delivered dose distribution causes the outcome. Delivered dose distributions are simulated with realistic treatment uncertainties operating at three different scales: independent noise at each pixel (dosimetric uncertainty), correlated noise sampled from a Gaussian Process prior (anatomical motion), and random shifts in x and y of the entire dose distribution (setup uncertainty).

The remaining assumptions are on the cause-effect relationships between variables. These are encapsulated by the DAG shown in figure 5.1, which will be the assumed causal structure for this simulation. It is drawn for only three pixels in this figure, but extends over all

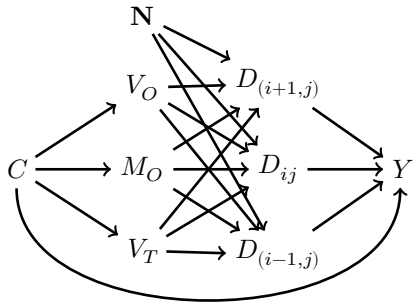
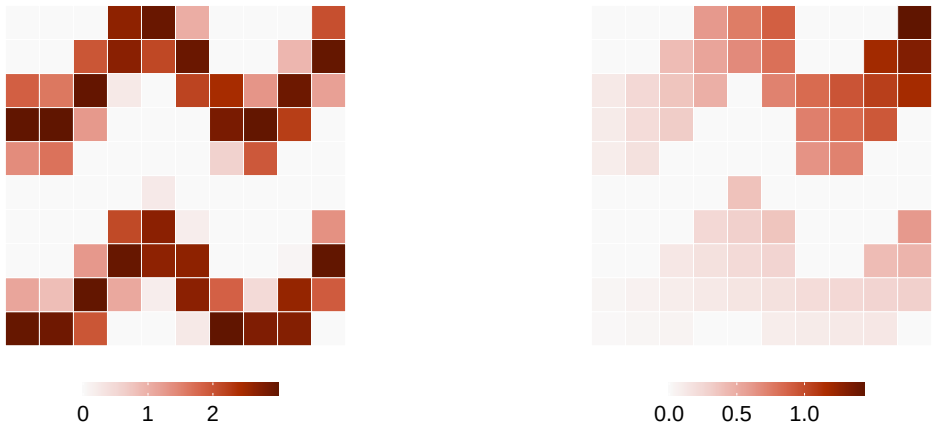


Figure 5.1. The causal structure assumed in this simulation, represented using a Direct Acyclic Graph (DAG). The estimand of interest is the Average Treatment Effect (ATE) of the delivered dose, D_{ij}, on the outcome, Y, at each pixel $i \in [1, N_x]$ and $j \in [1, N_y]$, where N_x and N_y is the number of pixels in x and y. C is a measured confounder, V_O is a parameter that controls the fall-off of dose around the organ ray, M_O is a parameter that controls for the magnitude of dose on the organ ray, V_T is a parameter that controls that fall-off of dose around the tumour, D_{ij} is the value of the delivered dose distribution at pixel location $x = i, y = j$, and $\mathbf{N}$ is a latent variable representing the parameters generating the treatment uncertainties. The DAG is drawn only for three pixels in this figure, but extends over all pixels in this work.

pixels in this work. The variables in figure 5.1 are defined as follows: Y is the continuous outcome variable of interest, C is a covariate acting as a measured confounder, V_O is a parameter that controls the fall-off of dose around the organ ray, M_O is a parameter that controls for the magnitude of dose on the organ ray, V_T is a parameter that controls that fall-off of dose around the tumour, D_{ij} is the value of the delivered dose distribution at pixel location $x = i, y = j$, and $\mathbf{N}$ is a latent variable representing the parameters generating the treatment uncertainties. $\mathbf{N}$ is assumed latent because whilst $\mathbf{D}$ can be estimated in reality, it is difficult to quantify the exact values to adjust for at each pixel. V_O, M_O and V_T represent three possible variables the clinician and treatment planning staff could intervene on in practice during dose planning; each is affected by C. C could represent a clinical variable such as age, and Y could represent a continuous outcome such as weight loss after treatment. The effect of V_O, M_O and V_T on Y is assumed to be mediated entirely by the delivered dose distribution. The high-dimensional delivered dose distribution, $\mathbf{D}$, has been expressed non-parametrically in figure 5.1 by assuming that all possible connections exist between Y, D_{ij}, V_O, M_O, V_T and $\mathbf{N}$.

5.3.3 Estimands

The estimands are the quantities to be estimated in the simulation. In this work, the estimands are defined as the ATE of the delivered dose, D_{ij}, on the outcome, Y, at each pixel $i \in [1, N_x]$ and $j \in [1, N_y]$, where N_x, N_y are the number of pixels in x and y. Mathematically, the estimands are expressed as array $\boldsymbol{\theta} \in \mathbb{R}^{N_x \times N_y}$. As mentioned in section 5.3.1, the clinician and treatment planning staff typically intervene on variables like V_O, M_O and V_T, so why am I interested in the ATE of D_{ij} (a variable of indirect control)? I am interested in the ATE of D_{ij} on Y because ultimately I am interested in the mechanism that causes the outcome, i.e. I am not just interested in determining, for example, the ATE of M_O on Y, but I want to know *how* M_O causes Y. The current possible interventions, such as a set similar to V_O, M_O and V_T, could be sub-optimal, and knowledge of the mechanism can help us



(a) (b)

Figure 5.2. (a) The estimand array produced using the function $f(x, y)$ defined in equation 5.1, where the number of pixels in the x- and y-directions is equal to $N_x = 10, N_y = 10$, respectively. (b) The interaction strength array produced using the function $h(x, y)$ defined in equation 5.5, where the number of pixels in the x- and y-directions is equal to $N_x = 10, N_y = 10$, respectively.

realise new interventions, for example, new dose interventions on previously unknown regions with given dose-response. Answers for this mechanistic question using observational data are not reliable outside of a causal framework.

The estimands are defined by the following function, $f(x, y)$, which is evaluated at the x- and y-coordinates of each pixel to define the estimand array, $\boldsymbol{\theta}$, in any required resolution,

$$
\begin{aligned}
g(x, y) &= 3 \sin\left(\frac{10x}{N_x} + \frac{3}{2}\sin\left(\frac{10y}{N_y}\right) - 0.5\right), \\
f(x, y) &= \begin{cases} g(x, y), & \text{if } g(x, y) > 0 \\ 0, & \text{otherwise.} \end{cases}
\end{aligned}
\tag{5.1}
$$

The function, $f(x, y)$, produces values in the range $[0, 3]$. Figure 5.2a shows the estimand array, $\boldsymbol{\theta} = \{\theta_{ij}\}$, calculated using equation 5.1, i.e. $\theta_{ij} = f(i, j)$, for $N_x, N_y = 10$. The estimand array will be constant for all patients generated in the simulation, and represents the spatial dose-response. The exact form of the estimand array, i.e. $f(x, y)$ in equation 5.1, is chosen such that the dose-response is spatially complex, inhomogeneous, positive (dose always increases the outcome value / odds), and sparse (zero effect in many pixels). In this work, $\boldsymbol{\theta}$ will be estimated using different methods in causal inference, and compared to the current voxel-based analysis methods in radiotherapy.

5.3.4 Simulation method

Generating equations

Each of the variables in the DAG assumed for this simulation, as shown in figure 5.1, are generated as a function of their parents and independent noise, i.e. using a Structural Causal

Model (SCM). The functions that will be used to generate each variable are defined as,

$$C = 3 + U_C, \tag{5.2a}$$

$$V_T = \frac{6C^2}{5} + 3 + U_{V_T}, \tag{5.2b}$$

$$V_O = (1 + \frac{U_{V_O} C}{15}) \times s(O_{x,y}, T_{x,y})^2, \tag{5.2c}$$

$$M_O = (\frac{C}{15} f_{D_T}(O_x, O_y) + U_{M_O}) \times (\frac{1}{2} + s(O_{x,y}, T_{x,y})), \tag{5.2d}$$

$$D_{ij} = P(i + L_x, j + L_y) + M_{ij} + U_{D_{ij}}, \tag{5.2e}$$

$$Y = 5C + U_Y + \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left(D_{ij}\theta_{ij} + \frac{C}{2}(D_{ij}\xi_{ij}) \right). \tag{5.2f}$$

Unless otherwise stated, the independent random noise for the i^{th} variable, U_i, is sampled from a mean-zero Gaussian distribution with unit variance, $U_i \sim N(0, 1)$. The form of these equations are designed such that they are mostly complex non-linear functions, as they may well be in reality, except for equation 5.2f which is chosen to be linear. As shown, the covariate C influences the magnitude of the variables V_T, V_O and M_O. In equations 5.2c and 5.2d, $s(O_{x,y}, T_{x,y}) = \frac{\sqrt{(T_x - O_x)^2 + (T_y - O_y)^2}}{\sqrt{(T_x + O_x)^2 + (T_y + O_y)^2}}$ is a non-linear function that scales V_O and M_O depending on the relative x- and y-coordinates of the organ, $O_{x,y}$, and the tumour, $T_{x,y}$. This scaling is used in equations 5.2c and 5.2d such that if the tumour is closer to the organ, the amount of dose I can spare to the organ is reduced and the dose fall-off is sharp. In addition, $f_{D_T}(x, y) = T_D \exp(\frac{(x - T_x)^2 + (y - T_y)^2}{V_T})$ is a Gaussian function describing the dose at and around the tumour.

The delivered dose distribution, $\mathbf{D} = \{D_{ij}\}$, will be composed using equation 5.2e. The first term in equation 5.2e, $P(i + L_x, j + L_y)$, is the shifted planned dose distribution evaluated at pixel $x = i$ and $y = j$, where L_x and L_y are the shifts in the x- and y-directions respectively. This shift represents set-up uncertainty, where L_x and L_y are sampled independently from $L_x, L_y \sim N(0, \sigma_L^2)$; where σ_L^2 is a parameter that will be varied in the simulation. The planned dose distribution function is written as,

$$P(x, y) = 2 + f_{D_T}(x, y) - f_{D_O}(x, y), \tag{5.3}$$

where $f_{D_O}(x, y) = M_O \exp(\frac{d_{ray}(x,y)^2}{V_O})$ is a Gaussian function describing the dose on and around the organ ray, where $d_{ray}(x, y)$ is the distance of point x, y from the organ ray as computed using Meijster et al. [117]'s distance transform algorithm. This creates a planned dose distribution similar to a VMAT plan, and ensures neighbouring voxels are correlated. For the conservation property of the planned dose distribution to hold (if I decrease dose in one region, I must increase it in another region), I take the total planned dose subtracted by the organ function, $\sum_i^{N_x} \sum_j^{N_y} f_{D_O}(i, j)$, and add it homogeneously across the dose dis-

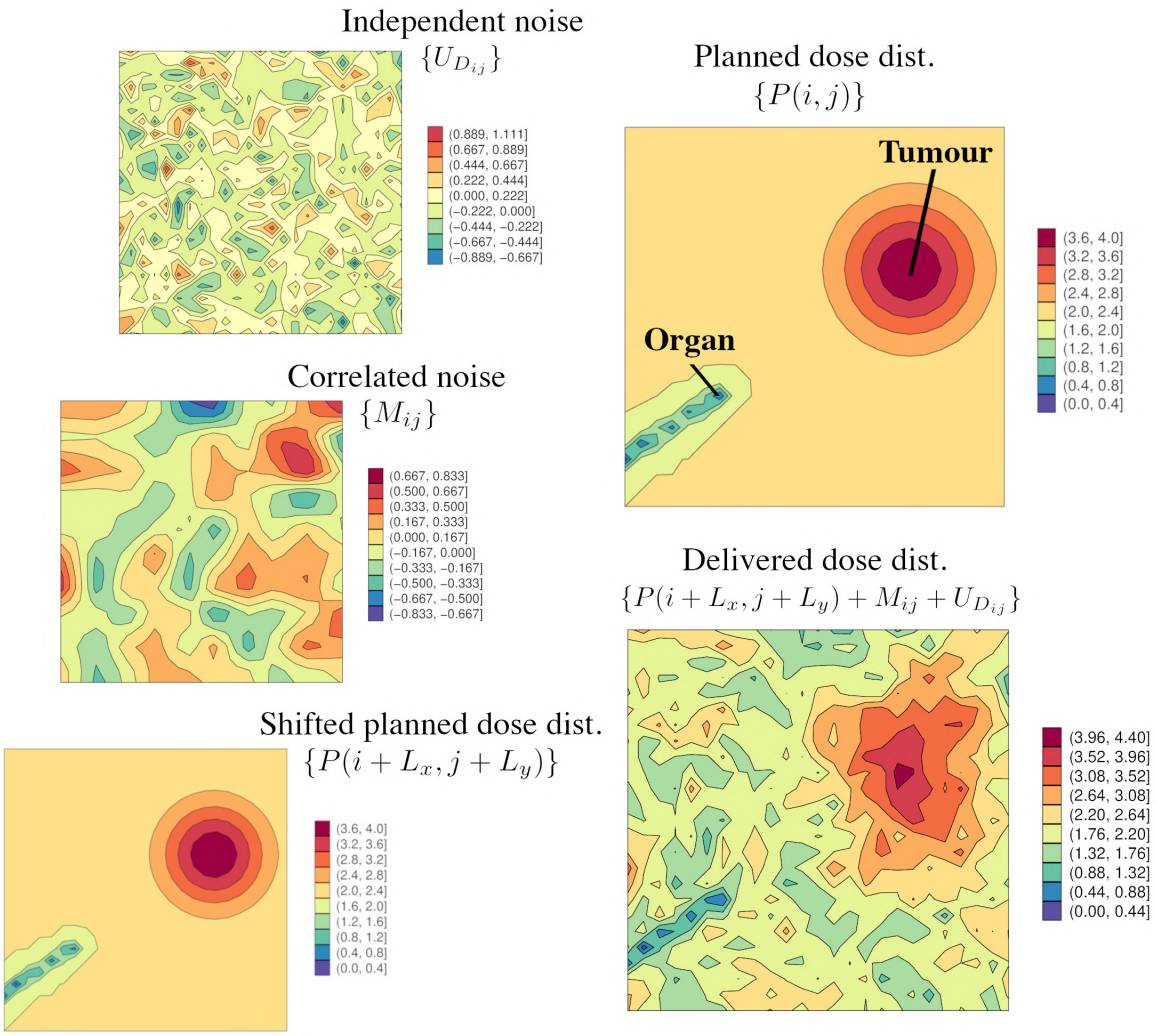


Figure 5.3. An example of a simulated planned dose distribution, a delivered dose distribution, and its components, for a patient. Simulated for $\sigma^2_{U_{D_{ij}}}, M_{\mathcal{GP}}, \sigma^2_{\mathcal{GP}}, \sigma^2_L = 3\text{e}^{-1}$. The tumour and organ locations are shown. The ray of low dose prior to the organ location is defined as the *organ ray*, as referred to in text. The organ ray is always directed towards the tumour. The equations correspond to their definitions in equation 5.2 and equation 5.3.

tribution by increasing V_T. By increasing V_T, the prescribed dose to the tumour location, $T_D = 4$, will remain unchanged. The second term in equation 5.2e, M_{ij}, represents correlated noise, $\mathbf{M} \in \mathbb{R}^{N_x \times N_y}$, evaluated at the pixel location. The correlated noise, $\mathbf{M}$, is sampled from a 2D mean-zero Gaussian process prior,

$$\mathbf{M} = M_{\mathcal{GP}} \times \mathcal{GP}(0, K(\mathbf{x}, \mathbf{x}')), \tag{5.4}$$

where $M_{\mathcal{GP}}$ is a parameter that will vary in the simulation, and $K(\mathbf{x}, \mathbf{x}')$ is a 2D radial basis function kernel parameterised by $\sigma^2_{\mathcal{GP}}$; a parameter that controls for the amount of correlation between two adjacent pixels, which will be varied in the simulation. The final term in equation 5.2e, $U_{D_{ij}}$, represents the independent Gaussian noise at each pixel (dosimetric uncertainty), which is sampled from $U_{D_{ij}} \sim N(0, \sigma^2_{D_{ij}})$ where $\sigma^2_{D_{ij}}$ is a parameter that will vary in the simulation. Figure 5.3 illustrates the planned dose distribution, the delivered dose distribution and each of its components.

