In vitro-based Modelling of Relative Biological Effectiveness in Proton Therapy

Erlend Lyngholm

Supervisors:

Kristian Smeland Ytre-Hauge

Therese Berge Sjursen

Master thesis in medical physics

Department of Physics and Technology

University of Bergen

June 14, 2021

Acknowledgements

First of all, I would like to thank my supervisor Kristian Smeland Ytre-Hauge for great guidance throughout this work. Thank you for taking your time to set up regular meetings with interesting discussions and for providing valuable feedback on my work.

Thanks to all my fellow students, who have made my time as a student at UiB a memorable time. Especially thanks to Erlend Aakvaag for countless lunch breaks filled with interesting conversations, both study-related and about everyday-life, not to mention a lot of laughter. You have been a valuable friend, keeping my motivation up throughout our years studying together. Thank you Maja Hustad for being such a good company on the longest and most trying days of exam reading.

Thanks to my sister Camilla Lyngholm for not going completely insane from living with me the past year and for all the good times we have shared during this time.

Finally, I want to thank my parents for all the love and support they have given me throughout my hole life, always checking in on me and expressing genuine interest in what I am doing and how I am at all times.

Bergen, June 2021

..

Abstract

Many cancer patients can benefit from proton therapy, as the treatment modality have been shown to provide a more conformal irradiation of the target while sparing surrounding healthy tissue, compared to photon therapy. The concept of the relative biological effectiveness allows for appliance of the large amount of experience from photon therapy when considering irradiation with protons. In proton therapy treatment today, a constant proton RBE of 1.1 relative to high-energy photons is applied, assuming protons to be 10% more effective for the same physical dose. However, the proton RBE have been shown to vary with multiple physical and biological factors including the deposited physical dose, irradiated tissue and radiation quality of the beam. Precise modelling of the proton RBE have therefore become an important field of study.

This work presents a comprehensive analysis of proton RBE dependencies, using a large up-to date database of *in vitro* data points from proton irradiation experiments. The analysis focus on the RBE dependence on the reference radiation fractionation sensitivity ($(\alpha/\beta)_x$), and the radiation quality of the beam, quantified by the linear energy transfer (LET). Additionally, potential differences between RBE of monoenergetic and broad energy proton beams (for a given average LET value) was explored, and the effects of variations in the data selection procedure were investigated. The widely used assumption of an inverse proportionality of RBE with $(\alpha/\beta)_x$ was investigated through linear fitting and the fits were compared with previous published models applying this assumption. The RBE-LET relationship was investigated by fitting of polynomials from 1st to 4th degree and polynomials of 1st and 2nd degree in combination with exponential functions. Additional fitting was performed on different restricted databases in terms of reduced range of included LET and $(\alpha/\beta)_x$ values. Furthermore, the database was seen to be highly imbalanced, i.e. data points were not evenly distributed over the $(\alpha/\beta)_x$ and LET range, and fitting was therefore also performed using an iterative sampling procedure to compensate for this. Selected

regression fits were implemented as RBE models and the RBE estimates from these, both as a function of LET, dose and $(\alpha/\beta)_x$ and for a simulated spread out Bragg peak (SOBP) scenario were compared to the estimates of three published RBE models.

Linear fitting on both the unrestricted and restricted databases showed a trend of a steeper increase in RBE with increasing LET when an SOBP beam is used, compared to applying a monoenergetic beam. Analysis of the $(\alpha/\beta)_x$ dependency implied that appliance of the $1/(\alpha/\beta)_x$ assumption might lead to overestimated RBE at $(\alpha/\beta)_x$ values below approximately 6 Gy, and underestimated RBE predictions for higher $(\alpha/\beta)_x$. Therefore, a less strong dependency of $(\alpha/\beta)_x$ for the proton RBE should be considered further. Fitting on restricted databases showed that the model output is largely affected by the included LET values in the model database, as lowering the upper limit of included LET values resulted in a smaller slope in the applied linear RBE-LET relationship, indicating a non-linear relationship. Fitting on a balanced database in terms of $(\alpha/\beta)_x$ largely affected the fitting output. It is however debatable whether the applied procedure is optimal, as some of the data points in regions with fewer data might be assigned too large weights in the fitting. Non-linear fitting generally yielded better performance than linear fitting on the database, although the differences were marginal, and a linear dependency could not be rejected. The quartic fitting function, f_{LQCq}, was implied to best model the data, and comparable performance was indicated for the fitting functions f_{LQ}, f_{LQC}, f_{LE} and f_{LQE}. Implementing three of the linear models and the f_{LQCq} and f_{LQC} models from this work in a SOBP scenario, all models gave RBE estimates that largely disagreed with a constant RBE=1.1. All the considered models showed only small differences from a constant RBE=1.1 in the entrance region. The models deviated more from each other at the proximal part of the SOBP, and all models predicted an increasing RBE across the SOBP. The RBE increase given by the f_{LQC} model was steeper than the other models, and although the f_{LQCq} model was above this model at the proximal part of the SOBP, the f_{LQC} model gave higher RBE estimates than the f_{LQCq} model at the distal

end of the SOBP. The RBE estimates of the linear model obtained with data having all values of LET generally gave higher RBE estimates than the two linear models obtained with stricter restrictions on both LET and $(\alpha/\beta)_x$. The analyses of the present work showed that database selection and regression procedure largely affects the outcome of RBE modelling. This can explain the observed differences between previously published RBE models.

Contents

1. Introduction

In 2020 there was 19.3 million new cases of cancer worldwide [1], and 1 in 6 deaths were caused by the disease. The main cancer treatment modalities include surgery, chemotherapy, immunotherapy and radiotherapy. About 50% of all cancer patients are eligible for radiotherapy at some point during the course of disease [2]. The modality can be used for curative or palliative treatments. While the latter aims at relieving the suffering of patients with non-curable cancer, the goal of curative radiotherapy is to kill or inactivate tumor cells while sparing the surrounding healthy tissue. External beam therapy was first performed with photon radiation, and this modality have been used to treat cancer patients for almost 13 decades. During the 20th century, radiotherapy using other particles such as protons, neutrons and electrons has also been explored and used for cancer treatment. The maximum dose deposition of protons is in the last centimeter of the range, forming the so-called Bragg peak. Protons therefore avoid exit doses and greatly reduce the entrance dose compared to photons, providing an intrinsic increase in dose conformity.

1.1 Radiotherapy

Radiotherapy is one of the most widely used modalities in cancer treatment. Radiation treatment with high-energetic photons have dominated the field for the past decades, and photon therapy is still the most frequently used modality today. Proton therapy has been established as an important radiotherapy modality, as it can offer improved sparing of normal tissue compared to conventional radiation therapy with photons. The main benefits of proton therapy are seen from the characteristics of the physical dose deposition of protons, depicted in Figure 1. While the increasing dose towards the tumor volume can provide sparing normal tissue before the target, the finite range of the proton beam can ensure little dose deposition beyond the target volume. In addition, protons have an increased biological effectiveness compared to photon therapy, meaning that protons produce more damage than photons from the same physical dose. This difference is referred to as the relative biological effectiveness (RBE), and is

accounted for in clinical proton therapy by applying a constant RBE-factor of 1.1, thus assuming protons to be 10% more efficient than photons.

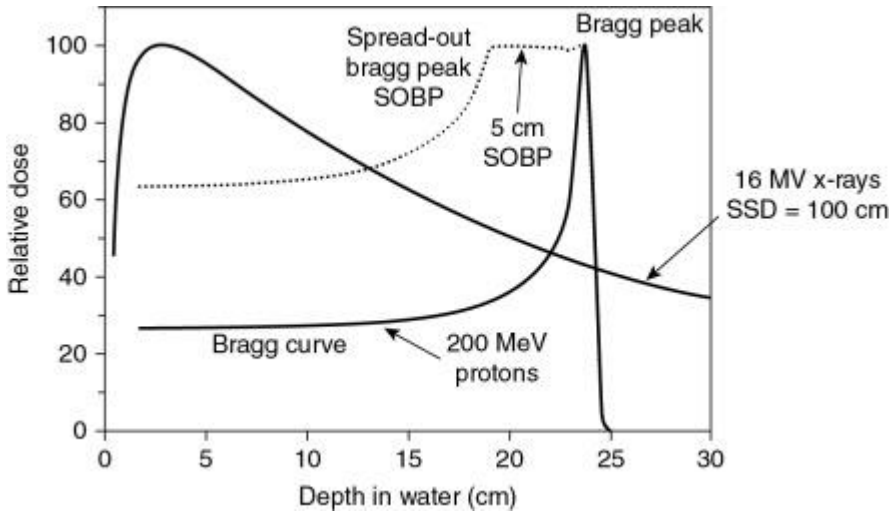

Figure 1: Depth dose curves for a 200 MeV proton beam: both unmodulated and with a 5 cm spread out Bragg peak (SOBP), compared with a 16 MV x-ray beam. The curves are normalized in each case to 100 at maximum dose. (Modified from Mohan and Grosshans [3])

More recent clinical and experimental data show that the applied and used constant RBE-factor is an oversimplification. An increasing amount of experience and data show that the proton RBE varies spatially within the patient, depending on numerous factors such as the linear energy transfer (LET), the dose level, physiological and biological factors and clinical endpoint [4]. The use of a generic, spatially invariant RBE-factor within tumors and normal tissues disregards the evidence of a spatially variable RBE. In Figure 2, evidence that the proton RBE varies with depth is shown for experimental *in vitro* data from irradiation of Chinese hamster cells with clinical proton beams and photon beams for reference.

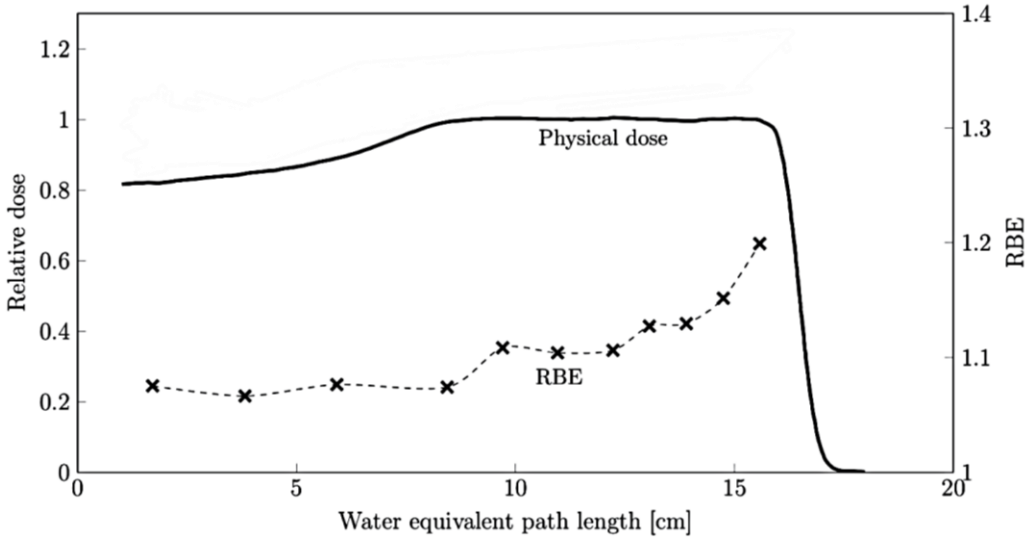

Figure 2: An example of a physical depth dose distribution for protons, found by Monte Carlo simulations. The variable RBE value is shown by the points and the dashed curve. The lines between the points are only for guidance. As seen, the RBE is therefore not constant but increasing with depth in this example. Data from *in vitro* cell irradiation experiments with Chinese hamster V79-WNRE cells, extracted from Wouters et al 2015 [5] and Polster et al 2015 [6]. (Modified from Rørvik et al 2019 [7])

Precise modelling of the proton RBE is therefore becoming an important task in order to enable better treatment planning and minimize treatment complications in proton therapy. Numerous RBE models have been developed over the past decade, aiming to account for the potential effects of a variable proton RBE. Most of the published proton RBE models are so-called phenomenological models. These are based on empirical data from *in vitro* proton irradiation of various cell lines and utilize the linear-quadratic (LQ) model with cell inactivation as a biological endpoint. The LQ model is widely used in radiation biology to describe the cell survival as a function of dose based on the tissue specific LQ parameters α and β. The modelling has been done using many different techniques, databases and endpoints and while a general agreement in RBE estimates is seen among most RBE models, there are also clear differences. The present uncertainties in the RBE models hampers the introduction of a variable RBE in clinical proton therapy treatment planning.

1.2 Project motivation

Proton RBE models generally predicts that the RBE increases with increasing LET and decreasing dose, while it is inversely proportional with the ratio of the tissue specific LQ parameters of the reference radiation, $(\alpha/\beta)_x$. However, there are still significant variations between the estimates of RBE from different models. As the models are based on different selections of data it is difficult to draw clear conclusions regarding the origin of the observed differences between RBE models, and which models gives the best predictions. In addition, new *in vitro* data has become available over the past few years which may contribute to more accurate RBE models.

The main goal of this project was therefore to use a holistic approach to proton RBE modelling by using all available *in vitro* data and explore how data selection, regression methods and model assumptions affect the RBE estimation. With this goal, a database containing *in vitro* proton data from published literature was collected and analysed. Data on heavier ions were also included in the database for comparative studies and future work. The collected proton data included both data points obtained with monoenergetic irradiation and data points obtained using a spread out Bragg peak (SOBP), i.e. a proton beam with a wide energy spectrum. The analysis focus on investigating the relationships of proton RBE with the type of irradiated tissue and the radiation quality of the beam by analysing the proton data in the framework of a phenomenological model. The tissue dependency is introduced through the LQ parameters of the photon radiation for each respective cell line, while the radiation quality is quantified by the LET.

To achieve the main objective stated above, the following secondary objectives were pursued:

- The RBE for a given averaged LET value might be different for monoenergetic and SOBP beam. Although this is a debated issue, it has not been thoroughly investigated. Many of the published proton RBE models are developed using

data only from experiments executed by a pure monoenergetic beam. This study therefore aims to identify different trends in RBE data based on experiments using monoenergetic irradiation and experiments using an SOBP.

- The widely used assumption that RBE is inversely proportional with $(\alpha/\beta)_x$ was originally tested on a small dataset [8]. Although many of the published RBE models apply this assumption, it's validity has not been tested in detail on a larger and up-to-date dataset. In this study a detailed analysis of the RBEs dependency on $(\alpha/\beta)_x$ was performed to test this assumption.
- Most published RBE models assume that the proton RBE is linearly dependent on LET. Although some experiments and analyses have indicated a non-linear LET dependency [9, 10], this have only to a small degree been explored. The present work therefore aims to determine if non-linear RBE models can better represent the available *in vitro* data compared to linear models.
- Previous models have included data covering different ranges in $(\alpha/\beta)_x$ and LET. This is likely to affect the RBE estimates and an analysis based on different database restrictions in terms of reducing the range of included $(\alpha/\beta)_x$ and LET values was performed with the objective of determining how this will affect the model output, potentially explaining differences between previously published RBE models.

2. Proton therapy

2.1 Radiation interactions and radiobiology

2.1.1 Proton interactions with matter

When protons traverse in matter they interact with the material primarily through three processes: inelastic Coulomb scattering, elastic Coulomb scattering and non-elastic nuclear reactions [11]. The protons interact with atomic electrons through inelastic collisions. This is the main source of proton energy loss. As the proton mass is far greater than the electron mass, the proton trajectory is not significantly affected by these interactions. The target electron is either excited to a higher energy state or ionized, depending on the amount of energy transferred in the collision. If the ionized electrons receive a sufficient amount kinetic energy from the initial proton, such secondary electrons may cause further ionization in the medium. The energy loss of a heavy charged particle is typically referred to as the particles stopping power and is well described by the Bethe-Block equation [12]:

$$-\frac{dE}{dx} = 2\pi N_a r_e^2 m_e c^2 \rho \frac{Z}{A}\frac{z^2}{\beta^2}\left[\ln\left(\frac{2m_e\gamma^2 v^2 W_{max}}{I^2}\right) - 2\beta^2 - \delta - 2\frac{C}{Z}\right] \quad (2.1)$$

where dE is the energy loss over a small distance dx given in cm, and the remaining variables are described as:

- N_a = Avogadro's number
- r_e = classical electron radius
- m_e = electron mass
- c = speed of light
- ρ = density of absorbing material
- Z = atomic number of absorbing material
- A = atomic mass of absorbing material

- $z =$ charge of the incident particle
- $\beta =$ relativistic velocity of the incident particle
- $\gamma =$ Lorentz factor
- $v =$ speed of the incident particle
- $W_{max} =$ maximum energy transfer in a single collision
- $I =$ mean excitation potential
- $\delta =$ density correction factor
- $C =$ shell correction factor

According to this formula, the stopping power is proportional to the square of the incident particle ion charge, while there is no dependence on the mass of the incident particle. The stopping power is inversely proportional to the velocity of the incident particle, meaning that the heavy ions will deposit most of their energy towards the end of their range. The resulting depth dose distribution of heavy ions is thus characterized by an exponential increase followed by a sharp dose fall-off, forming the characteristic Bragg peak curve which is discussed in more detail in section 2.1.4.

A proton passing close to an atomic nucleus will experience a repulsive Coulomb force resulting from the opposite charges of the proton and the nucleus. Due to the large mass of the nucleus, the proton is deflected from its originally straight-line trajectory. The proton energy loss from such elastic Coulomb interactions are typically small. Although a single scattering normally gives a negligible deflection, the sum of many such deflections can be significant. The term multiple Coulomb scattering (MCS) is commonly used for the observed effect of countless tiny deflections leading to a lateral broadening of the beam. A good implementation of MCS is essential for accurate calculation of dose distributions, i.e., in a treatment planning system.

The protons can also interact through non-elastic nuclear reactions with the atomic nucleus. Although they are less frequent, these interactions have a much more profound effect in terms of the faith of a single proton [11]. If the distance between the proton and the nucleus is sufficiently small and the Coulomb barrier is overcome,

the projectile proton enters the nucleus, which in turn may emit a proton, deuteron, triton or a heavier ion or one or more neutrons. The particles produced in nuclear interactions are called secondaries. These particles will contribute to the total dose delivered to the patient in a clinical setting. Secondary neutrons might cause dose depositions outside the target volume. As they do not carry charge, the neutrons do not interact electromagnetically and might travel beyond the range of the primary particles [13].

The three considered interactions of protons in matter are schematically illustrated in Figure 3. In addition to these, proton bremsstrahlung is theoretically possible, but this effect is negligible at therapeutic proton beam energies and it will not be considered here.

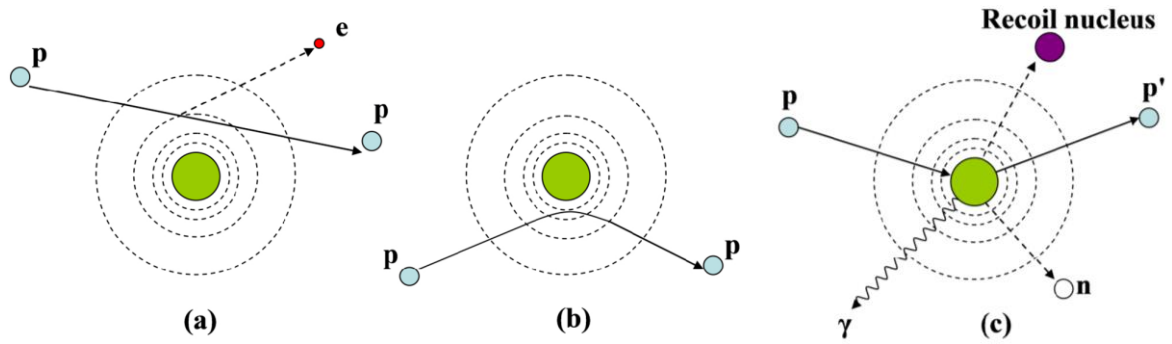

Figure 3: Proton interaction mechanisms: (a) energy loss via inelastic Coulomb interaction with atomic electron, (b) deflection of proton trajectory by repulsive Coulomb elastic scattering with nucleus, (c) removal of primal proton and creation of secondaries via non-elastic nuclear interaction (p=proton, e=electron, n=neutron, γ=gamma rays) [11].

2.1.2 Photon interactions with matter

Photon interactions are briefly summarized here as the present work revolves around the biological effects of protons compared to photons which is the standard radiation for radiotherapy. Photons are used as reference radiation in proton cell irradiation experiments to obtain the *in vitro* data used in proton RBE analysis and modelling.

As a photon beam traverse matter, attenuation of the beam by the absorbing material is caused by five major types of interactions: photodisintegration, coherent scattering, the photoelectric effect, the Compton effect, and the pair production [14]. The first mentioned interaction type is only important at very high photon energies (> 10 MeV) and will not be considered here. The other four processes are illustrated in Figure 4. In coherent scattering (or Rayleigh scattering), a photon is scattered on an atomic electron. No energy is changed into electronic motion and no energy is absorbed in the medium in this process. The coherent scattering is probable in high-atomic-number materials and with photons of low energy. The process is thus of little importance in radiation therapy. The photoelectric effect is a phenomenon in which a photon is absorbed by an atom, and as a result one of its orbital electrons is ejected. The kinetic energy of the ejected photoelectron is equal to $h\nu - E_B$, where E_B is the binding energy of the electron. After the ejection of a photoelectron, the atom is left in an excited state. The vacancy created in the shell where the ejected electron was positioned can be filled by an outer orbital electron with the emission of a characteristic x-ray. There is also the possibility of emission of Auger electrons, which will occur when the energy released as a result of the outer electron filling the vacancy is given to another electron in a higher shell which is subsequently ejected. In the Compton process, the photon interacts with an atomic electron as though it were a "free" electron, that is, the binding energy of the electron is much less than the energy of the incident photon. In this interaction, the electron receives some of the photon energy and is emitted at an angle θ, while the photon is scattered at an angle ϕ, with reduced energy. In the process of pair production, the photon interacts strongly with the electromagnetic field of an atomic nucleus and gives up all its energy in the process of creating a pair consisting of an electron and a positron. This process is only possible if the energy of the photon is greater than 1.02 MeV, which is the minimum energy required to create the electron-positron pair given that the rest mass energy of these particles are each equivalent to 0.51 MeV.