The continuous outcome, Y, for each patient will be generated using equation 5.2f. The

most important fact to state regarding this equation is that the affect of D_{ij} on Y is linear and parameterised by θ_{ij}; the ATE estimand at that pixel. The affect of the interaction between D_{ij} and the covariate C, on Y, is also linear, and is parameterised by $\frac{C}{2}\xi_{ij}$; as shown in equation 5.2f. $\boldsymbol{\xi} = \{\xi_{ij}\}$ represents the interaction strength array, where $\boldsymbol{\xi} \in \mathbb{R}^{N_x \times N_y}$. The interaction strengths will be defined by the following function, $z(x, y)$, which is evaluated at the x- and y-coordinates of each pixel to define the interaction strength array, $\boldsymbol{\xi}$, in any required resolution,

$$\begin{aligned} h(x, y) &= 3 \sin\left(\frac{xy}{20N_y}\right), \\ z(x, y) &= \begin{cases} h(x, y), & \text{if } h(x, y) > 0 \text{ and } g(x, y) > 0 \\ 0, & \text{otherwise,} \end{cases} \end{aligned} \tag{5.5}$$

where $g(x, y)$ is the estimand function defined in equation 5.1. The function, $z(x, y)$, produces values in the range $[0, 3]$. Figure 5.2b shows the interaction strength array, $\boldsymbol{\xi} = \{\xi_{ij}\}$, calculated using $z(x, y)$ in equation 5.5, i.e. $\xi_{ij} = z(i, j)$, for $N_x, N_y = 10$. The interaction strength array will be constant for all patients generated in the simulation. The exact form of the interaction strength array, i.e. $z(x, y)$ equation 5.5, is chosen such that the interaction is spatially complex, inhomogeneous, positive, and is non-zero only in regions where the estimand array, $\boldsymbol{\theta}$, is non-zero.

Estimation methods

The estimands of interest are the ATE of D_{ij} on Y, as specified by equation 5.1 and exemplified by figure 5.2a. In this work, estimation methods based in causal inference will be tested and compared to current voxel-based estimation methods in radiotherapy.

Three estimation methods based in causal inference will be tested: a pixel-wise sparse causal regression, a sparse causal regression, and a causal regression. Each estimation method requires the following assumptions:

1. The causal structure is known, i.e. the DAG in figure 5.1;
2. The data generating process can be modelled using a SCM;
3. The affect of the delivered dose on outcome at each pixel is linear;
4. Ignorability, causal consistency and positivity hold.

In regards to the first assumption, the DAG in figure 5.1 assumes that the delivered dose to each pixel causes the outcome, despite the fact that the delivered dose to many pixels will have no effect on outcome, as shown in equation 5.1 and figure 5.2a. In this case, the estimation methods should correctly estimate an ATE of zero at these pixels. The second assumption holds because of the form of the generating functions in equations 5.2, i.e. they each are written as a function of their parents in figure 5.1 and independent random noise.

The third assumption is satisfied from the form of equation 5.2f. The third assumption is quite strong, so it is worth noting that it is very easily relaxed, and just made here for simplicity. To explain the fourth assumption, ignorability, i.e. no hidden confounders, is satisfied given I know the causal structure and can identify an adjustment set not including latent variable $\mathbf{N}$; causal consistency, i.e. the outcome observed is actually the potential outcome under the observed treatment, is satisfied as I generate potential outcomes from equation 5.2f; and positivity, i.e. the conditional density for the treatment is non-negative for the covariate values, is satisfied due to the form of the generating functions in equation 5.2 [118].

The correct adjustment set must be identified in order to make unbiased estimates of the ATE of delivered dose at each pixel, D_{ij}, on the outcome Y. As the data generating process is assumed to be the same for all the estimation methods, the identified adjustment sets will be identical. The adjustment set can only be composed of observed variables. Recall that $\mathbf{N}$ in figure 5.1 is a latent variable representing the parameters generating the treatment uncertainties. Therefore, the adjustment set cannot contain $\mathbf{N}$. Leveraging the Back-Door criterion, the Front-Door criterion, and instrumental variables [99, 100], I find that no adjustment sets can be defined using the Front-Door criterion or instrumental variables. However, the following adjustment sets are identified as satisfying the Back-Door criterion:

$$\mathbf{Z} = \{\{D_{ij}\}_{i,j\in N_x,N_y}, \{D_{ij} \times C\}_{i,j\in N_x,N_y}, C\}; \tag{5.6a}$$

$$\mathbf{Z} = \{\{D_{ij}\}_{i,j\in N_x,N_y}, \{D_{ij} \times C\}_{i,j\in N_x,N_y}, V_O, M_O, V_T\}. \tag{5.6b}$$

Note that in equation 5.6, because I consider $\mathbf{N}$ to be latent, the adjustment sets contain the entire delivered dose distribution and all possible interactions between delivered dose and covariate. Therefore, the adjustment sets in equations 5.6a and 5.6b contain a total of $p = 2N_xN_y + 1$ and $p = 2N_xN_y + 3$ features, respectively. In addition, to simplify the adjustment sets, they have been composed parametrically, i.e. with the knowledge that the included interaction terms should be between the delivered dose and covariate. Both adjustment sets are sufficient, but the adjustment set in equation 5.6b will be used in the simulation. This is because in reality one could argue that M_O, V_O and V_T could have a direct effect on Y (despite us assuming otherwise), and therefore equation 5.6b is more robust to mis-specification of the causal structure.

The causal regression is defined as a linear regression with the identified adjustment set as features. The sparse causal regression is defined as a linear Adaptive Lasso, using the identified adjustment set as features. The Adaptive Lasso is chosen because of its oracle property (consistency in variable selection and parameter estimation) and because $\mathbf{D}$ can be considered sparse with respect to Y [103]. The parameters of the Adaptive Lasso can be

estimated from the following loss function,

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \left(\sum_{i=1}^{n} (y_i - \sum_{j=1}^{p} X_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} w_j |\beta_j| \right), \tag{5.7}$$

where the adaptive weight vector $\mathbf{w} \in \mathbb{R}^p$ has been added to the Lasso penalty term [104]. The adaptive weights are estimated from data as,

$$\hat{\mathbf{w}} = \frac{1}{|\hat{\boldsymbol{\beta}}^*|^\gamma}, \tag{5.8}$$

where $\gamma > 0$ is an additional hyper-parameter, and $\hat{\boldsymbol{\beta}}^*$ are the parameter estimates from an additional regression that is consistent in parameter estimation. In this work, I will estimate the adaptive weights using a Ridge regression with the same identified adjustment set as features. A Ridge regression is chosen to control for the effects of multicollinearity between pixels in $\mathbf{D}$ [105]. However, the adaptive weight for variables V_O, M_O and V_T are set to zero. This is done because V_O, M_O and V_T should be in the true adjustment set. The hyper-parameters for the Ridge and Adaptive Lasso, λ_{Ridge} and λ, respectively, will be selected through 10-fold cross-validation using the GLMNET library [119] in the R programming language [120]. I choose the values of λ_{Ridge} and λ that minimise MSE. The additional hyper-parameter, γ, is set to unit value to reduce total simulation time; however, performing cross-validation to select the value of γ will likely improve the estimation methods further. The pixel-wise sparse causal regression is setup exactly the same as the sparse causal regression, except two key differences: it is applied at each pixel (a total of $N_x N_y$ regressions), and the adaptive weight for the delivered dose at present pixel, D_{ij}, is set to zero. This method is introduced under the hypothesis that removing the adaptive weight from the present pixel D_{ij}, could reduce bias in estimation.

As mentioned, the causal inference methods will be compared to the statistical voxel-based method currently used in radiotherapy: the pixel-wise univariate linear regression. Specifically, the following two models will be applied to each pixel:

$$Y = \beta_0 + \hat{\theta}_{ij}^{PP} P(i,j); \tag{5.9a}$$

$$Y = \beta_0 + \hat{\theta}_{ij}^{PD} D_{ij}. \tag{5.9b}$$

Equation 5.9a is an estimation method which uses the planned dose at the pixel, $P(i,j)$, as calculated from equation 5.3. The fitted gradient, $\hat{\theta}_{ij}^{PP}$, is interpreted as the estimate of the ATE of delivered dose at that pixel. The estimation method in equation 5.9a will be referred to as *pixel-wise planned*, hence the PP notation. Equation 5.9b is an estimation method which uses the delivered dose at the pixel, D_{ij}, as calculated from equation 5.2e. The fitted gradient, $\hat{\theta}_{ij}^{PD}$, is interpreted as the estimate of the ATE of delivered dose at that pixel. The estimation method in equation 5.9b will be referred to as *pixel-wise delivered*, hence the PD notation. In both equations 5.9a and 5.9b, β_0 is the intercept of the model.

Parameters	Values to test
n_{obs}	50, 100, 500
N_x, N_y	10, 25
$\sigma^2_{U_{D_{ij}}}$	3e−12, 3e−2, 3e−1, 3
$M_{\mathcal{GP}}$	3e−12, 3e−2, 3e−1, 3
$\sigma^2_{\mathcal{GP}}$	3e−12, 3e−2, 3e−1, 3
σ^2_L	3e−12, 3e−2, 3e−1, 3

Table 5.1. The table of parameter values to be tested in this simulation. Each possible combination of these parameters will be tested for all estimations methods. n_{obs} refers to the number of patients simulated, N_x, N_y is the number of pixels in the x- and y-directions, respectively, $\sigma^2_{U_{D_{ij}}}$ is the variance of the independent noise at each pixel (held constant across all pixels), $M_{\mathcal{GP}}$ is the magnitude of the Gaussian process, $\sigma^2_{\mathcal{GP}}$ is the variance of the radial basis function kernel used in the Gaussian process, and σ^2_L is the variance used to generate L_x and L_y; set-up uncertainty shifts in x- and y-directions.

Simulation set-up

Patient data are simulated according to the functions defined in equation 5.2 (corresponding to the DAG in figure 5.1). For 150 different combinations of the parameters defined and presented in table 5.1, the simulation will be run for a total of 100 repetitions, $n_{sim} = 100$. The 150 different combinations correspond to testing each value of n_{obs}, cases where N_x and N_y are equivalent, each value of $\sigma^2_{U_{D_{ij}}}$, and jointly varying $M_{\mathcal{GP}}$, $\sigma^2_{\mathcal{GP}}$ and σ^2_L. For the k^{th} repetition, the ATE estimate at each pixel is given by $\hat{\theta}_{ijk}$. Monte Carlo estimates of the ATE at each pixel are then calculated as $\hat{\theta}_{ij} = \frac{1}{n_{sim}} \sum_{k=1}^{n_{sim}} \hat{\theta}_{ijk}$. This is used to form the Monte Carlo estimate array, $\hat{\boldsymbol{\theta}} = \{\hat{\theta}_{ij}\}$.

As the interest is in whether the estimation method can make unbiased estimates of $\boldsymbol{\theta}$, I will use bias as a performance measure. Specifically, I will calculate the bias at each pixel and visualise the results. I define the bias array as $\mathbf{B} = \{B_{ij}\}$ where,

$$B_{ij} = \mathbb{E}[\hat{\theta}_{ij}] - \theta_{ij} \approx \frac{1}{n_{sim}} \sum_{k=1}^{n_{sim}} \hat{\theta}_{ijk} - \theta_{ij}. \tag{5.10}$$

As a global metric to compare different simulation setups, the sum of MSE across all pixels, MSE_{tot}, will be used. This is defined as

$$\text{MSE}_{tot} = \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \text{MSE}_{ij}, \tag{5.11}$$

where,

$$\text{MSE}_{ij} = \mathbb{E}[(\hat{\theta}_{ij} - \theta_{ij})^2] \approx \frac{1}{n_{sim}} \sum_{k=1}^{n_{sim}} \left(\hat{\theta}_{ijk} - \theta_{ij}\right)^2. \tag{5.12}$$

The average (mean) MSE across all pixels will also be calculated to compare setups with different N_x, N_y. This is defined as

$$\text{MSE}_{avg} = \frac{\text{MSE}_{tot}}{N_x N_y}. \tag{5.13}$$

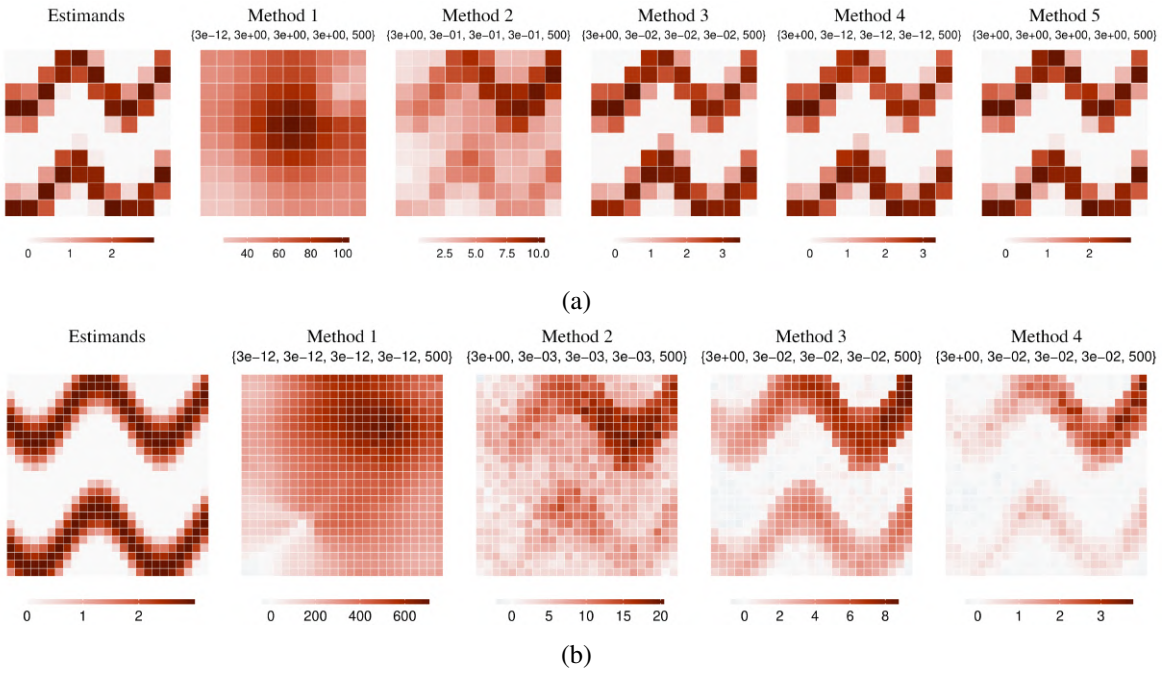


Figure 5.4. Monte Carlo estimates of the Average Treatment Effect (ATE) of the delivered dose, D_{ij}, on the outcome, Y, at each pixel for the best performing parameterisations of each estimation method. Estimands are shown too. (a) Uses constant values of $N_x, N_y = 10$, and (b) uses constant values of $N_x, N_y = 25$. Methods 1-5 correspond to the pixel-wise planned dose, pixel-wise delivered dose, pixel-wise sparse causal regression, sparse causal regression and the causal regression, respectively. Method 5, i.e. the causal regression, is removed from (b) due to failure to converge. For both (a) and (b), the best performing parameterisations are shown above each sub-figure as $\{\sigma^2_{U_{D_{ij}}}, M_{\mathcal{GP}}, \sigma^2_{\mathcal{GP}}, \sigma^2_L, n_{obs}\}$.