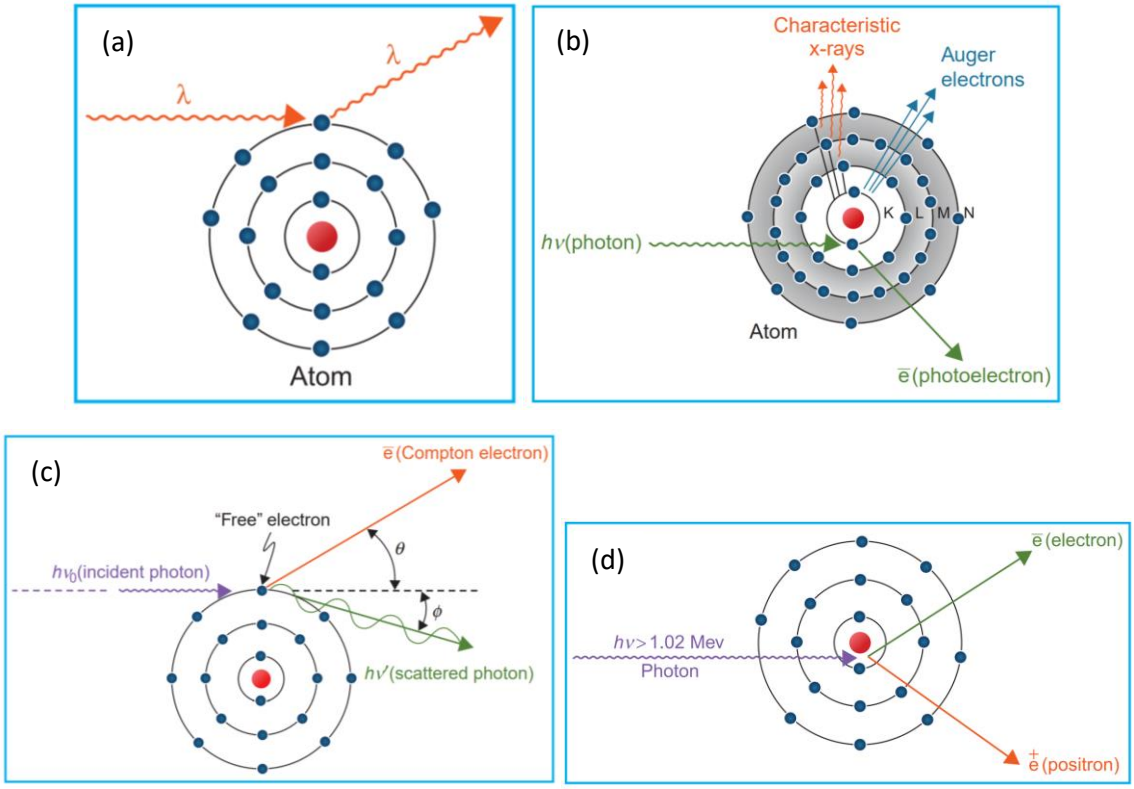

Figure 4: Schematic illustration of photon interaction mechanisms: (a) Coherent (Raleigh) scattering, (b) photoelectric effect, (c) Compton effect, (d) pair production [14].

2.1.3 Relative biological effectiveness

The use of proton therapy in cancer treatment has increased globally over the recent decades. The treatment modality can potentially provide a more conformal dose to the tumor, sparing more of the surrounding tissue, compared to photon radiation therapy [15]. Proton radiation is ionizing, and interacts with the tissue different to photons. Thus, the radiation quality for the two modalities is different. Giving the same amount of physical dose with photon- and proton therapy, the effect is normally higher for the latter. Thus, protons normally produce more biological damage for the same physical dose [4]. The potential difference in biological effectiveness must be considered when treating patients with different modalities. To acquire this, the relative biological effectiveness (RBE) is introduced as a scaling factor to reflect the variations in effect for the same physical dose. The RBE for a specific endpoint is defined as:

$$RBE(endpoint) \equiv \frac{D_x}{D_p}, \tag{2.2}$$

where D_x and D_p are the absorbed physical doses deposited by the reference photon radiation and the proton radiation, respectively.

There is a lot more gained experience in conventional photon therapy than in proton therapy. To utilize this experience in proton therapy, the physical doses must be scaled by the proton RBE to account for the difference in biological effectiveness. Today, proton therapy treatments are based on a proton RBE of 1.1 relative to high-energy photons. However, this generic, spatially invariant factor is only an assumption, disregarding the evidence that the proton RBE depends on a number of physical and biological factors. Experimental data have shown that the RBE varies with the Linear Energy Transfer (LET) (radiation quality?), deposited physical dose and physiological and biological properties including cell type, oxygen concentration and clinical endpoint [4]. In order to optimize treatment planning and minimize treatment complications, a better understanding of the RBE of proton radiation is needed. Over the last decades, numerous different RBE models have been developed to account for the potential effect of a variable RBE. The models can be divided into three major groups: Phenomenological models, plan-based models and mechanical models [16].

2.1.4 Depth-dose distributions

There are distinct differences in the dose depositions of photon and proton radiation. By comparing the depth-dose curves for the two radiation modalities in a water phantom, the differences can be studied in detail.

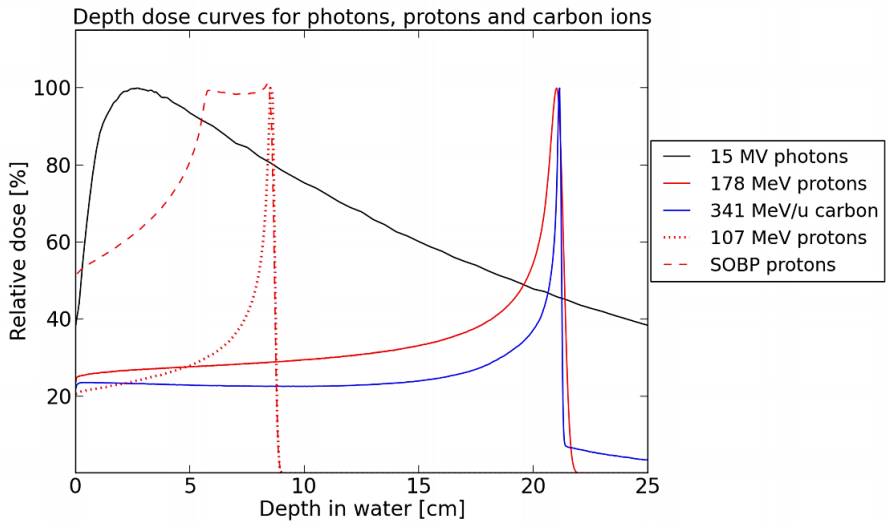

Figure 5: Depth dose curves for photons, protons and carbon ions, from Monte Carlo simulations. The curves represent the absorbed dose and do not take into account RBE values. The picture also shows a Spread Out Bragg Peak (SOBP) for protons covering a range of 3 cm with a homogeneous dose [13].

Figure 5 shows depth dose curves for photons, protons and carbon ions and illustrate the fundamental difference in dose deposition between photons and ions. The photon dose deposition increase with increasing depth up to the maximum dose located a few cm inside the phantom. After reaching the depth of maximum dose, the dose deposition begins to decrease. As seen in Figure 5, the dose deposition post to the maximum is slowly decreasing with depth, compared to the sharp build-up in front of the maximum. The depth dose curves for protons are characterized by a relatively low entrance dose. At some point of depth, the dose starts to increase exponentially with depth. After reaching the depth of maximum dose deposition (the Bragg peak), the curve has a sharp fall-off to zero. For both photon- and proton radiation, the depth of the dose maximum will depend on the beam energy. This is seen from the two depth dose curves for protons in Figure 5, showing that the position of the Bragg peak is a function of the primary beam energy. In proton therapy, the energy and intensity of the beam are gradually varied to form a series of overlapping Bragg Peaks, adding up to a homogenous dose deposited over the whole tumor or target volume. The sum of

overlapping Bragg peaks is called a Spread Out Bragg Peak (SOBP), illustrated by the dashed line in Figure 5. Thus, the depth dose curve for a therapeutic proton beam will have an increased entrance dose compared to the depth dose curve for monoenergetic protons. This is seen in Figure 5, where the depth dose curve for the SOBP even has a higher entrance dose than the one for photons. The dose deposition for carbon is similar to the one for protons, but the carbon curve has a more narrow Bragg beak, and a tail of relatively low doses post to the sharp dose fall-off. The tail in the distal end of the dose distribution is caused by nuclear fragments from inelastic collisions between the carbon ions and the nuclei in the material they traverse.

The characteristics observed from the (comparison of the) depth-dose curves indicates that proton therapy could potentially provide a more conformal irradiation of a tumor, and a lower dose to the surrounding healthy tissue, compared to conventional radiation therapy. The sharp dose fall-off after the Bragg peak can lower the dose to an organ at risk (OAR) positioned near the distal end of a tumor. These favourable qualities of the proton dose distribution are the main motivation for application of proton therapy in cancer treatment.

2.1.5 Linear energy transfer

Even though two different radiation modalities deposit the same physical macroscopic dose within the tissue, differences in the ionization density and track structure can cause different pattern of dose deposition [16]. This property is described by the radiation quality, which is normally quantified by the linear energy transfer (LET) of the radiation. The LET describes the rate of energy depositions along the particle track, thus describing the ionization density of a radiation beam. LET is defined as

$$LET = \frac{dE}{dl}, \tag{2.3}$$

where dE is the infinitesimal amount of mean energy transferred locally from the proton to the tissue, given in keV, and dl is an infinitesimal part of the proton track,

given in µm. (The LET reflects the biological effectiveness of the radiation. It is not a directly measurable quantity, and the experimental equivalent to LET is lineal energy (l).)

In a clinical setting, the radiation field within the patient consist of multiple types of particles with a range of energies, rather than only monoenergetic protons. The radiation quality can therefore be described by a dose weighted spectrum from protons of different LET values at every spatial location, instead of a single quantity. The dose weighted LET spectrum is defined such that the sum of all dose compositions is normalized to 1. Each LET entry in the spectrum is weighted by its dose deposition. For simplicity, it is most common to describe the spectrum by the dose averaged LET value, LET_d. This is the most commonly reported and applied LET value in radiation therapy, both in clinical settings and in radiobiological models.

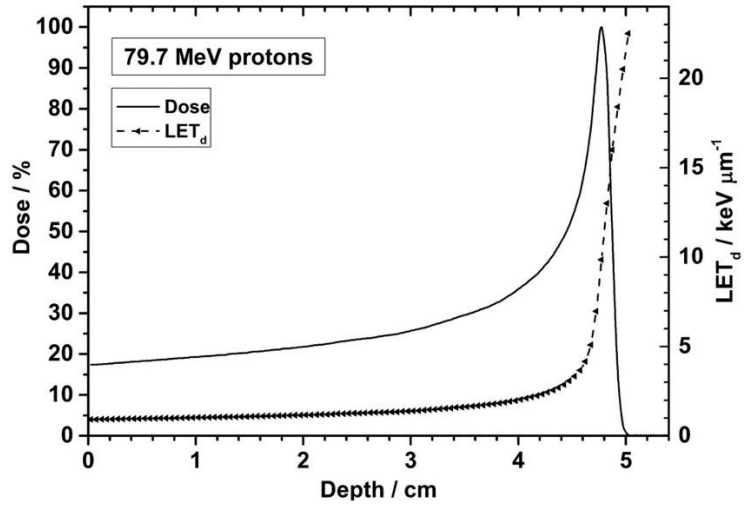

Figure 6: Depth-Dose and depth-LET curves for a 79.7 MeV monoenergetic proton beam. The LET increases slowly up to about one cm before the Bragg peak and steeply at points beyond the peak [17].

For ion radiation, the LET_d is inversely proportional to the beam energy. Thus, for a proton beam, the LET is increasing towards the distal end of the beam range, as shown in Figure 6. Experimental results show that the RBE increases with increasing LET [4]. Therefore, as LET values increase with the depth of the beam, the RBE increases towards the distal end of the SOBP [10]. The variation in RBE along the treatment

depth results in an increased biological dose at greater depths, as illustrated in Figure 7.

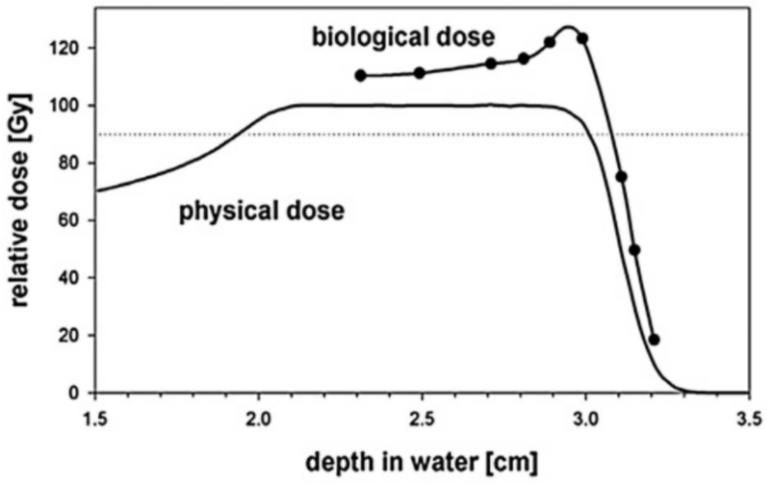

Figure 7: An example of a physical depth dose distribution for a SOBP, from Monte Carlo simulations. The biological dose is plotted above the physical dose, indicating the higher effectiveness for protons over photons [16].