5.4 Results

Figures 5.4a and 5.4b show the Monte Carlo estimate arrays for the best performing, i.e. lowest MSE_{tot}, parameterisations for each estimation method at $N_x, N_y = 10$ and $N_x, N_y = 25$, respectively. From figure 5.4a at $N_x, N_y = 10$, it is observed quickly that the best performing causal inference methodologies are able to approximate the estimands more accurately compared to the best performing, currently used, pixel-wise planned and pixel-wise delivered estimation methods. In fact, the causal regression in figure 5.4a, is able to make unbiased estimates of the ATE of the delivered dose, D_{ij}, on the outcome, Y, at each pixel, at this specific parameterisation. The sparse causal regression and the pixel-wise sparse causal regression are biased, albeit accurate, with the sparse causal regression being more accurate; achieving a $\text{MSE}_{tot} = 12.0$ compared to $\text{MSE}_{tot} = 21.2$ of the pixel-wise sparse causal regression. From figure 5.4b at $N_x, N_y = 25$, the first thing to note is that the causal regression failed to converge. This is due to a larger number of collinear features, and the method failing to combat the effects of multicollinearity. Also from figure 5.4b, the best performing pixel-wise sparse causal regression and the sparse causal regression methods both perform with a lower accuracy compared to $N_x, N_y = 10$, however, the sparse causal regression is again more accurate; achieving a $\text{MSE}_{tot} = 2329.7$ compared to $\text{MSE}_{tot} = 7367.3$ of the pixel-wise sparse causal regression. The best performing sparse causal regression for $N_x, N_y = 10$ and $N_x, N_y = 25$, had MSE_{avg} of 0.12 and 3.73, respectively; showing that increasing the number of pixels has had an affect on pixel-wise

ATE inference. In figures 5.4a and 5.4b, at this specific best performing parametrisation, the pixel-wise delivered dose method in comparison to the pixel-wise planned dose method is able to estimate the underlying dose-response pattern.

Figure 5.5 shows visualisations of the bias array, $\mathbf{B}$, of each estimation method for the simulation run with different parameterisations of $\sigma^2_{U_{D_{ij}}}$, $M_{\mathcal{GP}}$, $\sigma^2_{\mathcal{GP}}$, σ^2_L and constant values of $n_{obs} = 500$, and $N_x, N_y = 10$. Firstly, it is observed that the the pixel-wise delivered dose is able to estimate the underlying ATE with lower bias as $\sigma^2_{U_{D_{ij}}}$, $M_{\mathcal{GP}}$, $\sigma^2_{\mathcal{GP}}$ and σ^2_L increase. The bias at each pixel from the causal regression reduces as $\sigma^2_{U_{D_{ij}}}$, $M_{\mathcal{GP}}$, $\sigma^2_{\mathcal{GP}}$ and σ^2_L increase. The sparse causal regression performs well over all parameterisations, unlike the causal regression and pixel-wise sparse causal regression, which show stronger dependencies on $\sigma^2_{U_{D_{ij}}}$, $M_{\mathcal{GP}}$, $\sigma^2_{\mathcal{GP}}$ and σ^2_L. In addition, when the parameters $\sigma^2_{U_{D_{ij}}}$, $M_{\mathcal{GP}}$, $\sigma^2_{\mathcal{GP}}$ and σ^2_L are all set to 3e$-$12, i.e. when the delivered dose distribution is approximately equal to the planned dose distribution, the sparse causal regression handles the high degree of multicollinearity between adjacent pixels well. This indicates that using the Ridge regression to estimate the adaptive weights was successful, as the effects of multicollinearity have been reduced by adding the Ridge penalty; unlike for the causal regression, where the bias array contains much larger values. Interestingly, the sparse causal regression has a very low bias in the regions where the estimand is zero, and has a larger bias in the regions where the estimand is non-zero. This reflects the Adaptive Lassos oracle property of consistency in variable selection, however, utilisation of the Ridge regression does bias parameter estimates where parameters are deemed important. The same observations hold for the bias arrays produced by the estimation methods for $n_{obs} = 500$, and $N_x, N_y = 25$, as shown in figure 5.6.

Figures 5.4a and 5.4b showed many of the best performing estimation methods having value of $\sigma^2_{U_{D_{ij}}} = 3$ and $n_{obs} = 500$. To visualise the dependency of MSE$_{tot}$ on $\sigma^2_{U_{D_{ij}}}$, $M_{\mathcal{GP}}$, $\sigma^2_{\mathcal{GP}}$, σ^2_L and n_{obs}, figure 5.7a was produced for $N_x, N_y = 10$. From figure 5.7a, it is first observed that all estimation methods achieve lower MSE$_{tot}$ across all parameterisations compared to the pixel-wise planned dose method. MSE$_{tot}$ for the pixel-wise delivered dose method generally decreases as $M_{\mathcal{GP}}$, $\sigma^2_{\mathcal{GP}}$ and σ^2_L increase. However, the pixel-wise delivered dose method performs best (lowest MSE$_{tot}$) for $\sigma^2_{U_{D_{ij}}} = 3$. Of the causal inference methods, the causal regression at $n_{obs} = 500$ performs the best. However, the causal regression displays the strongest dependency on n_{obs}. As there are a total of $p = 2N_xN_y + 3 = 203$ features in the adjustment set, accurate estimation of parameters is not possible until $n_{obs} > p$ using the causal regression. The pixel-wise sparse causal regression and the sparse causal regression, due to the sparsity assumption holding true, display a weaker dependence of MSE$_{tot}$ on n_{obs} compared to the causal regression. Of the causal inference methods, the sparse causal regression displays the weakest dependence of MSE$_{tot}$ on n_{obs}. For all causal inference methods, MSE$_{tot}$ decreases as $\sigma^2_{U_{D_{ij}}}$ increases. The same observations hold for figure 5.7b, which shows the dependency of MSE$_{tot}$ against $\sigma^2_{U_{D_{ij}}}$, $M_{\mathcal{GP}}$, $\sigma^2_{\mathcal{GP}}$, σ^2_L and n_{obs} for $N_x, N_y = 25$ ($p = 1253$ features). However, the best performing method across all parameterisations is the sparse causal regression.

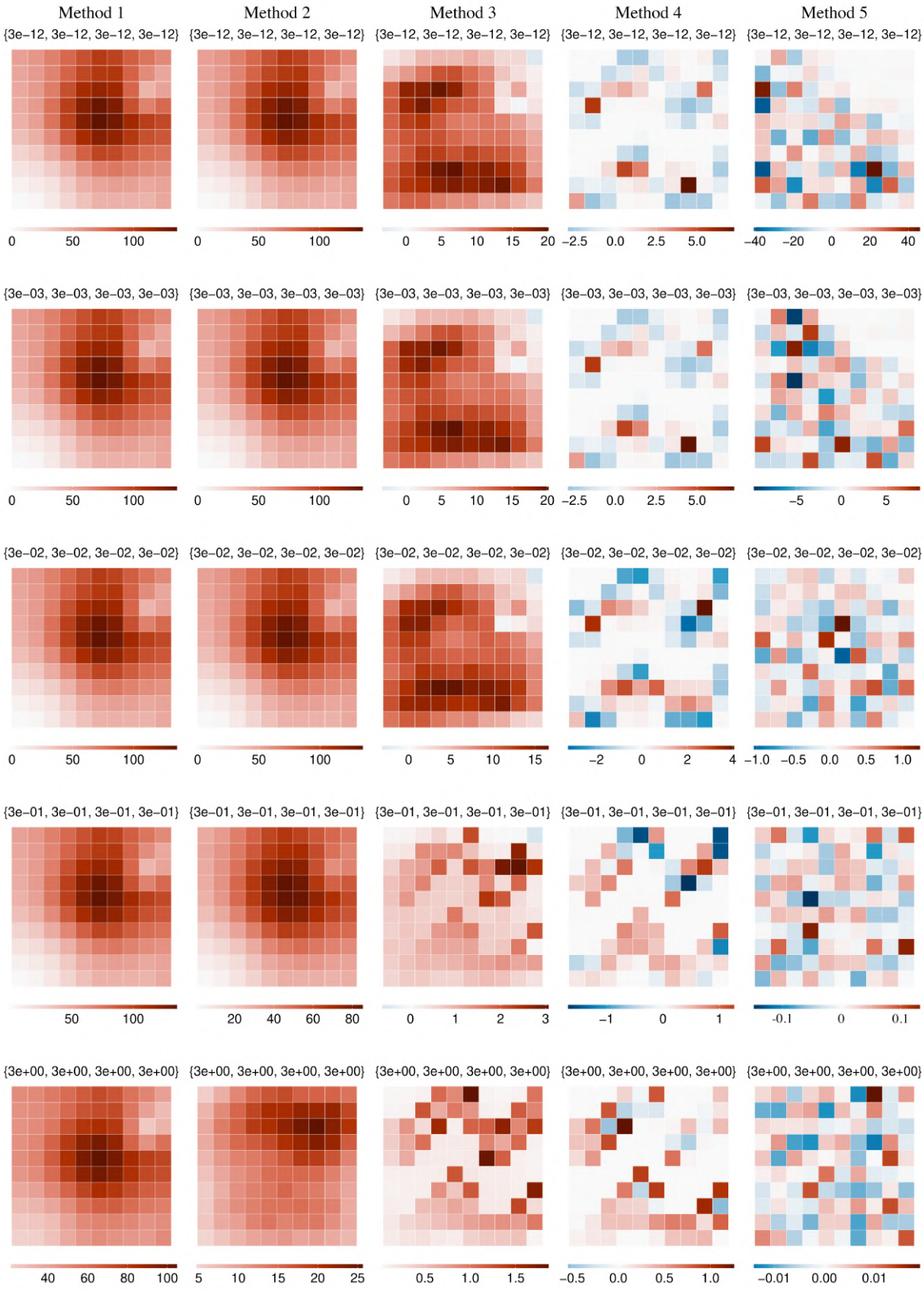


Figure 5.5. Visualisations of the Monte Carlo estimates of the bias array, $\mathbf{B}$, of each estimation method for the simulation run with different parameterisations of $\sigma^2_{U_{D_{ij}}}$, $M_{\mathcal{GP}}$, $\sigma^2_{\mathcal{GP}}$ and σ^2_L, with values shown above each sub-figure as $\{\sigma^2_{U_{D_{ij}}}, M_{\mathcal{GP}}, \sigma^2_{\mathcal{GP}}, \sigma^2_L\}$. Results are for constant values of $n_{obs} = 500$, and $N_x, N_y = 10$. The columns represent the results of estimation methods 1-5, which correspond to the pixel-wise planned dose, pixel-wise delivered dose, pixel-wise sparse causal regression, sparse causal regression and the causal regression, respectively.