The increased RBE at the distal end of the SOBP might cause an increased biological dose to OARs located distal to the target volume. In such cases, the biological dose to the OARs can be underestimated if a constant RBE of 1.1 is used. As a result, the normal tissue complication probability (NTCP) of a patient can increase. To prevent the underestimation of biological dose, it is therefore of clinical interest to take the variation of RBE along the treatment depth into account in the treatment planning. Several RBE models have been developed to predict the biological dose delivered to patients [6, 18].

2.1.6 Cell survival and the linear-quadratic model

When a cell is exposed to ionizing radiation, the DNA in the cell nucleus is the main target. The damage on the DNA caused by the radiation can be divided into two types, single track events and two-track events. A single track event cause a non-repairable damage, while the damage from a two-track event is repairable.

In RBE experiments, the most common measured endpoint is cell survival of irradiated *in vitro* cells. This is also the basis for most RBE models. Many mathematical models have been developed to describe the cell survival curve for tissue under the exposure of radiation. The linear-quadratic model (LQ model) is a general radiobiological model widely used in radiation biology to describe the cell survival. It can describe the effect of radiation on multiple endpoints, and provides a simple relationship between cell survival and delivered dose. According to the LQ model, the survival fraction of cells (S) after irradiation by a single dose D is given by:

$$S(D) = e^{-(\alpha D + \beta D^2)}, \tag{2.4}$$

where α and β are the LQ-model fitting parameters describing the dose-response to the radiation. The first term in the exponential describes the initial slope of the cell survival curve, that is, the linear component caused by single track events. The second term describes the quadratic component caused by two-track events. The LQ-parameters are found by regression fitting to experimental data. The ratio (α/β) is widely used to describe the fractionation sensitivity of different tissue types and organs, as it is possible to extract the ratio from clinical endpoints, not only cell survival data [7]. The LQ model has been used extensively to analyse and predict response to ionizing radiation both *in vitro* and *in vivo* [19]. The phenomenological RBE models are typically based on the LQ model [20].

Figure 8 shows the survival curves given by the LQ model for an experiment executed with protons and a reference experiment done with photons. The difference in steepness between the two curves stems from an RBE above 1. The α and β parameters are different for cell survival experiments executed with protons compared to photons, thus giving different cell survival curves. The α and β data in Figure 8 originates from Belli et al [21].

By calculating the ratio of the dose level evaluated at specific survival fraction where both modalities are isoeffective, the RBE can be found. This is exemplified in Figure 8, calculating the RBE for two specific survival fractions.

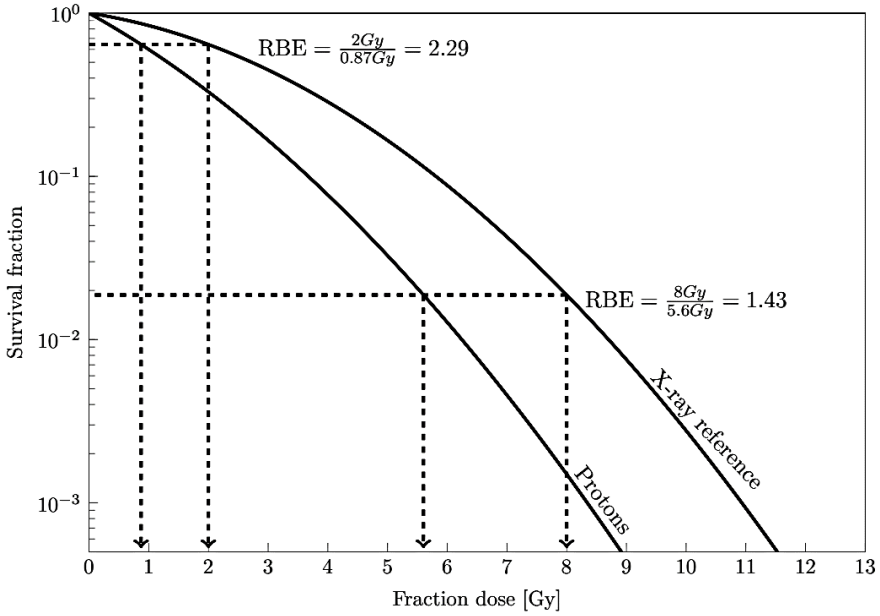

Figure 8: Schematic dose response curves of V79 hamster cells irradiated with monoenergetic protons and with x-rays as reference radiation [16].

2.1.7 Coupling the RBE with the LQ-model

Considering a specific endpoint where proton- and photon irradiation are isoeffective, the survival fraction for both modalities are mathematically equal to each other, and we can write:

$$S(D_p) = S(D_x), \tag{2.5}$$

where $S(D_p)$ and $S(D_x)$ are the survival fractions of proton and photon irradiations, respectively. We can describe this equation with the mathematical description of the LQ model of the proton and photon irradiation, as given by Equation (2.4):

$$e^{-(\alpha D_p + \beta D_p^2)} = e^{-(\alpha_x D_x + \beta_x D_x^2)}, \tag{2.6}$$

where D_p is the physical proton dose, α and β are the LQ parameters of the proton radiation, D_x is the physical photon dose and α_x and β_x are the LQ parameters of the photon dose. Solving this equation for D_x and inserting the result into the definition of RBE given in Equation (2.2) yields:

$$RBE\left(D_p, \alpha_x, \beta_x, \alpha, \beta\right) = \frac{1}{2D_p}\left(\sqrt{\left(\frac{\alpha_x}{\beta_x}\right)^2 + 4D_p\frac{\alpha_x}{\beta_x}\frac{\alpha}{\alpha_x} + 4D_p^2\frac{\beta}{\beta_x}} - \frac{\alpha_x}{\beta_x}\right). \quad (2.7)$$

With this equation, the proton RBE is now only a function of the proton dose and the LQ parameters of the proton radiation and the photon reference radiation. Generally, the RBE is highest at low doses and decreases with increasing dose, as seen in Figure 8. By evaluating Equation (2.7) at the upper and lower physical dose limits, we can obtain two independent functions describing the extreme RBE at low doses (RBE_{max}) and high doses (RBE_{min}):

$$RBE_{max} = \lim_{D_p \to 0} RBE = \frac{\alpha}{\alpha_x} \quad (2.8)$$

$$RBE_{min} = \lim_{D_p \to \infty} RBE = \sqrt{\frac{\beta}{\beta_x}} \quad (2.9)$$

With these two equations, we can now reformulate Equation (2.7) with respect to RBE_{max} and RBE_{min}:

$$\begin{aligned} &RBE\left(D_p, (\alpha/\beta)_x, RBE_{max}, RBE_{min}\right) \\ &= \frac{1}{2D_p}\left(\sqrt{\left(\frac{\alpha}{\beta}\right)_x^2 + 4D_p\left(\frac{\alpha}{\beta}\right)_x RBE_{max} + 4D_p^2 RBE_{min}^2} - \left(\frac{\alpha}{\beta}\right)_x\right), \end{aligned} \quad (2.10)$$

where $(\alpha/\beta)_x$, equivalent to α_x/β_x, is the treatment fractionation sensitivity of the reference radiation.

2.2 RBE modelling

2.2.1 RBE models in literature

As mentioned, the RBE models found in literature can be divided into three major groups. Phenomenological models try to describe the relationship of measurable empirical quantities and do not include any information or assumptions of cells on a subcellular level [7]. The majority of proposed RBE models belong to this group of models [22]. They rely on measurable input and output variables of cell irradiation experiments, typically the LET_d and the LQ parameters α and β of experiments. Appropriate assumptions are then made for RBE_{max} and RBE_{min} with free fitting parameters, which are determined by regression analysis to the data [7]. Most phenomenological RBE models employ a linear dependency of LET_d to calculate the biological dose. However, several experiments have indicated a possible non-linear trend [10].

Plan-based models are not directly based on cell experiments. As the name implies, they are instead based on information from treatment plans, like dose and LET_d distributions. For plan-based models it is assumed that he average RBE inside the target volume is 1.1, while the definition of RBE_{max} and RBE_{min} are linearly dependent of LET_d.

Mechanistic models aim to model the biological effects on a microscopic scale within the cells, instead of assuming and calculating relationships between experimental variables [7]. The microscopic dose distribution will give rise to biological local events within the nucleus, such as double strand breaks (DSB), which are estimated by the models. These events are quantified and used in the estimation of the overall cell survival [16].

2.2.2 RBE modelling from the LQ-model

Phenomenological models are based on empirical data from *in vitro* proton irradiation of various cell lines and utilize the LQ-model with cell inactivation as the biological endpoint. All LQ based RBE models have Equation (2.10) in common. They only differ by the definitions of the RBE_{max} and RBE_{min} functions, which for most models are found by regression to an experimental database [22]. The size and selection of the *in vitro* data and the regression technique vary between different models. The models also use different assumptions for the RBE_{max} and RBE_{min} functions, as they can be made dependent on physical quantities like the LET spectrum, the LET_d of the beam, or biological quantities such as the tissue specific $(\alpha/\beta)_x$ ratio of the photon radiation [22].

The RBE is known to vary with tissue type and cell line, and this dependency can be taken into consideration in the RBE models. To make a general RBE model which include a tissue specific parameter as input to RBE_{max} and RBE_{min}, the LQ parameters of the photon radiation is often used to quantify the tissue type.

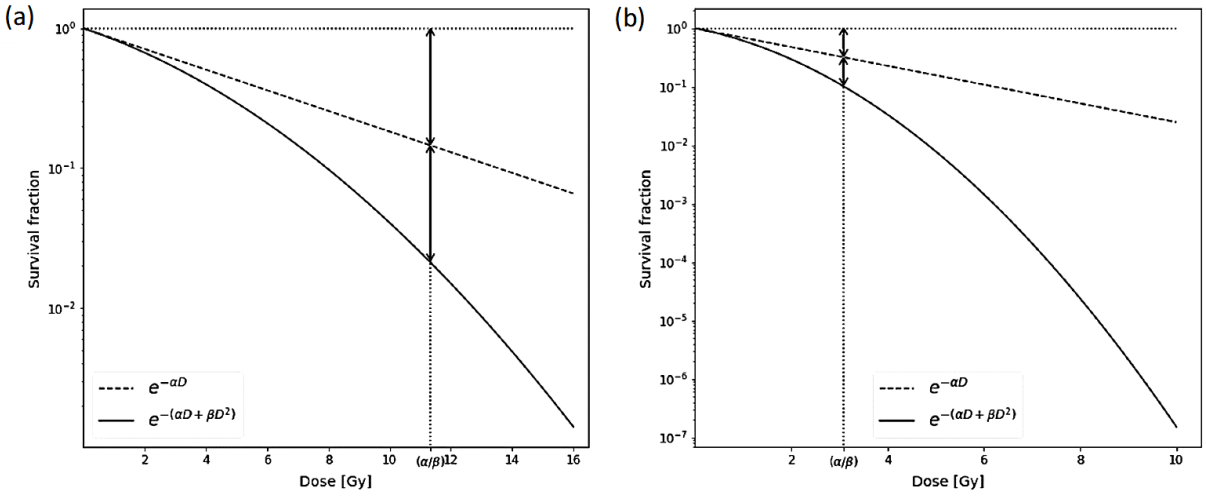

Figure 9: Cell survival curves of two different cell lines ((a) SCC25 cells from human squamous cell carcinoma of the tongue and (b) DLD1 cells from human colorectal tumor), irradiated with ^{60}Co x-rays. The two arrows in each plot show the dose at which the contribution to cell killing is equal for the linear and quadratic component. This dose defines the (α/β) ratio, shown at the x-

axis of both curves, with values (a) 11.333 Gy and (b) 3.083. The α and β data originates from (a) Bettega et al [23] and (b) Baggio et al [24].