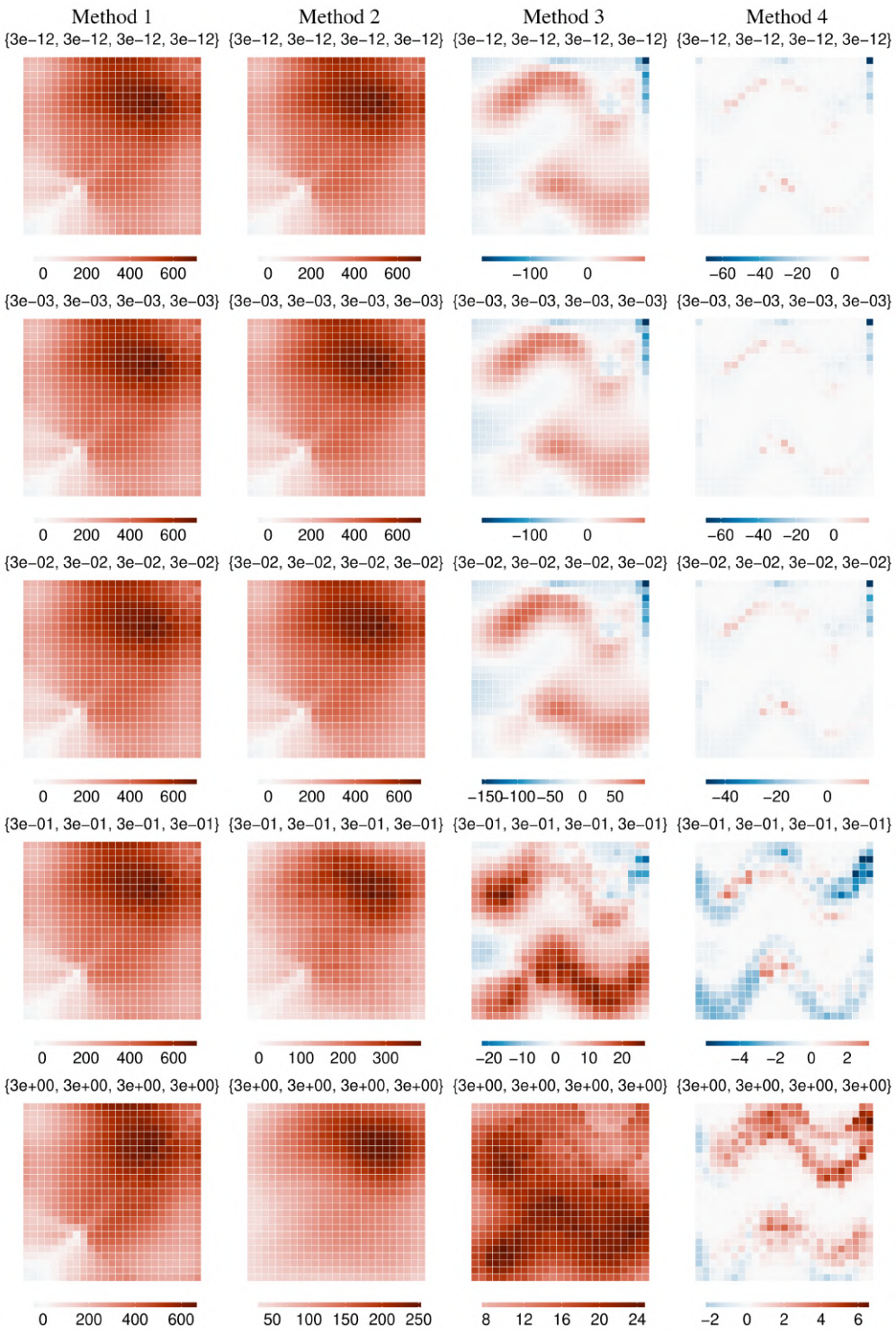
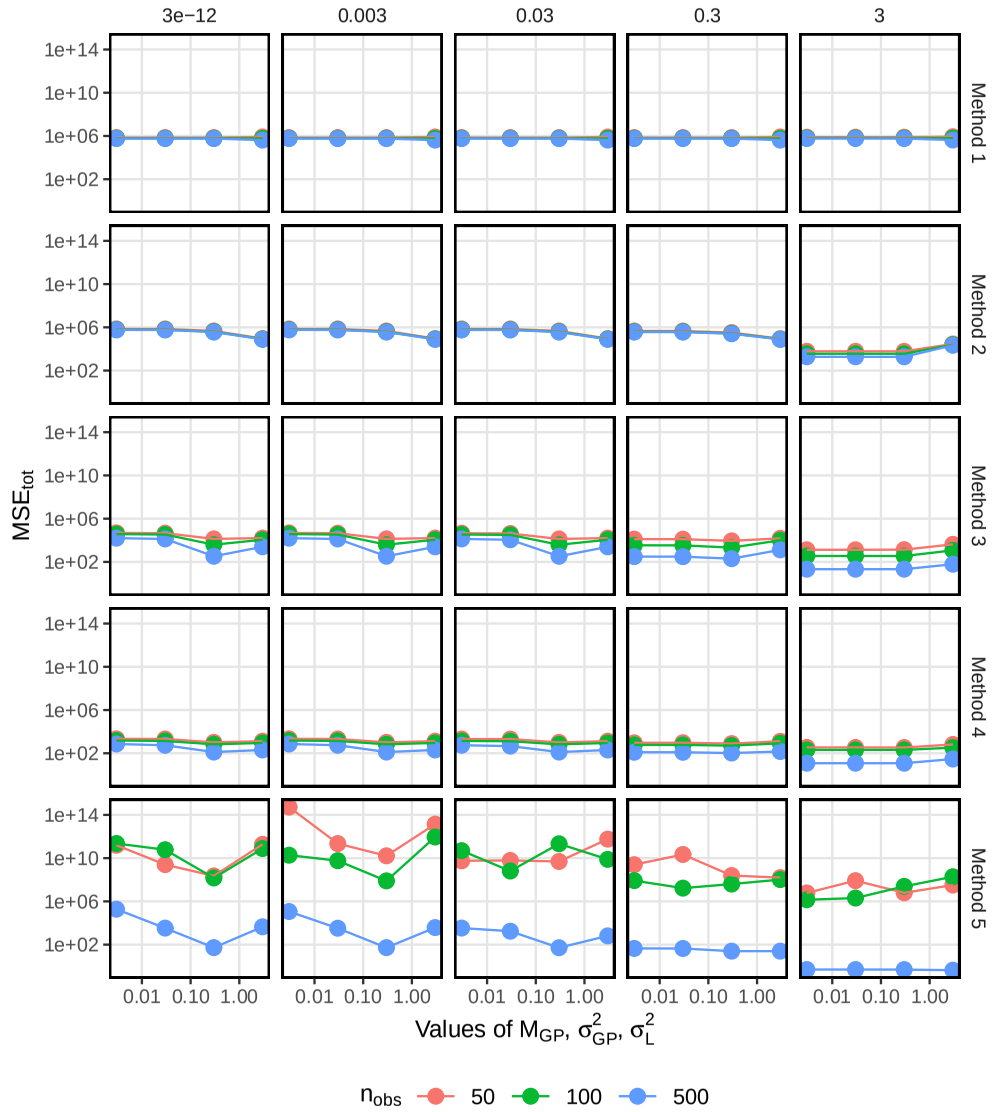


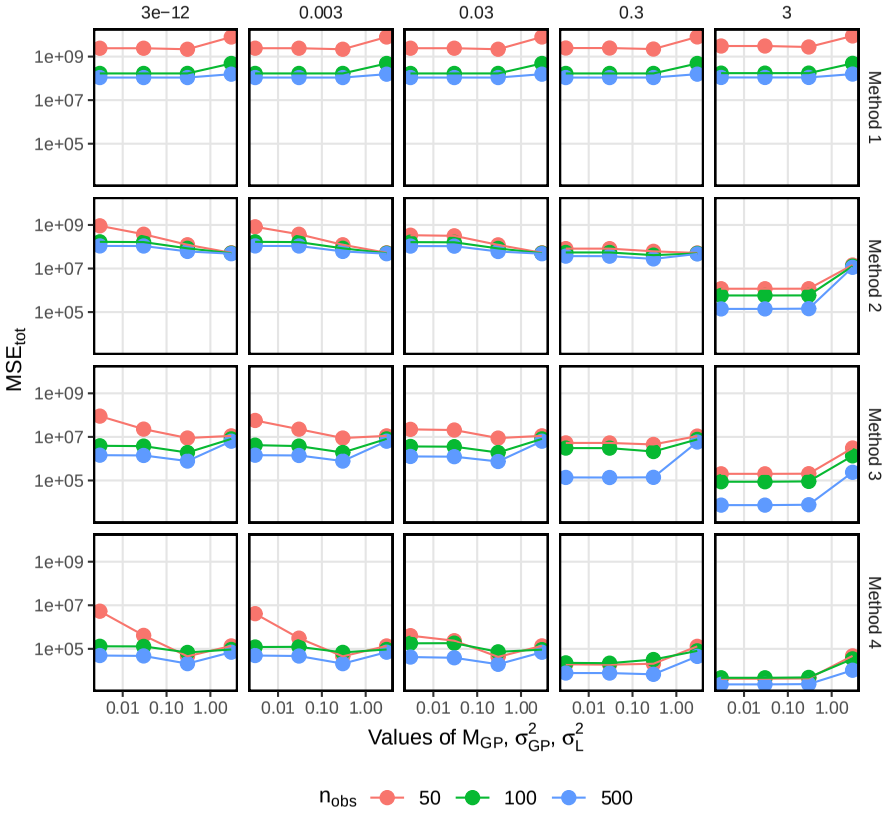
Figure 5.6. Visualisations of the Monte Carlo estimates of the bias array, $\mathbf{B}$, of each estimation method for the simulation run with different parameterisations of $\sigma^2_{U_{D_{ij}}}$, $M_{\mathcal{GP}}$, $\sigma^2_{\mathcal{GP}}$ and σ^2_L, with values shown above each sub-figure as $\{\sigma^2_{U_{D_{ij}}}, M_{\mathcal{GP}}, \sigma^2_{\mathcal{GP}}, \sigma^2_L\}$. Results are for constant values of $n_{obs} = 500$, and $N_x, N_y = 25$. The columns represent the results of estimation methods 1-4, which correspond to the pixel-wise planned dose, pixel-wise delivered dose, pixel-wise sparse causal regression and the sparse causal regression, respectively.

5.5 Discussion

In this novel work, I fused ideas in causal inference, such as SCMs, with ideas in statistics, such as sparse statistical modelling, to propose and evaluate estimation methods for pixel-wise ATE inference of a high-dimensional 2D continuous treatment on an outcome variable. I simulated radiotherapy treatment-outcome data from a simplified, yet realistic, data



(a)



(b)

Figure 5.7. Line plots of the total Mean Squared Error (MSE), MSE_{tot}, against values of $M_{\mathcal{GP}}$, $\sigma^2_{\mathcal{GP}}$ and σ^2_L for (a) $N_x, N_y = 10$ and (b) $N_x, N_y = 25$. Colours represent values of n_{obs}. The columns represent different values of $\sigma^2_{U_{D_{ij}}}$, with the exact value labelled. Methods 1-5 correspond to the pixel-wise planned dose, pixel-wise delivered dose, pixel-wise sparse causal regression, sparse causal regression and the causal regression, respectively.

generating process; ensuring many of the defining complexities of radiotherapy treatment-outcome data are preserved. Planned dose distributions were fit to an assumed randomly placed tumour and a fixed OAR, to create a treatment plan similar to VMAT. Delivered dose distributions were simulated with realistic geometric and dosimetric uncertainties: random shifts in x- and y-directions of the entire planned dose distribution, spatially correlated noise sampled from a Gaussian process prior, and independent noise at each pixel. Whilst delivered dose was assumed to be observed, parameters quantifying the geometric and dosimetric uncertainties were assumed to be latent. The outcome variable was generated by the delivered dose to a spatially complex set of pixels with inhomogeneous dose-response (ATE estimand), a covariate that also affects the delivered dose distribution, and a spatially inhomogeneous interaction between delivered dose and the covariate.

I tested 2 statistical estimation methods based on the current voxel-based analysis in radiotherapy (pixel-wise planned and pixel-wise delivered), and 3 estimation methods based in causal inference (pixel-wise sparse causal regression, sparse causal regression and a causal regression). I found that all methods based in causal inference performed with lower MSE_{tot} across all parameterisations tested in the simulation. The sparse causal regression fully exploited the oracle property of the Adaptive Lasso to simultaneously identify important pixels with dose-response and estimate ATE, and it was successful at both resolutions tested; although MSE_{avg} did increase when $N_x, N_y = 25$. As the sparsity assumption holds true, I found that the sparse estimation methods can obtain lower MSE_{tot} at a lower n_{obs} across all parameterisations of the simulation compared to the causal regression. I hypothesised in section 5.3.4 that the pixel-wise sparse causal regression could reduce bias in ATE estimates at each pixel, however, I have observed the opposite. It is thought that by setting the adaptive weight to zero for D_{ij} at the current pixel, as I loop through them, falsely identifies as D_{ij} as being in the true adjustment set; and this biases the estimate. I found that using a Ridge regression to estimate the values of the adaptive weights did indeed bias ATE estimates, however, its use in the sparse causal regression did indeed reduce the effects of multicollinearity, as the method performed well in the domain of $\sigma^2_{U_{D_{ij}}}, M_{\mathcal{GP}}, \sigma^2_{\mathcal{GP}}, \sigma^2_L = 3\text{e}^{-12}$ (where multicollinearity between adjacent pixels is strongest) compared to the other estimation methods.

Before being adapted to radiotherapy treatment-outcome data, voxel-based methods first originated in neuroimaging [63]. Furthermore, the use of sparse statistical models in the voxel-based method also first appeared in neuroimaging [73], before its application to radiotherapy data [74]. The benefits of sparse statistical models to the analysis of high-dimensional imaging / treatment data (variable selection, efficient estimation and interpretability) have been known for a while, however, it has yet to be fused with SCMs and used for causal inference for data of this type in neuroimaging or in radiotherapy. For the first time, sparsity and causal inference methodologies have been adapted to tackle the aims of the voxel-based analysis in radiotherapy treatment-outcome studies: (1) to identify sub-regions of organs with a given dose-response and (2) estimate the effect of radiotherapy voxel dose on outcome. In fact, leveraging the Adaptive Lasso and its oracle properties in the sparse causal

regression allowed us to tackle both these aims simultaneously. Given the sparsity assumption held in the simulated data, I showed that the sparse causal regression could perform well in the domain where $p >> n_{obs}$.

Although not applicable to the paradigm of the voxel-based analysis, where $p >> n_{obs}$, Nabi et al. [121] developed the technique 'Semi-Parametric Causal Sufficient Dimension Reduction' to analyse high-dimensional treatments. The authors tested this technique to estimate the causal effect of the combination of 5 continuous parotid DVH parameters on radiation-induced weight loss. One of the main ideas of Nabi et al. [121]'s work was that the high-dimensional treatment could be represented by a lower dimensional representation which preserves the cause-effect relationships. To utilise this idea to solve the aims of the voxel-based analysis in radiotherapy, one must be able to interpretably link the lower dimensional representation back to the original features. This could be challenging in general, yet especially challenging for high-dimensional causal inference methods that leverage the subject of representation learning in Artificial Intelligence (AI); where interpretability is a major challenge [122]. Whilst not interpretable, the current advances in causal representation learning include the work of Pryzant et al. [123], who tackled challenging social science questions in Natural Language Processing (NLP). The authors considered the case with text as a high-dimensional treatment and was able to adjust for confounding using adversarial learning, i.e. adjusting the loss function to learn a representation of the text that is not related to the confounders. An additional advantage of using representation learning is that a Convolutional Neural Network (CNN) could be used to construct the representation, which could potentially eliminate the role of image registration in the current voxel-based method [65]; a source of uncertainty. CNNs perform excellently on imaging data [124] and are known to be invariant under translations, with variants of the CNN also shown to be invariant under rotations [125] and scaling [126]; transformations that image registration in the voxel-based method seeks to account for.

In this work, the data generating process was setup to represent many of the defining complexities in radiotherapy treatment-outcome data: multiple correlated voxels, continuous dose values, presence of tumour and OAR, etc. However, in reality the data generating process could be much more complicated. There could be additional confounders, additional interactions, and additional clinical intervention variables that would affect the delivered dose distribution other than V_O, M_O and V_T, for example, DVH parameters for many different organs. Whilst many of these additional variables may be observed, some may be latent in practice. In addition, different mechanisms could be present. For example, I assumed the tumour location was random, however, there is data to suggest that geographical location could influence tumour location in head and neck cancer [127]. In addition, tumour location has been shown to be linked with survival outcome for different tumour sites [128, 129]. Despite the possibly more complicated data generating process, drawing the process as a DAG and identification of an adjustment set could still be possible. To check the assumed DAG is reasonable, one could check for the presence and extent of hidden bias using a sensitivity analysis.

One of the main assumptions in our choice of DAG in figure 5.1 is that the effect of V_O, M_O and V_T on Y is entirely indirect; it is mediated by the delivered dose distribution $\mathbf{D}$. V_O, M_O and V_T appear in our choice of adjustment set, however, if they do have a direct affect on Y then they must certainly be included. I also assume there is no interaction present between values of the delivered dose distribution, i.e. the effect of delivered dose on outcome at each voxel does not depend on the value of the dose elsewhere. This assumption is consistent with the early work by Schultheiss et al. [34] and Withers et al. [35] who introduced the concepts of serial and parallel organs and the Functional Subunit (FSU); a subvolume of the organ that contributes to its function via an independent dose-response function. However, evidence such as links between integral dose and fatigue [130], the possible immunogenic effect of radiotherapy on distant metastases [131, 132], and inflammatory cytokines induced by radiation that may induce a response larger than the specific radiation target [133], could all support the existence of interaction effects between values of the dose distribution. Regardless, if believed to be important the interaction terms between values of the delivered dose distribution can be included, and a sparse model consistent in variable selection can be used to identify the important variables to include.

In the simulation, I also assumed that the tumour and organ were of constant shape and size. Essentially, ignoring the effect of segmentations. As mentioned in section 5.3.1, segmentations will affect the planned dose distribution and may introduce confounding as they can mediate the effect of clinical factors. If these clinical factors are observed, they could be included in the adjustment set to make unbiased causal effect estimates. However, depending on the DAG, the segmentations may need to be included as a variable to control for confounding. In their raw form, segmentations can be represented using a high-dimensional 3D segmentation mask. In this case, advances in causal representation learning could again prove useful. For example, Veitch et al. [134] tackled challenging causal inference questions in NLP such as 'What is the causal effect of adding a theorem to a paper on the chance of conference acceptance?'. In this example, the paper's content is a high dimensional confounder, and to answer the question the confounder needs to be adjusted for. Veitch et al. [134] developed 'Causally Sufficient Embeddings', a low-dimensional representation that retains only the 'readable' parts of the confounder that are predictive of both the treatment and outcome, and is proven to be sufficient and efficient for causal inference. This methodology could be adapted to produce an embedding for the segmentation masks, which can then be adjusted for if necessary. However, if the outcome of interest is known to generate only from a specific OAR, then one could assume that only this organ segmentation affects the dose within it. In this case, the effect of the segmentations could be ignored if the dose within the organ is used instead of the full dose distribution, and normalised to a common frame of reference.

The sparsity assumption is valid in radiotherapy treatment-outcome studies for many outcomes where there is evidence to suggest a localised radiosensitive region of anatomy; rendering the remainder of the dose distribution to have no important effect for a given outcome. For example, radiation dose to the salivary glands (parotid, submandibular and sub-

lingual glands) is a known cause of xerostomia [135, 136]. On the other hand, there are some outcomes that are believed to be sparse, yet remain largely unknown. For example, whilst there is evidence to suggest that radiation dose to the posterior fossa, cerebellum and brainstem is linked with fatigue [137], a definite answer remains uncertain. A causal inference methodology that utilises sparsity, such as ours, could help shed light on outcomes such as radiation-induced fatigue. However, there are some technical challenges with the utilisation of sparse estimation methods for causal inference. For instance, if the sparse estimation method did not select the true set of non-zero parameters for the adjustment set, then the estimated causal effect could be biased. Whilst the Adaptive Lasso has the oracle property, it does so asymptotically. Therefore, whilst it is consistent in variable selection asymptotically, it could select the wrong adjustment set with finite data [138, 103]. In addition, I found that for the sparse causal regression, MSE_{avg} increased as I increased N_x and N_y. This is likely due to the total number of features growing as $p = 2N_xN_y + 3$, and number of important features growing proportionally. As N_X, N_y are increased further, and as the number of features in the true adjustment set exceeds the sample size, estimates will become noisier. Therefore, depending on the outcome studied and the sample size available, it is likely that the resolution of the delivered dose distribution will need to be carefully considered for efficient voxel-based causal inference.

In this work, I assumed that outcomes were generated via a linear function of the delivered dose distribution, the interactions, and the covariate in the SCM used. In the simulated data this assumption was true. However, in clinical data the relationships are likely to be non-linear. In this case, it is important to utilise non-parametric sparse estimation methods for the ATE. In addition, it is likely that the effect of dose at each voxel, whilst assumed to be constant across the population in this work, will be heterogeneous in clinical data. Therefore, non-parametric sparse estimation methods for heterogeneous treatment effects are likely needed. Innovations on this front include the work of Caron et al. [139], who developed the 'Sparse Bayesian Causal Forest', a non-parametric causal regression that leverages random forests to model heterogeneous non-linear relationships and includes a sparsity-inducing prior distribution for the splitting probabilities of each feature in the random forest; where non-important features are given a splitting probability close to zero. This method could be adapted for voxel-based causal inference with clinical data.

To model the system using a SCM, each variable must be a function of its parents in the DAG and some independent random noise. Whilst an appropriate model for the covariates and outcomes in clinical data, the planned dose at each voxel, which is traditionally used in voxel-based methods, is a deterministic function of its parents. To enable use of a SCM for voxel-based causal inference, I introduced the delivered dose distribution. The planned dose distribution is the expectation of the delivered dose distribution, where the expectation is over the number of treatment fractions [111]. The delivered dose distribution is known to be a better predictor of radiation-induced outcomes than the planned dose distribution [112]. The independent random noise at each pixel, $U_{D_{ij}}$, is ultimately the component of the delivered dose distribution that satisfies the SCM requirements for voxel-based causal

inference. Estimating the delivered dose distribution is possible, for example, Shelley et al. [112] estimated the delivered dose distribution using Cone Beam Computed Tomography, segmentations and a ray-tracing algorithm.

The primary goal of this work was to investigate if unbiased and consistent voxel-based causal inference is possible. Whilst simplifications were made when simulating data, many of the defining complexities of radiotherapy data were conserved. I found that unbiased and consistent voxel-based causal inference is possible under the discussed assumptions. I also found that the current estimation methods used in the voxel-based analysis, the per-pixel planned and pixel-wise delivered methods, resulted in biased estimates. Furthermore, all estimation methods that leveraged a causal framework performed with lower MSE_{tot}. I found that the sparse causal regression, which leveraged the Adaptive Lasso and its oracle properties, allowed us to tackle both aims of the voxel-based analysis simultaneously. The sparse causal regression performed well over all parameter space, and controlled for the effects of collinear features in the dose distribution. Whilst not a first introduction of causal inference to radiotherapy, this work is to the best of our knowledge a first introduction of causal inference to the voxel-based method. Therefore, the framework I proposed, from DAG setup, to the fusion of SCMs with sparse statistical models, can be built upon by future methods and scaled to the analysis of clinical radiotherapy data.

Causal inference in radiotherapy treatment-outcome data involves many of the most challenging topics in causal inference today: high-dimensional data, continuous treatments and heterogeneous treatment effects. It is an exciting and challenging direction for future research. In this work I was interested in ATE inference, however, in future research it would be extremely interesting to explore the highest level of Pearl's ladder of causation [140]: association, intervention and counterfactuals. Currently, I am at the intervention level, but being able to generate counterfactual outcomes for individuals, i.e. outcome predictions under hypothetical interventions, could be used in the future to find the optimal radiotherapy treatment for individuals.

5.6 Conclusion

The aim of this work was to use simulated data to investigate if unbiased and consistent voxel-based causal inference is possible, how, under what circumstances, and with what accuracy. I simulated simplified, yet realistic, radiotherapy treatment-outcome data. Despite our simplifications made for the simulated data, it captures many of the defining complexities of radiotherapy data: multiple correlated voxels, continuous dose values, presence of tumour and OAR, etc. I sought to establish methodologies that, for the first time, can estimate the causal effect of dose on outcome across the anatomy. I found that all methods based in causal inference performed better in the simulation compared to current voxel-based statistical methods in radiotherapy. In particular, one method I tested, the sparse causal regression, was able to accurately identify and ignore regions with zero causal effect. By ignoring the values of dose in these unimportant regions, accurate causal effect estimates

were possible in the important regions; where it was needed most. This method showed potential to work for lower sample sizes and also at higher resolutions. Whilst our work makes the first steps, further work is needed to scale up our methodologies for the analysis of clinical data.

This work has shown that leveraging sparse causal inference methods can benefit both the identification of regions of given dose-response and the estimation of treatment effects. Causal inference methodologies provide a powerful approach to tackling the aims of the voxel-based analysis and furthering our understanding of the mechanisms behind how complications arise. Therefore, adapting causal inference methodologies to the analysis of clinical radiotherapy treatment-outcome data could lead to new and impactful insights on the causes of treatment complications, and in turn, to improved optimisation of complication risk in radiotherapy treatment planning.

Chapter 6

Conclusions

The aim of current voxel-based approaches is to (1) identify sub-regions of organs with a given dose-response and (2) estimate the effect of radiotherapy voxel dose on outcome. However, this must be done from observational data, as using an RCT environment to identify the effect of radiotherapy treatment at different anatomical regions is not ethical; dose will never be randomised within a patient. I hypothesise that due to confounding present in observational data and no methodology yet to reliably adjust for it, I observe studies conflicting in location of radiosensitive sub-regions of organs and the effect of radiation dose on outcomes. Under certain assumptions, developing a causal framework provides methods to adjust for confounding. The aim of this thesis was to use simulated data to investigate if unbiased and consistent voxel-based causal inference is possible, how, under what circumstances, and with what accuracy.

Causal inference for each feature of a high-dimensional treatment is challenging for a number of reasons. Firstly, constructing a DAG for the relationships between covariates, the high-dimensional variable and the outcome, could be extremely challenging, and on occasion, it could be a completely unknown process. Secondly, analysing a high-dimensional variable with a number of features, p, that greatly exceeds the sample size, n, would lead to inefficient effect estimation. For many complications in radiotherapy treatment, it is believed that only a sub-set of the features of the high-dimensional treatment have an important effect on them, i.e. the treatment is sparse. Thereby, sparse estimation methods that leverage a causal framework, and hold the oracle property, could be a useful tool that (1) selects only the parts of the high-dimensional variable that are important and (2) enables efficient causal effect estimation.

In this novel work, I fused ideas in causal inference, such as SCMs, with ideas in statistics, such as sparse statistical modelling, to propose and evaluate estimation methods for pixel-wise ATE inference of a high-dimensional continuous treatment (radiotherapy) on an outcome variable. I simulated radiotherapy treatment-outcome data from a simplified, yet realistic, data generating process; ensuring many of the defining complexities of radiotherapy treatment-outcome data were preserved. Planned dose distributions were fit to an assumed randomly placed tumour and a fixed OAR, to create a treatment plan similar to VMAT. Delivered dose distributions were simulated with realistic geometric and dosimetric uncertainties: random shifts in x- and y-directions of the entire planned dose distribution, spatially correlated noise sampled from a Gaussian process prior, and independent noise at each

pixel. Whilst delivered dose was assumed to be observed, parameters quantifying the geometric and dosimetric uncertainties were assumed to be latent. The outcome variable was generated by the delivered dose to a spatially complex set of pixels with inhomogeneous dose-response (ATE estimand), a covariate that also affects the delivered dose distribution, and a spatially inhomogeneous interaction between delivered dose and the covariate.

I tested 2 statistical estimation methods based on the current voxel-based analysis method in radiotherapy (pixel-wise planned and pixel-wise delivered), and 3 estimation methods based in causal inference (pixel-wise sparse causal regression, sparse causal regression and a causal regression). I found that all methods based in causal inference performed better across all parameterisations tested in the simulation compared to current voxel-based statistical methods in radiotherapy. Exploiting the oracle property of the Adaptive Lasso to simultaneously identify important pixels with given dose-response and estimate ATE, was in general a successful technique over all parameterisations of the simulation tested. In fact, as the sparsity assumption generally holds true in radiotherapy treatment-outcome studies, I found that the sparse estimation methods achieve better performance at a lower n across all parameterisations compared to other estimation methods. The best estimation method over all parameterisations was the sparse causal regression. This method was able to assign a near zero effect to unimportant pixels, and whilst estimates elsewhere were biased, they were accurate; especially at lower resolutions.

In this work, the data generating process was setup to represent many of the defining complexities in radiotherapy treatment-outcome data: multiple correlated voxels, continuous dose values, presence of tumour and OAR, etc. However, in reality the data generating process could be much more complicated. Outcomes could be generated by complicated non-linear functions of dose, covariates and interactions; treatment effects are likely heterogeneous, rather than constant, in a patient population; and segmentations will affect the treatment plan. All factors we, and all current voxel-based methods, assume to be false. However, it may be possible to take account of these factors using non-parametric estimators for heterogeneous treatment effects and with current advances in representation learning for causal inference, as discussed in section 5.5. Regardless, the novel framework I proposed in this work, from DAG setup, to the fusion of SCMs with sparse statistical models, can be built upon by future methods and scaled to the analysis of clinical radiotherapy data.

This work has shown that leveraging sparse causal inference methods can benefit both the identification of regions of given dose-response and the estimation of treatment effects. Causal inference methodologies provide a powerful approach to tackling the aims of the voxel-based analysis and furthering our understanding of the mechanisms behind how complications arise. Therefore, adapting causal inference methodologies to the analysis of clinical radiotherapy treatment-outcome data could lead to new and impactful insights on the causes of treatment complications, and in turn, to improved optimisation of complication risk in radiotherapy treatment planning.

References

[1] Rifat Atun et al. “Expanding global access to radiotherapy”. In: *The Lancet Oncology* 16.10 (Sept. 2015), pp. 1153–1186. DOI: 10.1016/s1470-2045(15)00222-3. URL: https://doi.org/10.1016/s1470-2045(15)00222-3.

[2] Josep M. Borras et al. “The optimal utilization proportion of external beam radiotherapy in European countries: An ESTRO-HERO analysis”. In: *Radiotherapy and Oncology* 116.1 (July 2015), pp. 38–44. DOI: 10.1016/j.radonc.2015.04.018. URL: https://doi.org/10.1016/j.radonc.2015.04.018.

[3] *Global Cancer Observatory: Cancer Survival*. 2021. URL: https://gco.iarc.fr/survival/survmark/ (visited on 01/05/2021).

[4] Lawrence B. Marks et al. “Use of Normal Tissue Complication Probability Models in the Clinic”. In: *International Journal of Radiation Oncology∗Biology∗Physics* 76.3 (Mar. 2010), S10–S19. DOI: 10.1016/j.ijrobp.2009.07.1754. URL: https://doi.org/10.1016/j.ijrobp.2009.07.1754.

[5] I. Improta et al. “Bladder spatial-dose descriptors correlate with acute urinary toxicity after radiation therapy for prostate cancer”. In: *Physica Medica* 32.12 (Dec. 2016), pp. 1681–1689. DOI: 10.1016/j.ejmp.2016.08.013. URL: https://doi.org/10.1016/j.ejmp.2016.08.013.

[6] Peter van Luijk et al. “Sparing the region of the salivary gland containing stem cells preserves saliva production after radiotherapy for head and neck cancer”. In: *Science Translational Medicine* 7.305 (Sept. 2015), 305ra147–305ra147. DOI: 10.1126/scitranslmed.aac4441. URL: https://doi.org/10.1126/scitranslmed.aac4441.

[7] Jordan Houri et al. “Analyses of regional radiosensitivity of white matter structures along tract axes using novel white matter segmentation and diffusion imaging biomarkers”. In: *Physics and Imaging in Radiation Oncology* 6 (Apr. 2018), pp. 39–46. DOI: 10.1016/j.phro.2018.04.003. URL: https://doi.org/10.1016/j.phro.2018.04.003.

[8] Yvette Seppenwoolde, Katrien De Jaeger, Liesbeth J. Boersma, José S.A. Belderbos, and Joos V. Lebesque. “Regional differences in lung radiosensitivity after radiotherapy for non–small-cell lung cancer”. In: *International Journal of Radiation Oncol-*

ogy∗Biology∗Physics 60.3 (Nov. 2004), pp. 748–758. DOI: 10.1016/j.ijrobp.2004.04.037. URL: https://doi.org/10.1016/j.ijrobp.2004.04.037.