The LQ model is first and foremost a practical mathematical model for radiation response and not necessarily directly related to specific mechanisms. However, one idea for the mechanistic justification of the LQ model is that the linear component ($\exp(-\alpha D)$) might result from single-track events, while the quadratic component ($\exp(-\beta D^2)$) might arise from two-track events . Thus, with this interpretation, the first term in the exponential in Equation (2.4) describes unrepairable lethal damage, while the second term describes the repairable non-lethal damage. The ratio between the LQ parameters (α/β) is commonly used to describe the fractionation sensitivity of different tissues and organs. The ratio is defined as the dose at which the linear contribution to damage equals the quadratic contribution. Figure 9 shows the (α/β) ratio for two different cell survival curves given by the LQ model. Early responding tissues express their damage within a period of days to weeks after irradiation. The (α/β) ratio of such tissue is in the range 7-20 Gy [25]. For late responding tissues, which express their damage months to years after irradiation, the (α/β) ratio generally ranges from 0.5 to 6 Gy [25]. The dose response of tumor cells are generally thought to be similar to that of early responding normal tissues, sometimes with an even higher (α/β) ratio. However, there are evidence that some human tumor types exhibit low (α/β) ratios, perhaps even lower than for late normal-tissue reactions [25].

After the development of an RBE model, the model is typically compared to previous published models. An example of such comparison of phenomenological models is showed in Figure 10. Here, the two top ranked models from the study by Rørvik et al in 2017 [10] (discussed in detail in 2.2.5) are compared to the three established models published by Carabe et al [26], Wedenberg et al [8] and McNamara et al [27] by varying the LET_d and keeping the dose and $(\alpha/\beta)_x$ constant at clinical relevant levels. Most of the published RBE models consider the RBE_{max} to be inversely proportional to $(\alpha/\beta)_x$, as seen in equations (2.11) and (2.15). All the models in Figure 10 are obtained using this assumption, as seen from the decreasing slopes with increasing $(\alpha/\beta)_x$

values. In the 2017 study by Rørvik et al [10], the reliability of this widely used assumption for the dependency on $(\alpha/\beta)_x$ is questioned, as it might be too generic. Although this tissue dependency has previously been verified and included in many phenomenological models, it should potentially be revised in the light of new data and new experience. The present study includes an investigation of the validity of this assumption when a linear relationship between RBE_{max} and LET is applied.

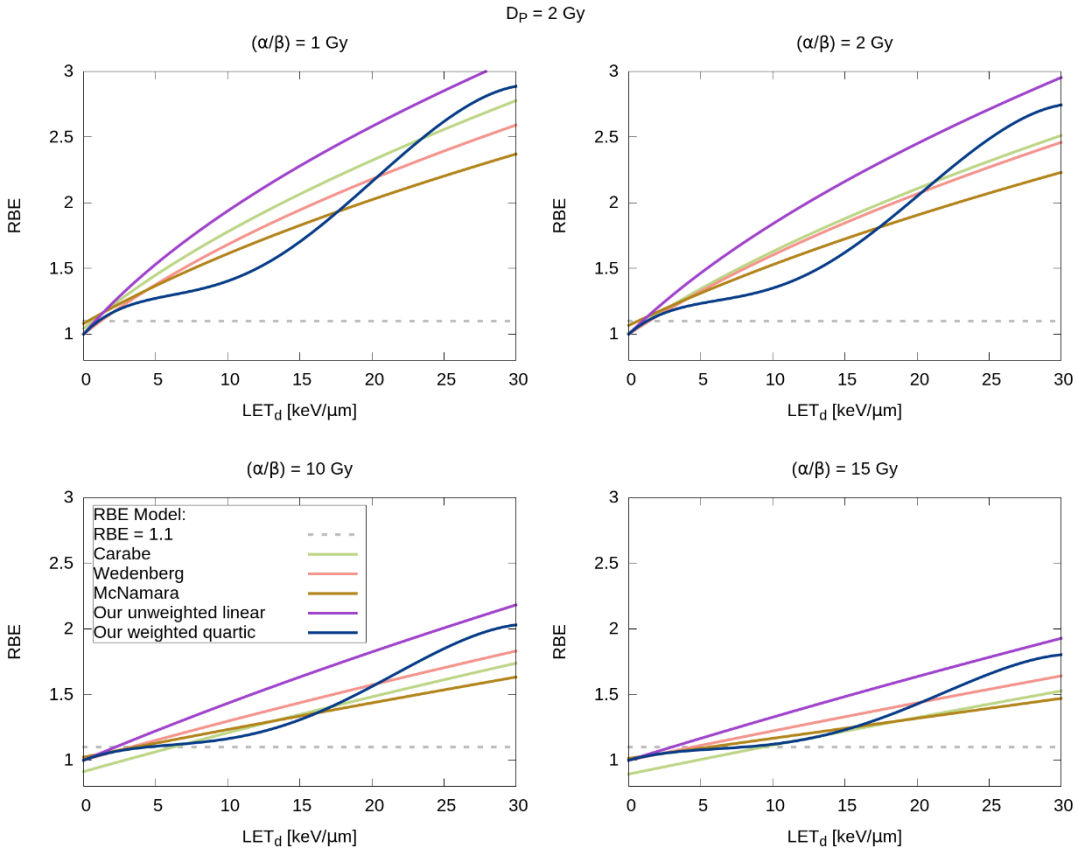

Figure 10: The RBE for different LET_d values for a monoenergetic beam with a physical proton dose of 2 Gy and 4 clinical relevant $(\alpha/\beta)_x$ values (Figure S4 in the appendix to Rørvik et al 2017 [10]).

2.2.3 Cell irradiation experiments for proton and ion therapy

As noted above, the phenomenological RBE models rely on an experimental *in vitro* database. The data is obtained from cell irradiation experiments. The method of such experiments is to expose biological samples of specific cell lines to proton radiation with specific combinations of dose and LET. The same cell line are also irradiated with

photons to obtain reference data for the proton RBE determinations. Introduction of material into the beam line enables mapping of the biological effect at different points on the Bragg curve.

The results from cell irradiation experiments have been compromised by numerous complicating factors such as adequate access to beam time, non-standardized irradiations, variations in experimental techniques and reporting, as well as a large span of reported biological responses [9]. The methods for mapping spatial variations in biological effectiveness have been time-consuming and often yielded inconsistent results with large uncertainties [9]. Thus, the data needed to develop accurate RBE models have been limited and difficult to obtain.

In Paganetti's review from 2014 [4], he systematically analysed hundreds of published clonogenic data points [4]. The review highlights the large spread of the existing clonogenic data and suggests a need for experimental protocol standardization and more complete reporting of fit parameters and errors. This was also addressed by Guan et al in 2015 [9], with a report describing their procedure to spatially map the biologic effectiveness of scanned proton beams with high accuracy and throughput while minimizing biological uncertainties. Their design is summarized below, as an example of a possible procedure for cell irradiation experiments:

Monte Carlo simulations were used to design an apparatus (jig) to attenuate proton energy in a stepwise fashion from the incident energy to the end of the range, as schematically illustrated in Figure 11a. The basis for this setup is that the LET increases as a function of depth along the Bragg curve. 96-well plates were used to allow the simultaneous irradiation of biologic samples to multiple dose-LET combinations. The wells were grouped into 12 columns, and each of them were simultaneously exposed to a different combination of dose and LET, such that all 8 wells in a column were intended to receive the same dose-LET combination. The method enables the acquisition of 12 times the amount of data from a single exposure. The design can be considered a multi-step range shifter consisting of 12 steps. The thickness of steps

varies from zero to a maximum value, producing protons of increasing LET. The jig was custom-fabricated from Lucite and directly mounts into the snout of the scanning beam gantry. The biological sample plate is placed on top of the jig with the beam directed upwards, as seen in Figure 11b. The thickness of the material interposed in the proton beam path varies in increasingly finer increments towards the end of the proton range. This was done to increase the resolution in regions of high dose and LET gradients.

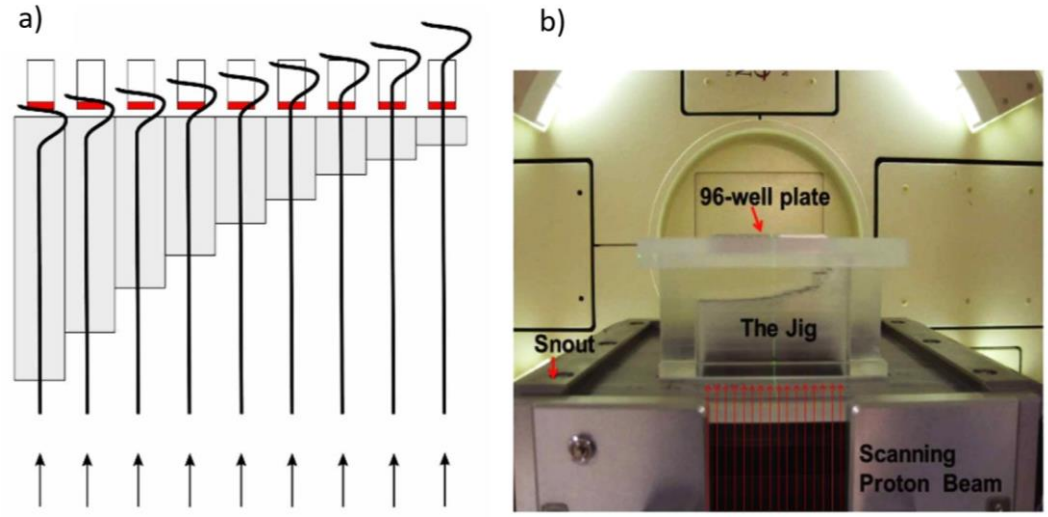

Figure 11: Design of the irradiation device used by Guan et al [9]. (a) Schematic diagram of the jig concept illustrating the strategy for the simultaneous irradiation of biological samples in the 96-well plate with protons at different depths of the Bragg curve. The grey bars indicate the Lucite, and the culture medium is indicated by red. The jig is designed so that the steps matches the columns of the 96-well plate, varying the position along the Bragg curve. The illustration shows only 9 columns and the step dimensions are not to scale. (b) The jig directly mounted onto the scanning beam gantry. The 96-well plates are inserted into a precisely milled slot in the jig holder designed to minimize positioning errors. Protons are incident from below.

2.2.4 The McNamara model

In 2015, a phenomenological LQ-based RBE model was proposed by McNamara et al [27]. The model was derived using the database from Paganetti's review in 2014 [4], which was the most comprehensive collection of proton RBE experimental data at the time. The database consists of 369 data points from 76 different studies. In order to develop a biophysical model relevant to clinical proton therapy, restrictions were

applied to the primary dataset. Only data points with $LET_d < 20\ keV/\mu m$ and $(\alpha/\beta)_x < 30\ Gy$ were used un the fit, excluding 84 data points from the original dataset. The LQ model is assumed valid for proton doses ranging from 1-10 Gy, so one data point only valid for doses between 8-24 Gy was excluded from the dataset. Additionally, two other data points were excluded. One of them was considered an outlier, the other was reported in the same study, and was removed for the sake of consistency. RBE_{max} were assumed to have a linear relationship with respect to LET_d as well as a dependence on $(\alpha/\beta)_x$, while RBE_{min} was assumed to have a dependence on $\sqrt{(\alpha/\beta)_x}$,

$$RBE_{max}[(\alpha/\beta)_x, LET_d] = p_0 + \frac{p_1}{(\alpha/\beta)_x} LET_d \tag{2.11}$$

$$RBE_{min}[(\alpha/\beta)_x, LET_d] = p_2 + p_3 \sqrt{\left(\frac{\alpha}{\beta}\right)_x} LET_d, \tag{2.12}$$

where p_{0-3} are the fit parameters for the model. Both assumptions are in accordance with the LQ model. The Matlab *NonLinearModel.fit* algorithm was used to estimate the fit parameters using an iterative procedure. The RBE uncertainties calculated from Paganetti [4] were used to weight the data in the fitting procedure. To avoid the fitted model from extending to the complex plane, three additional data points with $(\alpha/\beta)_x \sim 0$ were removed from the primary dataset.