[9] Alan McWilliam, Jason Kennedy, Clare Hodgson, Eliana Vasquez Osorio, Corinne Faivre-Finn, and Marcel van Herk. "Radiation dose to heart base linked with poorer survival in lung cancer patients". In: *European Journal of Cancer* 85 (Nov. 2017), pp. 106–113. DOI: 10.1016/j.ejca.2017.07.053. URL: https://doi.org/10.1016/j.ejca.2017.07.053.

[10] Klaus-Rüdiger Trott, Wolfgang Doerr, Angelica Facoetti, John Hopewell, Johannes Langendijk, Peter van Luijk, Andrea Ottolenghi, and Vere Smyth. "Biological mechanisms of normal tissue damage: Importance for the design of NTCP models". In: *Radiotherapy and Oncology* 105.1 (Oct. 2012), pp. 79–85. DOI: 10.1016/j.radonc.2012.05.008. URL: https://doi.org/10.1016/j.radonc.2012.05.008.

[11] J W Hopewell. "The Volume Effect in Radiotherapy—Its Biological Significance". In: *The British Journal of Radiology* 70.Special-Issue-1 (Nov. 1997), S32–S40. DOI: 10.1259/bjr.1997.0006. URL: https://doi.org/10.1259/bjr.1997.0006.

[12] Andrew Green, Eliana Vasquez Osorio, Marianne C. Aznar, Alan McWilliam, and Marcel van Herk. "Image Based Data Mining Using Per-voxel Cox Regression". In: *Frontiers in Oncology* 10 (July 2020). DOI: 10.3389/fonc.2020.01178. URL: https://doi.org/10.3389%2Ffonc.2020.01178.

[13] Alexander L. Jenkins, Thomas Soares Mullen, Corinne Johnson-Hart, Andrew Green, Alan McWilliam, Marianne Aznar, Marcel van Herk, and Eliana Vasquez Osorio. "Novel methodology to assess the effect of contouring variation on treatment outcome". In: *Medical Physics* 48.6 (Apr. 2021), pp. 3234–3242. DOI: 10.1002/mp.14865. URL: https://doi.org/10.1002/mp.14865.

[14] Veerle A.B. van den Bogaard et al. "Validation and Modification of a Prediction Model for Acute Cardiac Events in Patients With Breast Cancer Treated With Radiotherapy Based on Three-Dimensional Dose Distributions to Cardiac Substructures". In: *Journal of Clinical Oncology* 35.11 (Apr. 2017), pp. 1171–1178. DOI: 10.1200/jco.2016.69.8480. URL: https://doi.org/10.1200/jco.2016.69.8480.

[15] J. Price, E. Hall, C. West, and D. Thomson. "TORPEdO – A Phase III Trial of Intensity-modulated Proton Beam Therapy Versus Intensity-modulated Radiotherapy for Multitoxicity Reduction in Oropharyngeal Cancer". In: *Clinical Oncology* 32.2 (Feb.

2020), pp. 84–88. DOI: 10.1016/j.clon.2019.09.052. URL: https://doi.org/10.1016/j.clon.2019.09.052.

[16] C. R. Hansen et al. “NTCP model validation method for DAHANCA patient selection of protons versus photons in head and neck cancer radiotherapy”. In: *Acta Oncologica* 58.10 (Aug. 2019), pp. 1410–1415. DOI: 10.1080/0284186x.2019.1654129. URL: https://doi.org/10.1080/0284186x.2019.1654129.

[17] Claudia E. Rübe et al. “History of Radiation Oncology”. In: *Encyclopedia of Radiation Oncology*. Springer Berlin Heidelberg, 2013, pp. 314–325. DOI: 10.1007/978-3-540-85516-3_441. URL: https://doi.org/10.1007/978-3-540-85516-3_441.

[18] Xiufang Tian, Kun Liu, Yong Hou, Jian Cheng, and Jiandong Zhang. “The evolution of proton beam therapy: Current and future status (Review)”. In: *Molecular and Clinical Oncology* (Nov. 2017). DOI: 10.3892/mco.2017.1499. URL: https://doi.org/10.3892/mco.2017.1499.

[19] Bernard Gottschalk. “Physics of Proton Interactions in Matter”. In: *Proton therapy physics*. Ed. by Harald Paganetti. Boca Raton, FL: CRC Press/Taylor & Francis, 2012. Chap. 2, pp. 19–61. ISBN: 978-1-4398-3645-3.

[20] Robert R. Wilson. “Radiological Use of Fast Protons”. In: *Radiology* 47.5 (Nov. 1946), pp. 487–491. DOI: 10.1148/47.5.487. URL: https://doi.org/10.1148/47.5.487.

[21] J. H. LAWRENCE, C. A. TOBIAS, J. L. BORN, R. K. McCOMBS, J. E. ROBERTS, H. O. ANGER, B. V. LOW-BEER, and C. B. HUGGINS. “Pituitary irradiation with high-energy proton beams: a preliminary report”. In: *Cancer Res* 18.2 (Feb. 1958), pp. 121–134.

[22] Christian Karger. “Biological Models in Treatment Planning”. In: *New Technologies in Radiation Oncology*. Springer-Verlag Berlin Heidelberg, 2006. Chap. 18, pp. 221–233.

[23] Filip T. Troicki et al. “Normal Tissue Complication Probability (NTCP)”. In: *Encyclopedia of Radiation Oncology*. Springer Berlin Heidelberg, 2013, pp. 560–560. DOI: 10.1007/978-3-540-85516-3_341. URL: https://doi.org/10.1007/978-3-540-85516-3_341.

[24] Frank Ellis. “Tolerance Dosage in Radiotherapy with 200 kV X rays”. In: *The British Journal of Radiology* 15.180 (Dec. 1942), pp. 348–350. DOI: 10.1259/0007-1285-15-180-348. URL: https://doi.org/10.1259/0007-1285-15-180-348.

[25] John William Hopewell and Klaus-Rüdiger Trott. “Volume effects in radiobiology as applied to radiotherapy”. In: *Radiotherapy and Oncology* 56.3 (Sept. 2000), pp. 283–288. DOI: 10.1016/s0167-8140(00)00236-x. URL: https://doi.org/10.1016/s0167-8140(00)00236-x.

[26] Eberhard Scherer. *Radiopathology of Organs and Tissues*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1991, pp. 347–368. ISBN: 978-3-642-83416-5.

[27] Tod W. Speer et al. “Dose Volume Histogram (DVH)”. In: *Encyclopedia of Radiation Oncology*. Springer Berlin Heidelberg, 2013, pp. 166–166. DOI: 10.1007/978-3-540-85516-3_659. URL: https://doi.org/10.1007/978-3-540-85516-3_659.

[28] John T. Lyman. “Complication Probability as Assessed from Dose-Volume Histograms”. In: *Radiation Research Supplement* 8 (Nov. 1985), S13. DOI: 10.2307/3583506. URL: https://doi.org/10.2307/3583506.

[29] P. Rubin and G. Casarett. “A Direction for Clinical Radiation Pathology”. In: *Frontiers of Radiation Therapy and Oncology*. S. Karger AG, 1972, pp. 1–16. DOI: 10.1159/000392794. URL: https://doi.org/10.1159/000392794.

[30] Lionel Cohen. “The tissue volume factor in radiation oncology”. In: *International Journal of Radiation Oncology∗Biology∗Physics* 8.10 (Oct. 1982), pp. 1771–1774. DOI: 10.1016/0360-3016(82)90300-5. URL: https://doi.org/10.1016/0360-3016(82)90300-5.

[31] B. Emami et al. “Tolerance of normal tissue to therapeutic irradiation”. In: *International Journal of Radiation Oncology∗Biology∗Physics* 21.1 (May 1991), pp. 109–122. DOI: 10.1016/0360-3016(91)90171-y. URL: https://doi.org/10.1016/0360-3016(91)90171-y.

[32] C. Burman, G.J. Kutcher, B. Emami, and M. Goitein. “Fitting of normal tissue tolerance data to an analytic function”. In: *International Journal of Radiation Oncology∗Biology∗Physics* 21.1 (May 1991), pp. 123–135. DOI: 10.1016/0360-3016(91)90172-z. URL: https://doi.org/10.1016/0360-3016(91)90172-z.

[33] G.J. Kutcher, C. Burman, L. Brewster, M. Goitein, and R. Mohan. “Histogram reduction method for calculating complication probabilities for three-dimensional treatment planning evaluations”. In: *International Journal of Radiation Oncology∗Biology∗Physics* 21.1 (May 1991), pp. 137–146. DOI: 10.1016/0360-3016(91)90173-2. URL: https://doi.org/10.1016/0360-3016(91)90173-2.

[34] Timothy E. Schultheiss, Colin G. Orton, and R. A. Peck. “Models in radiotherapy: Volume effects”. In: *Medical Physics* 10.4 (July 1983), pp. 410–415. DOI: 10.1118/1.595312. URL: https://doi.org/10.1118/1.595312.

[35] H. Rodney Withers, Jeremy M.G. Taylor, and Boguslaw Maciejewski. “Treatment volume and tissue tolerance”. In: *International Journal of Radiation Oncology∗Biology∗Physics* 14.4 (Apr. 1988), pp. 751–759. DOI: 10.1016/0360-3016(88)90098-3. URL: https://doi.org/10.1016/0360-3016(88)90098-3.

[36] Andrzej Niemierko and Michael Goitein. “Calculation of normal tissue complication probability and dose-volume histogram reduction schemes for tissues with a critical element architecture”. In: *Radiotherapy and Oncology* 20.3 (Mar. 1991), pp. 166–176. DOI: 10.1016/0167-8140(91)90093-v. URL: https://doi.org/10.1016/0167-8140(91)90093-v.

[37] A. Jackson, G. J. Kutcher, and E. D. Yorke. “Probability of radiation-induced complications for normal tissues with parallel architecture subject to non-uniform irradiation”. In: *Medical Physics* 20.3 (May 1993), pp. 613–625. DOI: 10.1118/1.597056. URL: https://doi.org/10.1118/1.597056.

[38] Andrzej Niemierko and Michael Goitein. “Modeling of normal tissue response to radiation: The critical volume model”. In: *International Journal of Radiation Oncology∗Biology∗Physics* 25.1 (Jan. 1993), pp. 135–145. DOI: 10.1016/0360-3016(93)90156-p. URL: https://doi.org/10.1016/0360-3016(93)90156-p.

[39] Jamie A. Dean et al. “Normal tissue complication probability (NTCP) modelling using spatial dose metrics and machine learning methods for severe acute oral mucositis resulting from head and neck radiotherapy”. In: *Radiotherapy and Oncology* 120.1 (July 2016), pp. 21–27. DOI: 10.1016/j.radonc.2016.05.015. URL: https://doi.org/10.1016/j.radonc.2016.05.015.

[40] Shalini K. Vinod, Michael G. Jameson, Myo Min, and Lois C. Holloway. “Uncertainties in volume delineation in radiation oncology: A systematic review and recommendations for future studies”. In: *Radiotherapy and Oncology* 121.2 (Nov. 2016), pp. 169–179. DOI: 10.1016/j.radonc.2016.09.009. URL: https://doi.org/10.1016/j.radonc.2016.09.009.

[41] Marjan Sharabiani, Enrico Clementel, Nicolaus Andratschke, and Coen Hurkmans. “Generalizability assessment of head and neck cancer NTCP models based on the TRIPOD criteria”. In: *Radiotherapy and Oncology* 146 (May 2020), pp. 143–150. DOI: 10.1016/j.radonc.2020.02.013. URL: https://doi.org/10.1016/j.radonc.2020.02.013.

[42] Avraham Eisbruch, Randall K. Ten Haken, Hyungjin M. Kim, Lon H. Marsh, and Jonathan A. Ship. “Dose, volume, and function relationships in parotid salivary glands following conformal and intensity-modulated irradiation of head and neck cancer”. In: *International Journal of Radiation Oncology∗Biology∗Physics* 45.3 (Oct. 1999), pp. 577–587. DOI: 10.1016/s0360-3016(99)00247-3. URL: https://doi.org/10.1016%2Fs0360-3016%2899%2900247-3.

[43] Issam El Naqa, Jeffrey Bradley, Angel I. Blanco, Patricia E. Lindsay, Milos Vicic, Andrew Hope, and Joseph O. Deasy. “Multivariable modeling of radiotherapy outcomes, including dose–volume and clinical factors”. In: *International Journal of Radiation Oncology∗Biology∗Physics* 64.4 (Mar. 2006), pp. 1275–1286. DOI: 10.1016/j.ijrobp.2005.11.022. URL: https://doi.org/10.1016/j.ijrobp.2005.11.022.

[44] Ivo Beetz et al. “Development of NTCP models for head and neck cancer patients treated with three-dimensional conformal radiotherapy for xerostomia and sticky saliva: The role of dosimetric and clinical factors”. In: *Radiotherapy and Oncology* 105.1 (Oct. 2012), pp. 86–93. DOI: 10.1016/j.radonc.2011.05.010. URL: https://doi.org/10.1016/j.radonc.2011.05.010.

[45] Tsair-Fwu Lee et al. “Using Multivariate Regression Model with Least Absolute Shrinkage and Selection Operator (LASSO) to Predict the Incidence of Xerostomia after Intensity-Modulated Radiotherapy for Head and Neck Cancer”. In: *PLoS ONE* 9.2 (Feb. 2014). Ed. by Nils Cordes, e89700. DOI: 10.1371/journal.pone.0089700. URL: https://doi.org/10.1371%2Fjournal.pone.0089700.

[46] Tsair-Fwu Lee, Shyh-An Yeh, Pei-Ju Chao, Liyun Chang, Chien-Liang Chiu, Hui-Min Ting, Hung-Yu Wang, and Yu-Jie Huang. “Normal tissue complication probability modeling for cochlea constraints to avoid causing tinnitus after head-and-neck intensity-modulated radiation therapy”. In: *Radiation Oncology* 10.1 (Sept. 2015). DOI: 10.1186/s13014-015-0501-x. URL: https://doi.org/10.1186/s13014-015-0501-x.

[47] Susan Cheraghi, Alireza Nikoofar, Mohsen Bakhshandeh, Samideh Khoei, Saeid Farahani, Hamid Abdollahi, and Seied Rabi Mahdavi. “Normal tissue complication probability modeling of radiation-induced sensorineural hearing loss after head-and-neck radiation therapy”. In: *International Journal of Radiation Biology* 93.12 (Oct. 2017), pp. 1327–1333. DOI: 10.1080/09553002.2017.1385872. URL: https://doi.org/10.1080/09553002.2017.1385872.

[48] Miranda E.M.C. Christianen et al. “Predictive modelling for swallowing dysfunction after primary (chemo)radiation: Results of a prospective observational study”. In: *Radiotherapy and Oncology* 105.1 (Oct. 2012), pp. 107–114. DOI: 10.1016/j.radonc.2011.08.009. URL: https://doi.org/10.1016/j.radonc.2011.08.009.

[49] Miranda E.M.C. Christianen et al. “Swallowing sparing intensity modulated radiotherapy (SW-IMRT) in head and neck cancer: Clinical validation according to the model-based approach”. In: *Radiotherapy and Oncology* 118.2 (Feb. 2016), pp. 298–303. DOI: 10.1016/j.radonc.2015.11.009. URL: https://doi.org/10.1016/j.radonc.2015.11.009.