The values of the fit coefficients that best fit the experimental data were found to be: $p_0 = 0.99064$ (standard error (SE) 0.014125), $p_1 = 0.35605$ (SE 0.015038), $p_2 = 1.1012$ (SE 0.0059972) and $p_3 = -0.0038703$ (SE 0.00091303). Thus, by putting the fit coefficients in Equations (2.11) and (2.12) and inserting the results into Equation (2.10), the RBE were found to be given as:

$$RBE\left[D_p, \left(\frac{\alpha}{\beta}\right)_x, LET_d\right]$$
$$= \frac{1}{2D_p}\left(\sqrt{\left(\frac{\alpha}{\beta}\right)_x^2 + 4D_p\left(\frac{\alpha}{\beta}\right)_x\left(0.99064 + \frac{0.35605}{(\alpha/\beta)_x}LET_d\right) + 4D_p^2\left(1.1012 - 0.0038703\sqrt{(\alpha/\beta)_x}LET_d\right)^2}\right. \quad (2.13)$$
$$\left. - \left(\frac{\alpha}{\beta}\right)_x\right)$$

The fitted model was thoroughly compared to two previously published LQ based RBE models published by Carabe et al in 2012 [26] and Wedenberg et al in 2013 [8], which have different assumptions regarding the relationship between LET, β, α and $(\alpha/\beta)_x$. The Carabe model (CAR model) applies a linear relationship between RBE_{min}, RBE_{max} and LET_d with a slope depending on $(\alpha/\beta)_x$. The Wedenberg model (WED model) is based on the assumption that there is a linear relationship between α and LET_d with a slope depending on $(\alpha/\beta)_x$, while β is assumed to be independent of LET_d. The predictability of the model was also tested against new experimental RBE data published by Guan et al in 2015 [9], after Paganetti's 2014 review.

The model comparison and predictability testing using the Guan data are both explained in detail in the review concerning the development of the McNamara model (MCN model) [27]. While the latter is not considered here, some key results from the model comparison are summarized below.

Predictions of RBE as a function of LET_d for different values of $(\alpha/\beta)_x$ and a dose of 2 Gy showed that for all three models, the RBE increases with increasing LET_d with a somewhat linear relationship between RBE and LET_d (especially at large $(\alpha/\beta)_x$) and a steeper slope occurring at low $(\alpha/\beta)_x$. At $(\alpha/\beta)_x \lesssim 5\ Gy$, the MCN model predicts RBE values lower than the other two models for $LET_d \gtrsim 5\ keV/\mu m$, while it predicts slightly higher RBE values at $LET_d \lesssim 5\ keV/\mu m$. For $(\alpha/\beta)_x \geq 10\ Gy$, the MCN model agrees better with the WED model for low LET_d values, while at high LET_d it is similar to the CAR model.

The three models were used to predict the RBE for cell survival as a function of $(\alpha/\beta)_x$ for four different LET_d values and a dose of 2 Gy, showing that the RBE decreases with increasing $(\alpha/\beta)_x$ of all three models. The decrease in RBE was most significant at low $(\alpha/\beta)_x$ values, especially at large LET_d values, where the model curves are steeper. The MCN model had closer agreement with the WED model, especially at $LET_d \leq 5\ keV/\mu m$. Compared to the two other models, the CAR model predicts higher RBE values at $(\alpha/\beta)_x < 2\ Gy$ and much lower RBE values for $(\alpha/\beta)_x \gtrsim 5\ Gy$.

Regarding RBE for cell survival as a function of dose for different $(\alpha/\beta)_x$ values and an LET_d of 2.5 $keV/\mu m$, predictions showed a decreasing RBE with increasing dose for $(\alpha/\beta)_x < 2\ Gy$ for all models. According to the CAR model, RBE increases with increasing dose for large $(\alpha/\beta)_x$, while the WED model predicts a decrease in RBE with increasing dose for all $(\alpha/\beta)_x$ values considered. The slope of the MCN model converges to zero for high $(\alpha/\beta)_x$ values. Overall, the MCN model showed better agreement with the CAR model at low $(\alpha/\beta)_x$.

The CAR and WED models are based on fits through a small subset of the data used in the fitting of the MCN model. It was also pointed out that all previously published phenomenological models were based on limited experimental datasets and they may have to be revised [27].

2.2.5 The Rørvik models

In 2017, Rørvik et al [10] investigated if biological dose models including non-linear LET dependencies should be considered, in contrast to the linear dependency of LET_d employed by most phenomenological models in the calculation of the biological dose. To do this, an LET spectrum based dose model was introduced.

The RBE-LET relationship was investigated by fitting of polynomials from 1st to 5th degree to a database of 85 data points from aerobic in vitro experiments. The database was obtained by applying a list of criteria to the database from Paganetti [4] and in a

literature search of more recent articles. Experiments with modulated SOBP or laser accelerated protons were excluded, along with extreme LET values above $40\ keV/\mu m$ and experiments including hypoxic cells. Also excluded from the database were experiments with cells having $(\alpha/\beta)_x$ above $25\ Gy$ and studies which did not report the uncertainties of α, α_x and β_x. In addition to this, two experiments with irradiation doses exclusively above $8\ Gy$ were excluded because such high dose levels are not clinically relevant.

The polynomials were fitted to the database with both unweighted and weighted regression, with the latter taking the experimental uncertainties into account. Statistical testing was performed to decide whether higher degree polynomials provided better fits to the data compared to lower degrees. A tissue dependent phenomenological biological dose model for proton therapy with the full LET spectrum as a parameter for the radiation quality was proposed. The model was made tissue and dose dependent by Equation (2.10). The dependency on the full dose weighted LET spectrum were included by weighting the spectrum with $r_{max}(L)$, which is a biological weighting function (BWF) based on data from in vitro cell experiments. The BWF formalism for the LET spectrum is then similar to the formalism used in microdosimetric RBE models [28]:

$$RBE_{max}\big(\boldsymbol{d}(L)\big) = \int_0^\infty r_{max}(L)\boldsymbol{d}(L)dL \tag{2.14}$$

The LET spectrum of a monoenergetic proton beam can be expressed as a Dirac delta function centered around the LET_d of the beam. This property was used to find a quantitative description of $r_{max}(L)$, by making a regression fit to a monoenergetic database with discrete LET_d and corresponding $RBE_{max}(\boldsymbol{d}(L))$ values. The investigation of the RBE-LET relationship through the testing of the fitted polynomials aimed to determine the model function of RBE_{max}, while RBE_{min} was assumed to be independent of LET_d and $(\alpha/\beta)_x$, i.e. $RBE_{min} = 1$.

The result of the statistical testing showed that the linear fit was the most elaborate BWF for the unweighted dataset. By using a continuous and linear increasing fit for also the high LET region, the RBE_{max} of the model can be formulated as:

$$RBE_{max}(LET_d) = 1 + \frac{0.645}{(\alpha/\beta)_x} LET_d \tag{2.15}$$

For the weighted dataset, the quartic model was shown to be the best fitting polynomial to the data. The model thus needs to be based on the full LET spectrum, and the RBE_{max} of the model was formulated as:

$$RBE_{max}(\boldsymbol{d}(L)) = \int_0^{\infty} r_{max}(L)\boldsymbol{d}(L)dL, \tag{2.16}$$

where $\boldsymbol{d}(L)$ is the full LET spectrum and $r_{max}(L)$ is the BWF, defined as:

$$r_{max}(L) = \begin{cases} 1 + \frac{Gy}{(\alpha/\beta)_x}\left(0.587\left(\frac{keV}{\mu m}\right)^{-1} L - 0.0808\left(\frac{keV}{\mu m}\right)^{-2} L^2 + 0.00564\left(\frac{keV}{\mu m}\right)^{-3} L^3 - 9.92\cdot 10^{-5}\left(\frac{keV}{\mu m}\right)^{-4} L^4\right), L < 37.0 keV/\mu m. \\ 1 + \frac{10.5Gy}{(\alpha/\beta)_x}, L \geq 37.0 keV/\mu m. \end{cases} \tag{2.17}$$

These two models were compared to three previously published models by successively varying the dose, LET_d and $(\alpha/\beta)_x$ while the other variables were kept constant at clinical relevant levels.

2.2.6 Other RBE models

In 2018, Rørvik et al [22] published a study in which eleven published phenomenological RBE models and two plan-based models were explored. The paper provided a list of the dependencies of RBE_{max} and RBE_{min} that was applied to develop the models, as well as the number of data points and the cell lines that was used to derive them. The resulting overview of the considered models in the study by Rørvik et al [22] are summarized in Table 1.

Table 1: All phenomenological RBE models that was explored by Rørvik [22] listed in chronological order according to date of publication. The two plan-based models are given at the bottom of the table. The dose-weighted energy spectrum and the dose-weighted LET spectrum are denoted by d(E) and d(L), respectively. (Table 1 from Rørvik et al 2018 [22].)

Model reference	**Model abbreviation**	**Cell-lines**	**Number of data points**	RBE_{max} **dependencies**	RBE_{min} **dependencies**
Belli et al 1997 [29]	BEL	V-79	6	d(E)	d(E)
Wilkens and Oelfke 2004 [30]	WIL	V-79	19	LET_d	1
Tilly et al 2005 [31]	TIL (TIL2/TIL10)	V-79 / Multiple	7 / 4	LET_d, $(\alpha/\beta)_x$	1
Chen and Ahmad 2012 [32]	CHE	V-79	14	LET_d	1
Carabe et al 2012 [26]	CAR	V-79	44	LET_d, $(\alpha/\beta)_x$	LET_d, $(\alpha/\beta)_x$
Wedenberg et al 2013 [8]	WED	Multiple	19 (24)	LET_d, $(\alpha/\beta)_x$	1
Jones 2015b [33]	JON	Multiple	28 (200)	LET_d, α_x	LET_d, β_x
McNamara et al 2015 [27]	MCN	Multiple	285	LET_d, $(\alpha/\beta)_x$	LET_d, $(\alpha/\beta)_x$
Mairani et al 2017 [34]	MAI	Multiple	25 (31)	LET_d, $(\alpha/\beta)_x$	1
Rørvik et al 2017 [10]	RØR (RØRU / RØRW)	Multiple	85	LET_d, $(\alpha/\beta)_x$ / d(L), $(\alpha/\beta)_x$	1
Peeler 2016 [35]	PLR	Multiple	48	LET_d, $(\alpha/\beta)_x$	LET_d, $(\alpha/\beta)_x$
Frese et al 2011 [36]	FRE	Plan-based	0	LET_d, α_x	1
Unkelbach et al 2016 [37]	UNK	Plan-based	0	LET_d	LET_d

The RBE models are developed with experimental databases containing various number of datapoints from 6-285, and the range of $(\alpha/\beta)_x$ and LET_d values used to derive the models differs considerably [22]. Except from the BEL, WIL, CHE and UNK models, all the phenomenological and plan-based models are made tissue dependent through the reference radiation parameters of the LQ model. This dependency is based on the $(\alpha/\beta)_x$ ratio, with the exception of the JON model, which uses α_x and β_x separately.

The WIL, TIL2, CHE and CAR models are based on proton data from a single cell line, yielding a narrow range of $(\alpha/\beta)_x$ values. The data used in the TIL10 and PLR models and the proton data used in the JON model were obtained with two different cell lines, while the remaining models used data from 5 to 33 cell lines, yielding greater ranges of $(\alpha/\beta)_x$ values. Although the MCN and RØR models include a large number of different cell lines, the databases are still dominated by V-79 cells and other Chinese hamster cells, which have low $(\alpha/\beta)_x$ values.

As seen in Table 1, all the phenomenological models except from BEL and RØRW use LET_d as a parameter for the radiation quality of the beam. The TIL, CAR, MCN and PLR models used data only from LET_d values below $20\ keV/\mu m$. This is in line with the Paganetti 2014 article [4], which states that LET_d values above $20\ keV/\mu m$ are rarely present in clinical proton therapy, even for the most extreme cases. While the WIL, CAR, MCN and PLR databases are dominated by data points with $LET_d < 5\ keV/\mu m$, the BEL, TIL, CHE, WED and MAI databases first start at $7.7\ keV/\mu m$, showing notable differences between the databases in the low LET_d data.

Regarding the model assumptions used in the regression analysis of RBE_{max} and RBE_{min}, all models included a variable RBE_{max} while BEL, CAR, JON, MCN, PLR and UNK were the only ones assuming a variable RBE_{min}. Except from JON, all these models have an effective model function for RBE_{min} that is almost 1 for all $(\alpha/\beta)_x$ and LET_d values. As seen in Table 1, the remaining models assume that $RBE_{min} = 1$. Apart from BEL, CHE, RØRW and PLR, all models are linearly dependent on LET_d. The two last mentioned models assume a non-linear dependency mainly due to the inclusion of the cell survival data from Guan et al 2015 [9], which indicated a non-linear LET dependency. All the models that are made tissue dependent on the LQ parameters of the reference radiation assume that RBE_{max} increases with decreasing $(\alpha/\beta)_x$ value, while they deviate on the tissue dependency of RBE_{min}.

Fitting of the models to the databases were carried out using unweighted regression techniques for the majority of the models. Weighted regression based on the uncertainty of the data points is only applied by the MCN and RØRW models.

In 2016, Mairani et al [38] published a phenomenological LQ based RBE model for helium ions in a clinically relevant range of doses. The database used consisted of *in vitro* data points from experiments performed with asynchronous cells, with around 20 cell lines represented. The coupling of the helium RBE with the LQ model were done as in Equations (2.5)-(2.7), and two ratios were defined:

$$RBE_\alpha \equiv \frac{\alpha_{He}}{\alpha_x} \quad , \quad R_\beta \equiv \frac{\beta_{He}}{\beta_x} \tag{2.18}$$

yielding the following helium RBE dependency:

$$RBE\left[(\alpha/\beta)_x, D_{He}, RBE_\alpha, R_\beta\right] = -\frac{1}{2D_{He}}\left(\frac{\alpha}{\beta}\right)_x + \frac{1}{D_{He}}\sqrt{\frac{1}{4}\left(\frac{\alpha}{\beta}\right)_x^2 + RBE_\alpha\left(\frac{\alpha}{\beta}\right)_x D_{He} + R_\beta D_{He}^2} \tag{2.19}$$

Comprehensive data analysis and regression procedures were carried out for the parametrization of RBE_α and R_β. It was assumed that α_{He} approaches α_x for decreasing LET values, and that the initial slope was affected by the cell line via the inverse relationship with $(\alpha/\beta)_x$, so that the RBE_α of cell lines with high $(\alpha/\beta)_x$was less dependent on LET (L), namely:

$$RBE_\alpha(L|(\alpha/\beta)_x) = 1 + \left[k_0 + \left(\frac{\beta}{\alpha}\right)_x\right] \cdot f(L). \tag{2.20}$$

Several fitting functions were used, with $f(L)$ of increasing complexity and number of parameters, to improve the description of the experimental data until no relevant improvements could be found. The procedure used to link experimental RBE_α data with models were based on the Jackknife (JK) resampling technique, which is considered to be a robust procedure for the estimation of statistical population

parameters. The dataset was highly inhomogeneous, and α, β and LET were mixed up either having experimental uncertainties or not. To deal with this, the JK technique was coupled to a data resampling technique which took into account the data uncertainties, if available.

The R_β ratio showed a large degree of dispersion when plotted against LET, as seen in Figure 12. To estimate a trend component in the R_β-$LET(L)$ relation, a filter was applied to the data to remove fluctuating components, yielding an estimate of the slow-moving trend of the data. A Gaussian filter was adopted for the computation of a weighted running average of R_β ratios. The smoothed R_β curve was then fitted minimizing the unweighted sum of squared differences computed between the calculated running mean values and the following parametrization:

$$R_\beta(L) = b_0 \exp\left[-\left(\frac{L - b_1}{b_2}\right)^2\right]. \tag{2.21}$$

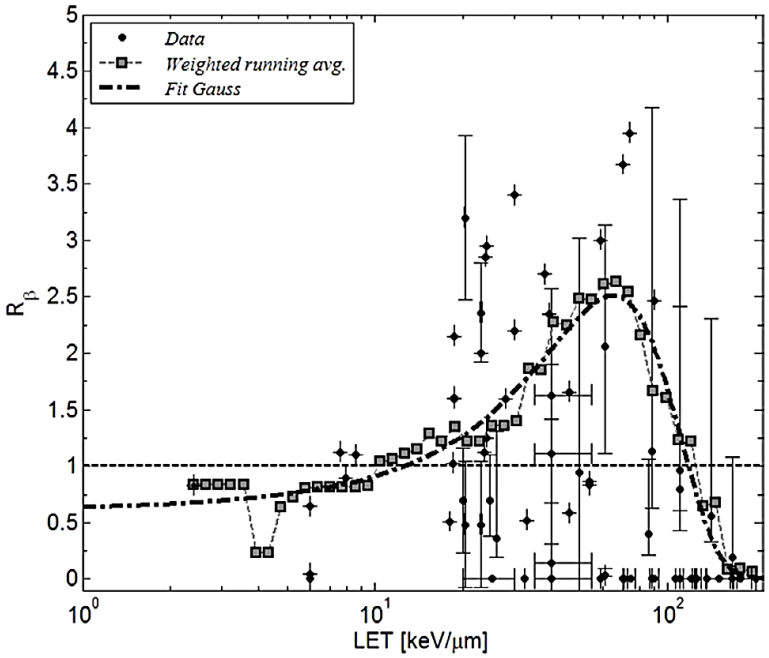

Figure 12: Comparison between experimental data (points with error bars), weighted running average values (squared connected by a dashed line) and Equation (2.21) (dashed-dotted line) of R_β as a function of LET [38].

Abolfath et al [39] introduced an approach for global fitting of the high-throughput and high accuracy clonogenic cell-survival data published by Guan et al [9]. Their phenomenological model was based on an extension of the microdosimetric kinetic model, originally proposed by Hawkins [40, 41]. Instead of individually fitting cell survival curves with measured data for each LET value separately, Abolfath et al [39] carried out a global fit to the measured data, aiming to reduce the overall uncertainty. The fitting procedure may be more reliable, as it takes into account a correlation among the survival curves. The method is based on a three-parameter global fitting. An optimization procedure was developed, allowing fitting of a 2D surface in a 3D parameter space spanned by dose, LET and cell surviving fraction (SF), using two independent variables D (macroscopic dose) and L (LET as a linear function of the dose-averaged lineal energy). Specific forms of polynomials that may be used to fit the data is provided by the following equation:

$$-\ln(SF) = \sum_{i=1}^{n} \sum_{j=1}^{i} b_{i,j} L^{i-j} D^{j}. \qquad (2.22)$$

The result of the global fitting using linear polynomials is shown in Figure 13. In this procedure, the LQ parameters are considered $\alpha = \alpha_0 + \alpha_1 LET$ and $\beta = \beta_0$ to obtain an optimal surface in the 3D space that fit the experimental points.

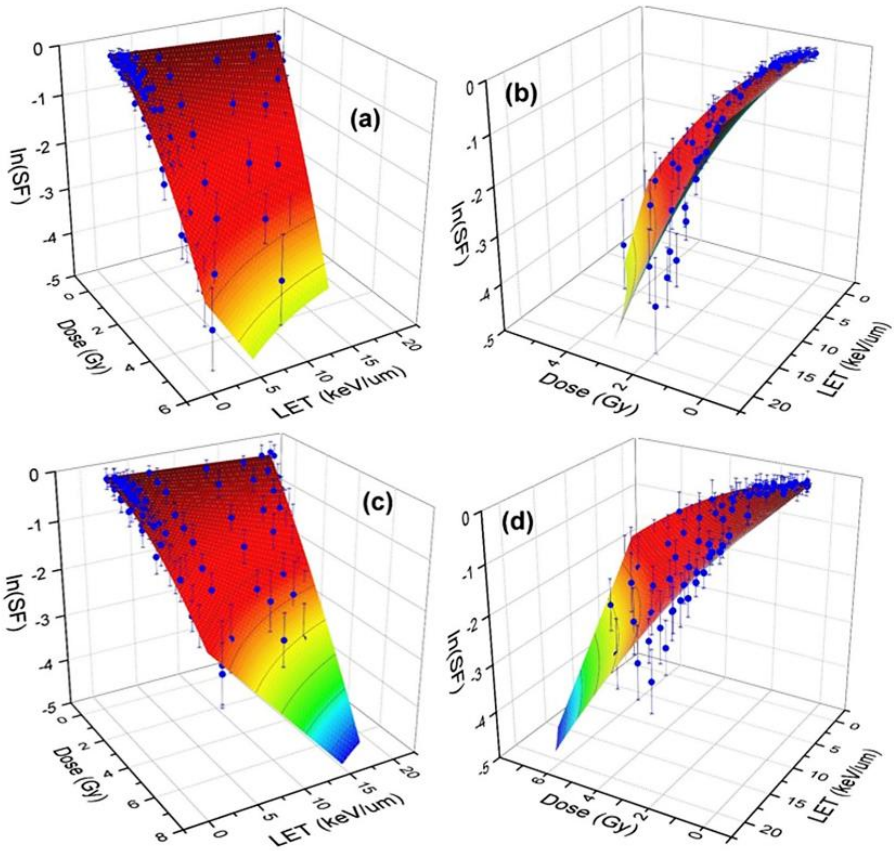

Figure 13: Experimental data of clonogenic cell survival fraction (SF) corresponding to H460 (a,b) and H1437 (c,d) cells are shown in blue dots with experimental error bars in logarithmic scale. A linear LET model corresponding to $\alpha = \alpha_0 + \alpha_1 L$ and $\beta = \beta_0$ was used to obtain a globally optimal surface fitted to the experimental data. The left and right figures show the views from low and high LETs [39].

3. Methods

3.1 Experimental data

3.1.1 Collecting the data

The database was mainly collected from two existing databases. Both of them are collections of experimental data from published literature. The first one can be found in the appendix (table A1-A4) in Paganetti's review from 2014 [4]. This database contains clonogenic data from proton experiments only. If the data was incompletely reported in the original articles, the missing values were found through suitable fits or simulations, as explained in detail in the review [4]. The second database used is the Particle Irradiation Data Ensemble (PIDE3.1) provided by GSI [42]. The datapoints in this database are mainly from irradiation experiments with protons, but it also contains data from experiments using other ions, in particular carbon ions. For most of the datapoints in the PIDE database, additional values are given for all four LQ-parameters α_x, β_x, α and β, which reflects fits to the raw data. When collecting the database, if the fit-values were provided, they were used instead of the values from the original publication with the exception of re-fits giving $\alpha_x = 0$. Some datapoints were given in both the PIDE- and the Paganetti database, and these datapoints were only included once in the database.

In addition to the datapoints from the Paganetti and the PIDE databases, 24 datapoints from the database used by Odin Alvestad in his master-thesis were included [43]. 24 datapoints from a more recent article by Mara et al [44], that was published after the PIDE and Paganetti databases were also included in the database.

3.1.2 Data selection and filtering

It was necessary to know the irradiation condition used to obtain the data. That is, knowing if a data point is obtained with monoenergetic irradiation or by using an SOBP. This information was required to perform separate analysis using only

monoenergetic data and only SOBP data in order to investigate how RBE is affected by the radiation modality used. The PIDE database provides information about the irradiation condition for each data point, while no such information is given in the Paganetti database. Thus, for the Paganetti data this information had to be found in the publications that the data points originated from. If the publication was not accessible, or it was not clear from the article which irradiation condition was used to obtain the data, the data point were not included in the database.

Some of the data points had negative values for one or more of the LQ-parameters α_x, β_x, α, and β. These datapoints were excluded from the database, as they were seen as less relevant and the inclusion of such data was considered an unnecessary overcomplication of the analysis.