[50] Kim Wopken et al. “Development of a multivariable normal tissue complication probability (NTCP) model for tube feeding dependence after curative radiotherapy/chemo-radiotherapy in head and neck cancer”. In: *Radiotherapy and Oncology* 113.1 (Oct. 2014), pp. 95–101. DOI: 10.1016/j.radonc.2014.09.013. URL: https://doi.org/10.1016/j.radonc.2014.09.013.

[51] Naoyuki Kanayama et al. “External validation of a multifactorial normal tissue complication probability model for tube feeding dependence at 6 months after definitive radiotherapy for head and neck cancer”. In: *Radiotherapy and Oncology* 129.2 (Nov. 2018), pp. 403–408. DOI: 10.1016/j.radonc.2018.09.013. URL: https://doi.org/10.1016/j.radonc.2018.09.013.

[52] Kim Wopken et al. “Development and Validation of a Prediction Model for Tube Feeding Dependence after Curative (Chemo-) Radiation in Head and Neck Cancer”. In: *PLoS ONE* 9.4 (Apr. 2014). Ed. by Richard E. Burney, e94879. DOI: 10.1371/journal.pone.0094879. URL: https://doi.org/10.1371/journal.pone.0094879.

[53] Sophie Otter et al. “Evaluation of the Risk of Grade 3 Oral and Pharyngeal Dysphagia Using Atlas-Based Method and Multivariate Analyses of Individual Patient Dose Distributions”. In: *International Journal of Radiation Oncology∗Biology∗Physics* 93.3 (Nov. 2015), pp. 507–515. DOI: 10.1016/j.ijrobp.2015.07.2263. URL: https://doi.org/10.1016/j.ijrobp.2015.07.2263.

[54] Kristin J. Redmond et al. “Association between radiation dose to neuronal progenitor cell niches and temporal lobes and performance on neuropsychological testing in children: a prospective study”. In: *Neuro-Oncology* 15.3 (Jan. 2013), pp. 360–369. DOI: 10.1093/neuonc/nos303. URL: https://doi.org/10.1093/neuonc/nos303.

[55] Andrew H. Zureick et al. “Left hippocampal dosimetry correlates with visual and verbal memory outcomes in survivors of pediatric brain tumors”. In: *Cancer* 124.10 (Mar. 2018), pp. 2238–2245. DOI: 10.1002/cncr.31143. URL: https://doi.org/10.1002/cncr.31143.

[56] Laura Cella, Raffaele Liuzzi, Manuel Conson, Vittoria D’Avino, Marco Salvatore, and Roberto Pacelli. “Development of multivariate NTCP models for radiation-induced hypothyroidism: a comparative analysis”. In: *Radiation Oncology* 7.1 (Dec. 2012). DOI: 10.1186/1748-717x-7-224. URL: https://doi.org/10.1186/1748-717x-7-224.

[57] Marianne Feen Rønjom, Carsten Brink, Søren M. Bentzen, Laszlo Hegedüs, Jens Overgaard, and Jørgen Johansen. “Hypothyroidism after primary radiotherapy for head and neck squamous cell carcinoma: Normal tissue complication probability modeling with latent time correction”. In: *Radiotherapy and Oncology* 109.2 (Nov. 2013), pp. 317–322. DOI: 10.1016/j.radonc.2013.06.029. URL: https://doi.org/10.1016/j.radonc.2013.06.029.

[58] Marianne F. Rønjom, Carsten Brink, Søren M. Bentzen, Laszlo Hegedüs, Jens Overgaard, Jørgen B. B. Petersen, Hanne Primdahl, and Jørgen Johansen. “External validation of a normal tissue complication probability model for radiation-induced hypothyroidism in an independent cohort”. In: *Acta Oncologica* 54.9 (Aug. 2015), pp. 1301–1309. DOI: 10.3109/0284186x.2015.1064160. URL: https://doi.org/10.3109/0284186x.2015.1064160.

[59] Ellen X. Huang et al. “Modeling the Risk of Radiation-Induced Acute Esophagitis for Combined Washington University and RTOG Trial 93-11 Lung Cancer Patients”. In: *International Journal of Radiation Oncology∗Biology∗Physics* 82.5 (Apr. 2012), pp. 1674–1679. DOI: 10.1016/j.ijrobp.2011.02.052. URL: https://doi.org/10.1016/j.ijrobp.2011.02.052.

[60] Robin Wijsman, Frank Dankers, Esther G.C. Troost, Aswin L. Hoffmann, Erik H.F.M. van der Heijden, Lioe-Fee de Geus-Oei, and Johan Bussink. “Multivariable normal-tissue complication modeling of acute esophageal toxicity in advanced stage non-small cell lung cancer patients treated with intensity-modulated (chemo-)radiotherapy”. In: *Radiotherapy and Oncology* 117.1 (Oct. 2015), pp. 49–54. DOI: 10.1016/j.radonc.2015.08.010. URL: https://doi.org/10.1016/j.radonc.2015.08.010.

[61] Frank J.W.M. Dankers et al. “External validation of an NTCP model for acute esophageal toxicity in locally advanced NSCLC patients treated with intensity-modulated (chemo-

)radiotherapy". In: *Radiotherapy and Oncology* 129.2 (Nov. 2018), pp. 249–256. DOI: 10.1016/j.radonc.2018.07.021. URL: https://doi.org/10.1016/j.radonc.2018.07.021.

[62] John Ashburner and Karl J. Friston. "Voxel-Based Morphometry—The Methods". In: *NeuroImage* 11.6 (June 2000), pp. 805–821. DOI: 10.1006/nimg.2000.0582. URL: https://doi.org/10.1006/nimg.2000.0582.

[63] J. L. Whitwell. "Voxel-Based Morphometry: An Automated Technique for Assessing Structural Changes in the Brain". In: *Journal of Neuroscience* 29.31 (Aug. 2009), pp. 9661–9664. DOI: 10.1523/jneurosci.2160-09.2009. URL: https://doi.org/10.1523/jneurosci.2160-09.2009.

[64] Marnix G. Witte, Wilma D. Heemsbergen, Román Bohoslavsky, Floris J. Pos, Abrahim Al-Mamgani, Joos V. Lebesque, and Marcel van Herk. "Relating Dose Outside the Prostate With Freedom From Failure in the Dutch Trial 68 Gy vs. 78 Gy". In: *International Journal of Radiation Oncology∗Biology∗Physics* 77.1 (May 2010), pp. 131–138. DOI: 10.1016/j.ijrobp.2009.04.040. URL: https://doi.org/10.1016/j.ijrobp.2009.04.040.

[65] G. Palma, S. Monti, and L. Cella. "Voxel-based analysis in radiation oncology: A methodological cookbook". In: *Physica Medica* 69 (Jan. 2020), pp. 192–204. DOI: 10.1016/j.ejmp.2019.12.013. URL: https://doi.org/10.1016/j.ejmp.2019.12.013.

[66] Jermome Friedman Trevor Hastie Robert Tibshirani. *The elements of statistical learning : data mining, inference, and prediction*. New York: Springer, 2009, pp. 683–693. ISBN: 978-0-387-84857-0.

[67] Giuseppe Palma, Serena Monti, Amedeo Buonanno, Roberto Pacelli, and Laura Cella. "PACE: A Probabilistic Atlas for Normal Tissue Complication Estimation in Radiation Oncology". In: *Frontiers in Oncology* 9 (Mar. 2019). DOI: 10.3389/fonc.2019.00130. URL: https://doi.org/10.3389/fonc.2019.00130.

[68] Alan McWilliam, Jonathan Khalifa, Eliana Vasquez Osorio, Kathryn Banfill, Azadeh Abravan, Corinne Faivre-Finn, and Marcel van Herk. "Novel Methodology to Investigate the Effect of Radiation Dose to Heart Substructures on Overall Survival". In: *International Journal of Radiation Oncology∗Biology∗Physics* 108.4 (Nov. 2020), pp. 1073–1081. DOI: 10.1016/j.ijrobp.2020.06.031. URL: https://doi.org/10.1016/j.ijrobp.2020.06.031.

[69] D. Westreich and S. Greenland. “The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients”. In: *American Journal of Epidemiology* 177.4 (Jan. 2013), pp. 292–298. DOI: 10.1093/aje/kws412. URL: https://doi.org/10.1093/aje/kws412.

[70] Judea Pearl. “Causal inference in statistics: An overview”. In: *Statistics Surveys* 3.none (Jan. 2009). DOI: 10.1214/09-ss057. URL: https://doi.org/10.1214/09-ss057.

[71] Christos Davatzikos. “Why voxel-based morphometric analysis should be used with great caution when characterizing group differences”. In: *NeuroImage* 23.1 (Sept. 2004), pp. 17–20. DOI: 10.1016/j.neuroimage.2004.05.010. URL: https://doi.org/10.1016/j.neuroimage.2004.05.010.

[72] N. K. Batmanghelich, B. Taskar, and C. Davatzikos. “Generative-Discriminative Basis Learning for Medical Imaging”. In: *IEEE Transactions on Medical Imaging* 31.1 (Jan. 2012), pp. 51–69. DOI: 10.1109/tmi.2011.2162961. URL: https://doi.org/10.1109/tmi.2011.2162961.

[73] M Carroll, G CECCHI, I RISH, R GARG, and A RAO. “Prediction and interpretation of distributed neural activity with sparse models”. In: *NeuroImage* 44.1 (Jan. 2009), pp. 112–122. DOI: 10.1016/j.neuroimage.2008.08.020. URL: https://doi.org/10.1016/j.neuroimage.2008.08.020.

[74] Wei Jiang et al. “Machine Learning Methods Uncover Radiomorphologic Dose Patterns in Salivary Glands that Predict Xerostomia in Patients with Head and Neck Cancer”. In: *Advances in Radiation Oncology* 4.2 (Apr. 2019), pp. 401–412. DOI: 10.1016/j.adro.2018.11.008. URL: https://doi.org/10.1016/j.adro.2018.11.008.

[75] Yue Guo et al. “Spatial Radiation Dose Influence on Xerostomia Recovery and Its Comparison to Acute Incidence in Patients With Head and Neck Cancer”. In: *Advances in Radiation Oncology* 5.2 (Mar. 2020), pp. 221–230. DOI: 10.1016/j.adro.2019.08.009. URL: https://doi.org/10.1016/j.adro.2019.08.009.

[76] Serena Monti et al. “Voxel-based analysis unveils regional dose differences associated with radiation-induced morbidity in head and neck cancer patients”. In: *Scientific Reports* 7.1 (Aug. 2017). DOI: 10.1038/s41598-017-07586-x. URL: https://doi.org/10.1038/s41598-017-07586-x.

[77] Tom Vercauteren, Xavier Pennec, Aymeric Perchant, and Nicholas Ayache. “Symmetric Log-Domain Diffeomorphic Registration: A Demons-Based Approach”. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008*.

Springer Berlin Heidelberg, 2008, pp. 754–761. DOI: 10.1007/978-3-540-85988-8_90. URL: https://doi.org/10.1007/978-3-540-85988-8_90.

[78] William Beasley et al. “Image-based Data Mining to Probe Dosimetric Correlates of Radiation-induced Trismus”. In: *International Journal of Radiation Oncology∗Biology∗Physics* 102.4 (Nov. 2018), pp. 1330–1338. DOI: 10.1016/j.ijrobp.2018.05.054. URL: https://doi.org/10.1016/j.ijrobp.2018.05.054.

[79] Marc Modat, Gerard R. Ridgway, Zeike A. Taylor, Manja Lehmann, Josephine Barnes, David J. Hawkes, Nick C. Fox, and Sébastien Ourselin. “Fast free-form deformation using graphics processing units”. In: *Computer Methods and Programs in Biomedicine* 98.3 (June 2010), pp. 278–284. DOI: 10.1016/j.cmpb.2009.09.002. URL: https://doi.org/10.1016/j.cmpb.2009.09.002.

[80] Andriy Myronenko and Xubo Song. “Point Set Registration: Coherent Point Drift”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32.12 (Dec. 2010), pp. 2262–2275. DOI: 10.1109/tpami.2010.46. URL: https://doi.org/10.1109/tpami.2010.46.

[81] Oscar Acosta, Gael Drean, Juan D Ospina, Antoine Simon, Pascal Haigron, Caroline Lafond, and Renaud de Crevoisier. “Voxel-based population analysis for correlating local dose and rectal toxicity in prostate cancer radiotherapy”. In: *Physics in Medicine and Biology* 58.8 (Mar. 2013), pp. 2581–2595. DOI: 10.1088/0031-9155/58/8/2581. URL: https://doi.org/10.1088/0031-9155/58/8/2581.

[82] J.-P. Thirion. “Image matching as a diffusion process: an analogy with Maxwell’s demons”. In: *Medical Image Analysis* 2.3 (Sept. 1998), pp. 243–260. DOI: 10.1016/s1361-8415(98)80022-4. URL: https://doi.org/10.1016/s1361-8415(98)80022-4.

[83] Eugenia Mylona et al. “Voxel-Based Analysis for Identification of Urethrovesical Subregions Predicting Urinary Toxicity After Prostate Cancer Radiation Therapy”. In: *International Journal of Radiation Oncology∗Biology∗Physics* 104.2 (June 2019), pp. 343–354. DOI: 10.1016/j.ijrobp.2019.01.088. URL: https://doi.org/10.1016/j.ijrobp.2019.01.088.

[84] Gaël Dréan, Oscar Acosta, Caroline Lafond, Antoine Simon, Renaud de Crevoisier, and Pascal Haigron. “Interindividual registration and dose mapping for voxelwise population analysis of rectal toxicity in prostate cancer radiotherapy”. In: *Medical Physics* 43.6 (May 2016), pp. 2721–2730. DOI: 10.1118/1.4948501. URL: https://doi.org/10.1118/1.4948501.

[85] Giuseppe Palma et al. “A Voxel-Based Approach to Explore Local Dose Differences Associated With Radiation-Induced Lung Damage”. In: *International Journal of Radiation Oncology∗Biology∗Physics* 96.1 (Sept. 2016), pp. 127–133. DOI: 10.1016/j.ijrobp.2016.04.033. URL: https://doi.org/10.1016/j.ijrobp.2016.04.033.

[86] Giuseppe Palma et al. “Spatial Dose Patterns Associated With Radiation Pneumonitis in a Randomized Trial Comparing Intensity-Modulated Photon Therapy With Passive Scattering Proton Therapy for Locally Advanced Non-Small Cell Lung Cancer”. In: *International Journal of Radiation Oncology∗Biology∗Physics* 104.5 (Aug. 2019), pp. 1124–1132. DOI: 10.1016/j.ijrobp.2019.02.039. URL: https://doi.org/10.1016/j.ijrobp.2019.02.039.

[87] M A Hernán. “A definition of causal effect for epidemiological research”. In: *Journal of Epidemiology & Community Health* 58.4 (2004), pp. 265–271. ISSN: 0143-005X. DOI: 10.1136/jech.2002.006361. eprint: https://jech.bmj.com/content/58/4/265.full.pdf. URL: https://jech.bmj.com/content/58/4/265.

[88] Judea Pearl. “Causation, Action, and Counterfactuals”. In: *Logic and Scientific Methods*. Springer Netherlands, 1997, pp. 355–375. DOI: 10.1007/978-94-017-0487-8_18. URL: https://doi.org/10.1007/978-94-017-0487-8_18.

[89] Judea Pearl. “Causal inference in statistics: An overview”. In: *Statistics Surveys* 3.none (2009), pp. 96–146. DOI: 10.1214/09-SS057. URL: https://doi.org/10.1214/09-SS057.

[90] David Sontag. “Lecture 15: Causal Inference, Part 2”. MIT Lecture - Machine Learning for Healthcare. 2020. URL: https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-s897-machine-learning-for-healthcare-spring-2019/lecture-videos/lecture-15-causal-inference-part-2/.

[91] Jennifer Hill. “Bayesian Nonparametric Modeling for Causal Inference”. In: *Journal of Computational and Graphical Statistics* 20 (Mar. 2011), pp. 217–240. DOI: 10.1198/jcgs.2010.08162.

[92] Susan Athey and Guido Imbens. *Recursive Partitioning for Heterogeneous Causal Effects*. 2015. arXiv: 1504.01132 [stat.ML].

[93] Stefan Wager and Susan Athey. *Estimation and Inference of Heterogeneous Treatment Effects using Random Forests*. 2017. arXiv: 1510.04342 [stat.ME].

[94] Patrik Hoyer, Dominik Janzing, Joris Mooij, Jonas Peters, and Bernhard Schölkopf. "Nonlinear causal discovery with additive noise models". In: *Advances in Neural Information Processing Systems 21 - Proceedings of the 2008 Conference*. Jan. 2008, pp. 689–696.

[95] Corwin M. Zigler, Francesca Dominici, and Yun Wang. "Estimating causal effects of air quality regulations using principal stratification for spatially correlated multivariate intermediate outcomes". In: *Biostatistics* 13.2 (Jan. 2012), pp. 289–302. DOI: 10.1093/biostatistics/kxr052. URL: https://doi.org/10.1093/biostatistics/kxr052.

[96] Fredrik D. Johansson, Uri Shalit, and David Sontag. *Learning Representations for Counterfactual Inference*. 2018. arXiv: 1605.03661 [stat.ML].

[97] Uri Shalit, Fredrik D. Johansson, and David Sontag. *Estimating individual treatment effect: generalization bounds and algorithms*. 2017. arXiv: 1606.03976 [stat.ML].

[98] David Lopez-Paz, Robert Nishihara, Soumith Chintala, Bernhard Schölkopf, and Léon Bottou. *Discovering Causal Signals in Images*. 2017. arXiv: 1605.08179 [stat.ML].

[99] Judea Pearl. "Causal diagrams for empirical research". In: *Biometrika* 82.4 (1995), pp. 669–688. DOI: 10.1093/biomet/82.4.669. URL: https://doi.org/10.1093/biomet/82.4.669.

[100] Phillip Green Wright. "The Tariff on Animal and Vegetable Oils". In: *International Commercial Policies* (1928).

[101] Stephen R Cole, Robert W Platt, Enrique F Schisterman, Haitao Chu, Daniel Westreich, David Richardson, and Charles Poole. "Illustrating bias due to conditioning on a collider". In: *International Journal of Epidemiology* 39.2 (Nov. 2009), pp. 417–420. DOI: 10.1093/ije/dyp334. URL: https://doi.org/10.1093/ije/dyp334.

[102] Judea Pearl. "On the Testability of Causal Models with Latent and Instrumental Variables". In: *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1995, pp. 435–443. ISBN: 1558603859.

[103] Hui Zou. "The Adaptive Lasso and Its Oracle Properties". In: *Journal of the American Statistical Association* 101.476 (Dec. 2006), pp. 1418–1429. DOI: 10.1198/016214506000000735. URL: https://doi.org/10.1198/016214506000000735.

[104] R. Tibshirani. "Regression Shrinkage and Selection via the Lasso". In: *Journal of the Royal Statistical Society (Series B)* 58 (1996), pp. 267–288.

[105] Arthur E. Hoerl and Robert W. Kennard. “Ridge Regression: Biased Estimation for Nonorthogonal Problems”. In: *Technometrics* 12.1 (Feb. 1970), pp. 55–67. DOI: 10.1080/00401706.1970.10488634. URL: https://doi.org/10.1080/00401706.1970.10488634.

[106] Hastie. *Statistical Learning With Sparsity The Lasso And Generalizations*. English. Paperback. TF INDIA, 2017. Chap. 2, pp. 7–28. ISBN: 978-1138044975.

[107] Nicolai Meinshausen and Peter Bühlmann. “High-dimensional graphs and variable selection with the Lasso”. In: *The Annals of Statistics* 34.3 (June 2006). DOI: 10.1214/009053606000000281. URL: https://doi.org/10.1214/009053606000000281.

[108] Jianqing Fan and Runze Li. “Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties”. In: *Journal of the American Statistical Association* 96.456 (Dec. 2001), pp. 1348–1360. DOI: 10.1198/016214501753382273. URL: https://doi.org/10.1198/016214501753382273.

[109] *Biases Archive*. 2021. URL: https://catalogofbias.org/biases/ (visited on 06/23/2021).

[110] Lisa Van den Bosch et al. “Key challenges in normal tissue complication probability model development and validation: towards a comprehensive strategy”. In: *Radiotherapy and Oncology* 148 (July 2020), pp. 151–156. DOI: 10.1016/j.radonc.2020.04.012. URL: https://doi.org/10.1016/j.radonc.2020.04.012.

[111] Marcel van Herk. “Errors and margins in radiotherapy”. In: *Seminars in Radiation Oncology* 14.1 (Jan. 2004), pp. 52–64. DOI: 10.1053/j.semradonc.2003.10.003. URL: https://doi.org/10.1053/j.semradonc.2003.10.003.

[112] L.E.A. Shelley et al. “Delivered dose can be a better predictor of rectal toxicity than planned dose in prostate radiotherapy”. In: *Radiotherapy and Oncology* 123.3 (June 2017), pp. 466–471. DOI: 10.1016/j.radonc.2017.04.008. URL: https://doi.org/10.1016/j.radonc.2017.04.008.

[113] Vincent Grégoire et al. “Delineation of the neck node levels for head and neck tumors: A 2013 update. DAHANCA, EORTC, HKNPCSG, NCIC CTG, NCRI, RTOG, TROG consensus guidelines”. In: *Radiotherapy and Oncology* 110.1 (Jan. 2014), pp. 172–181. DOI: 10.1016/j.radonc.2013.10.010. URL: https://doi.org/10.1016/j.radonc.2013.10.010.

[114] Coen Rasch, Isabelle Barillot, Peter Remeijer, Adriaan Touw, Marcel van Herk, and Joos V Lebesque. “Definition of the prostate in CT and MRI: a multi-observer study”. In: *International Journal of Radiation Oncology∗Biology∗Physics* 43.1

(Jan. 1999), pp. 57–66. DOI: 10.1016/s0360-3016(98)00351-4. URL: https://doi.org/10.1016/s0360-3016(98)00351-4.

[115] Wendy Jeanneret-Sozzi, Raphaël Moeckli, Jean-François Valley, Abderrahim Zouhair, Esat Mahmut Ozsahin, René-Olivier Mirimanoff, and on Behalf SASRO∗ of SASRO∗. "The Reasons for Discrepancies in Target Volume Delineation". In: *Strahlentherapie und Onkologie* 182.8 (Aug. 2006), pp. 450–457. DOI: 10.1007/s00066-006-1463-6. URL: https://doi.org/10.1007/s00066-006-1463-6.

[116] M Mukesh et al. "Interobserver variation in clinical target volume and organs at risk segmentation in post-parotidectomy radiotherapy: can segmentation protocols help?" In: *The British Journal of Radiology* 85.1016 (Aug. 2012), e530–e536. DOI: 10.1259/bjr/66693547. URL: https://doi.org/10.1259/bjr/66693547.

[117] A. Meijster, J. B. T. M. Roerdink, and W. H. Hesselink. "A General Algorithm for Computing Distance Transforms in Linear Time". In: *Mathematical Morphology and its Applications to Image and Signal Processing*. Kluwer Academic Publishers, 2002, pp. 331–340. DOI: 10.1007/0-306-47025-x_36. URL: https://doi.org/10.1007/0-306-47025-x_36.

[118] M.A. Hernan and J.M. Robins. "Observational Studies". In: *Causal Inference*. Taylor & Francis, 2020, pp. 25–37.

[119] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. "Regularization Paths for Generalized Linear Models via Coordinate Descent". In: *Journal of Statistical Software* 33.1 (2010), pp. 1–22. URL: https://www.jstatsoft.org/v33/i01/.

[120] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2021. URL: https://www.R-project.org/.

[121] Razieh Nabi, Todd McNutt, and Ilya Shpitser. *Semiparametric Causal Sufficient Dimension Reduction Of High Dimensional Treatments*. 2020. arXiv: 1710.06727 [stat.ME].

[122] Leilani H. Gilpin, David Bau, Ben Z. Yuan, Ayesha Bajwa, Michael Specter, and Lalana Kagal. *Explaining Explanations: An Overview of Interpretability of Machine Learning*. 2019. arXiv: 1806.00069 [cs.AI].

[123] Reid Pryzant, Kelly Shen, Dan Jurafsky, and Stefan Wagner. "Deconfounded Lexicon Induction for Interpretable Social Science". In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. Association for

Computational Linguistics, 2018. DOI: 10.18653/v1/n18-1146. URL: https://doi.org/10.18653/v1/n18-1146.

[124] Rikiya Yamashita, Mizuho Nishio, Richard Kinh Gian Do, and Kaori Togashi. "Convolutional neural networks: an overview and application in radiology". In: *Insights into Imaging* 9.4 (June 2018), pp. 611–629. DOI: 10.1007/s13244-018-0639-9. URL: https://doi.org/10.1007/s13244-018-0639-9.

[125] Jinpyo Kim, Wooekun Jung, Hyungmo Kim, and Jaejin Lee. *CyCNN: A Rotation Invariant CNN using Polar Mapping and Cylindrical Convolution Layers*. 2020. arXiv: 2007.10588 [cs.CV].

[126] Nanne van Noord and Eric Postma. "Learning scale-variant and scale-invariant features for deep image classification". In: *Pattern Recognition* 61 (Jan. 2017), pp. 583–592. DOI: 10.1016/j.patcog.2016.06.005. URL: https://doi.org/10.1016/j.patcog.2016.06.005.

[127] Japhet M. Gilyoma, Peter F. Rambau, Nestory Masalu, Neema M. Kayange, and Phillipo L. Chalya. "Head and neck cancers: a clinico-pathological profile and management challenges in a resource-limited setting". In: *BMC Research Notes* 8.1 (Dec. 2015). DOI: 10.1186/s13104-015-1773-9. URL: https://doi.org/10.1186/s13104-015-1773-9.

[128] Fotios Loupakis et al. "Primary Tumor Location as a Prognostic Factor in Metastatic Colorectal Cancer". In: *JNCI: Journal of the National Cancer Institute* 107.3 (Feb. 2015). DOI: 10.1093/jnci/dju427. URL: https://doi.org/10.1093/jnci/dju427.

[129] Hyun Woo Lee, Chang-Hoon Lee, and Young Sik Park. "Location of stage I-III non-small cell lung cancer and survival rate: Systematic review and meta-analysis". In: *Thoracic Cancer* 9.12 (Sept. 2018), pp. 1614–1622. DOI: 10.1111/1759-7714.12869. URL: https://doi.org/10.1111/1759-7714.12869.

[130] N. Joseph et al. "PO-0754: Whole body Integral dose is associated with radiotherapy related fatigue in prostate cancer". In: *Radiotherapy and Oncology* 119 (Apr. 2016), S352–S353. DOI: 10.1016/s0167-8140(16)32004-7. URL: https://doi.org/10.1016/s0167-8140(16)32004-7.

[131] David Eriksson and Torgny Stigbrand. "Radiation-induced cell death mechanisms". In: *Tumor Biology* 31.4 (May 2010), pp. 363–372. DOI: 10.1007/s13277-010-0042-8. URL: https://doi.org/10.1007/s13277-010-0042-8.

[132] Franz Rödel, Benjamin Frey, Gabriele Multhoff, and Udo Gaipl. “Contribution of the immune system to bystander and non-targeted effects of ionizing radiation”. In: *Cancer Letters* 356.1 (Jan. 2015), pp. 105–113. DOI: 10.1016/j.canlet.2013.09.015. URL: https://doi.org/10.1016/j.canlet.2013.09.015.

[133] Silvia C Formenti and Sandra Demaria. “Systemic effects of local radiotherapy”. In: *The Lancet Oncology* 10.7 (July 2009), pp. 718–726. DOI: 10.1016/s1470-2045(09)70082-8. URL: https://doi.org/10.1016/s1470-2045(09)70082-8.

[134] Victor Veitch, Dhanya Sridhar, and David M. Blei. *Adapting Text Embeddings for Causal Inference*. 2020. arXiv: 1905.12741 [cs.LG].

[135] H.-J. Guchelaar, A. Vermes, and J. H. Meerwaldt. “Radiation-induced xerostomia: pathophysiology, clinical course and supportive treatment”. In: *Supportive Care in Cancer* 5.4 (July 1997), pp. 281–288. DOI: 10.1007/s005200050075. URL: https://doi.org/10.1007/s005200050075.

[136] Christopher M Nutting et al. “Parotid-sparing intensity modulated versus conventional radiotherapy in head and neck cancer (PARSPORT): a phase 3 multicentre randomised controlled trial”. In: *The Lancet Oncology* 12.2 (Feb. 2011), pp. 127–136. DOI: 10.1016/s1470-2045(10)70290-4. URL: https://doi.org/10.1016/s1470-2045(10)70290-4.

[137] Sarah L. Gulliford et al. “Dosimetric explanations of fatigue in head and neck radiotherapy: An analysis from the PARSPORT Phase III trial”. In: *Radiotherapy and Oncology* 104.2 (Aug. 2012), pp. 205–212. DOI: 10.1016/j.radonc.2012.07.005. URL: https://doi.org/10.1016/j.radonc.2012.07.005.

[138] Benedikt M. Pötscher and Ulrike Schneider. “On the distribution of the adaptive LASSO estimator”. In: *Journal of Statistical Planning and Inference* 139.8 (Aug. 2009), pp. 2775–2790. ISSN: 0378-3758. DOI: 10.1016/j.jspi.2009.01.003. URL: http://dx.doi.org/10.1016/j.jspi.2009.01.003.

[139] Alberto Caron, Gianluca Baio, and Ioanna Manolopoulou. *Sparse Bayesian Causal Forests for Heterogeneous Treatment Effects Estimation*. 2021. arXiv: 2102.06573 [stat.ME].

[140] Judea Pearl. “The seven tools of causal inference, with reflections on machine learning”. In: *Communications of the ACM* 62.3 (Feb. 2019), pp. 54–60. DOI: 10.1145/3241036. URL: https://doi.org/10.1145/3241036.