One of the main tasks in the project was to model the RBE based on fitting of RBE_{max} (α/α_x) values and $(\alpha/\beta)_x$ ratios calculated from the data. For datapoints with $\alpha_x = 0$, the RBE_{max} cannot be calculated (α/α_x goes to infinity). By the same means, if $\beta_x = 0$, the $(\alpha/\beta)_x$ ratio cannot be calculated. Consequently, all data points with one of these features were excluded from the database.

For 35 of the datapoints, the reference radiation were uncertain or not given at all. 18 of these were from Belli et al 2008 [45], and had both ^{60}Co and ^{137}Cs given as reference radiation. It was decided to use ^{60}Co as reference radiation for these datapoints. The remaining 17 datapoints had no information about the reference radiation. This included 2 datapoints from Wulf et al 1985 [46], 2 datapoints from Okayasu et al 2006 [47], 7 data points from Cox et al 1977 [48] and 6 data points from Kraft et al 1985 [49]. For the former two data points, no information about the reference radiation were found in the article, and they were excluded from the database. For the remaining 15 data points, the reference radiations were listed as x-rays in the articles, without any further specification. It was assumed that 200kVp x-rays are representative as radiation quality, and this was used as the reference radiation for these 15 datapoints.

Two datapoints from proton irradiation experiments were removed due to extreme LET-values, associated with large uncertainties. Both of them originated from Belli et al 1989 [50]. For proton radiation, the theoretical maximum LET_d value of $83 keV/\mu m$. One of the data points had $LET_d = 88.8 keV/\mu m$. This datapoint was thereby excluded from the database. The other data point had an LET_d value of $63.7 keV/\mu m$. As no other datapoints had LET values in this range, it was considered an outlier and excluded.

Three of the proton data points originating form Schuff et al 2002 [51] had very high values for RBE_{max}, compared to the rest of the proton data. They had RBE_{max} values of 35.0, 80.0 and 167.5, while the highest value for the remaining data points were 16.5. The data points were considered as outliers and excluded from the database, as they clearly stand out from the others with their deviating values for RBE_{max}.

3.1.3 Normalizing the LET_d

The reported photon reference radiation varies between different experiments. The photons used have different energies, thus different LET values. Consequently, the reference radiations have an RBE_x value relative to each other. This has to be considered when RBE values are calculated using Equation (2.10) because the α_x and β_x refer to the reference radiation used. For a database consisting of data from multiple experiments with multiple endpoints, the RBE_x for the considered endpoints is typically not known. According to Paganetti et al 2014 [4], it is therefore not feasible to correct the RBE values deduced from these experiments based on an average RBE_x of the reference radiation. Instead, the LET_d values can be normalized by calculating the relative LET_d^* [4].

The database is collected from many different experiments, thus reporting a number of different photon reference radiations, 20 in total: ^{137}Cs and ^{60}Co γ-rays, $50 kVp$, $100 kVp$, $120 kVp$, $130 kVp$, $145 kVp$, $150 kVp$, $180 kVp$, $200 kVp$, $210 kVp$, $220 kVp$, $225 kVp$, $240 kVp$, $250 kVp$ and $300 kVp$ x-rays and $4 MV$, $6 MV$, $10 MV$ and $15 MV$

photons. The reported LET_d values were normalized to the LET_d of Cobalt-60 using the following equation [10]:

$$LET^* = LET_p - LET_x + LET_{^{60}Co} \quad (3.1)$$

where LET^* is the normalized LET_d value and LET_p, LET_x and $LET_{^{60}Co}$ are the LET_d values of the proton beam, the photon reference radiation and ^{60}Co, respectively.

The LET_x values were found in the previous published articles by Howard et al 2017 [52] and Mairani et al 2016 [38]. For the reference radiations in the kVp range that were not considered in any of these articles, linear interpolation was used to obtain the LET_x value. Only one LET_x value were given for reference radiation in the MV range, that is the value for $6MV$ photons from Howard et al 2017 [52]. Thus, there were not enough information available to use interpolation here, and the given value for $6MV$ was used for all the reference radiations in the MV range. The LET_x values used for normalization are given in Table 2.

Table 2: LET_x values used to normalize the LET_d values of the database. Interpolation was used to estimate the values that were not listed.

Reference radiation	$LET_d[keV/\mu m]$	Reference
^{137}Cs γ-rays	0.8	Howard et al 2017 [52]
$6MV$ photons	0.2	Howard et al 2017 [52]
^{60}Co γ-rays	0.4	Mairani et al 2016 [38]
$80kVp$	1.549	Mairani et al 2016 [38]
$100kVp$	1.443	Mairani et al 2016 [38]
$200kVp$	1.164	Mairani et al 2016 [38]
$220kVp$	1.127	Mairani et al 2016 [38]
$240kVp$	1.092	Mairani et al 2016 [38]
$250kVp$	1.075	Mairani et al 2016 [38]

$50kVp$	1.708	From interpolation
$120kVp$	1.387	From interpolation
$130kVp$	1.359	From interpolation
$145kVp$	1.317	From interpolation
$150kVp$	1.304	From interpolation
$180kVp$	1.220	From interpolation
$210kVp$	1.145	From interpolation
$225kVp$	1.118	From interpolation
$300kVp$	0.99	From interpolation
$4MV$	0.2	Howard et al 2017 [52]
$10MV$	0.2	Howard et al 2017 [52]
$15MV$	0.2	Howard et al 2017 [52]

For three of the data points from proton experiments, the calculation of LET^* resulted in negative values. This was the case for two data points originating from Matsuura et al 2010 [53] and one from Williams et al 1978 [54]. These data points were considered as outliers due to the extreme LET values and were therefore excluded from the database.

3.2 RBE modelling

The methodology described in this section aims to pursue the objectives of the present work that were outlined in section 1.2. The proton database was analysed in the framework of a phenomenological RBE model to explore how data selection, regression methods and model assumptions affect the RBE estimation.

3.2.1 Database

The ranges and distributions of RBE_{max}, LET^* and $(\alpha/\beta)_x$ values in the proton database were visualized by histograms, with separate plots for monoenergetic data and SOBP data. The proton LET^* were also plotted against the $(\alpha/\beta)_x$ values for further investigation of the relationship between the distributions of these two variables.

3.2.2 Linear regression

Fitting of proton RBE data without database restrictions

The curve_fit() function provided in the SciPy library in Python was used to obtain fits for RBE_{max} as a function of LET^* and $(\alpha/\beta)_x$ for the proton data, applying a linear relationship between RBE_{max} and LET^*. The fitting method, which uses non-linear least squares to fit a function to data, was successively applied to all the proton data, the monoenergetic proton data and the SOBP proton data. In accordance with the MCN model (Equations (2.11), (2.13)) and the Rørvik model obtained with an unweighted dataset (RORU model, Equation (2.15)), the intercept at $LET^* = 0\ keV/\mu m$ was assumed to be equal to 1 for all three fits, and the RBE_{max} was assumed to be inversely proportional with $(\alpha/\beta)_x$. The fitting functions were thus assumed to have the form:

$$RBE_{max} = 1 + \frac{k}{(\alpha/\beta)_x} LET^*, \tag{3.2}$$

where k is the fit parameter. In addition to the fit parameters, curve_fit() returns the variances of the parameter estimates. These were used to determine the standard deviation errors of the obtained k values, as explained in the documentation for curve_fit() in the SciPy Reference Guide.

Analysis of RBE dependence on $(\alpha/\beta)_x$

As shown in Equation (3.2), the RBE_{max} is often assumed to be inversely proportional with $(\alpha/\beta)_x$. To investigate the validity of this assumption for our proton data, the data points were divided into seven subsets based on their $(\alpha/\beta)_x$ values. Each subset

contained data points with $(\alpha/\beta)_x$ values within one of the following intervals (in units Gy): $[0,3)$, $[3,6)$, $[6,9)$, $[9,12)$, $[12,15)$, $[15,20)$ and $(\alpha/\beta)_x \geq 20$. For each subset, the ordinary least squares linear regression method LinearRegression() provided in the Scikit Learn library in Python was used to obtain fits for the RBE_{max}-LET^* relationship. A linear dependency of RBE_{max} on LET^* was applied, and the intercept was assumed to be equal to 1 for all the fits. In order to investigate a potential dependency on the irradiation condition, the fitting procedure was repeated using only the monoenergetic data in each subset and using only the SOBP data in each subset.

The resulting fitted lines were successively compared to the RORU model, the MCN model and the fits for RBE_{max} as a function of LET^* and $(\alpha/\beta)_x$ that were previously obtained using Equation (3.2).

Fitting on restricted databases

Different database restrictions in terms of reduced ranges of included $(\alpha/\beta)_x$ and LET^* values were imposed on the proton data points. The applied combinations of restrictions on $(\alpha/\beta)_x$ and LET^* values are given in Table 3. For each restricted database, the curve_fit() function provided in the Scipy library in Python was used to obtain fits for RBE_{max} as a function of LET^* and $(\alpha/\beta)_x$ according to the fitting function given in Equation (3.2). Separate fitting was done for the monoenergetic data and the SOBP data in each of the restricted databases. The resulting fit parameters were compared in order to investigate how restrictions on the data used in the fitting affect the model output in terms of RBE_{max} dependencies affecting predictions of RBE_{max} values.

Table 3: Database restrictions on $(\alpha/\beta)_x$ and LET^* applied to the proton data. The restrictions on $(\alpha/\beta)_x$ are combined with each of the restrictions on LET^* given in the same row. 1st to 4th row: Restrictions on the upper limits of the ranges of included $(\alpha/\beta)_x$ and LET^* values. 5th to 7th row: Restriction on the lower limit of the range of included $(\alpha/\beta)_x$ values and restrictions on the upper limits of the ranges of included $(\alpha/\beta)_x$ and LET^* values. 8th to 11th row: Restriction on the lower limit of the range of included LET^* values and restrictions on the

upper limits of the ranges of included $(\alpha/\beta)_x$ and LET^* values. 12th to 14th row: Restrictions on the upper and lower limits of the ranges of included $(\alpha/\beta)_x$ and LET^* values.

$(\alpha/\beta)_x$ values [Gy]	LET^* values [$keV/\mu m$]			
All	All	[0,20)	[0,10)	[0,5)
[0,20)	All	[0,20)	[0,10)	[0,5)
[0,10)	All	[0,20)	[0,10)	[0,5)
[0,5)	All	[0,20)	[0,10)	[0,5)
≥ 5	All	[0,20)	[0,10)	[0,5)
[5,20)	All	[0,20)	[0,10)	[0,5)
[5,10)	All	[0,20)	[0,10)	[0,5)
All	≥ 5	[5,20)	[5,10)	
[0,20)	≥ 5	[5,20)	[5,10)	
[0,10)	≥ 5	[5,20)	[5,10)	
[0,5)	≥ 5	[5,20)	[5,10)	
≥ 5	≥ 5	[5,20)	[5,10)	
[5,20)	≥ 5	[5,20)	[5,10)	
[5,10)	≥ 5	[5,20)	[5,10)	

Fitting on balanced databases

The amount of data points with $(\alpha/\beta)_x$ and LET^* values in the lower part of the range is considerably larger than the number of data points with higher values for $(\alpha/\beta)_x$ and LET^*, especially for the SOBP data (as seen in Figure 14 and Figure 15). To take this into consideration in the fitting of RBE_{max} as a function of LET^* and $(\alpha/\beta)_x$, actions were taken to balance the data used in the fitting.

To account for the large amount of data points with low $(\alpha/\beta)_x$, the proton data points with $(\alpha/\beta)_x < 20\ Gy$ were divided into four subsets, each of them containing data points with $(\alpha/\beta)_x$ values in the intervals [0,5), [5,10), [10,15) and [15,20), respectively. Five random data points were picked from each of the subsets, and the 20

randomly chosen data points were used to obtain a fit for RBE_{max} as a function of LET^* and $(\alpha/\beta)_x$ using Equation (3.2). This was repeated 100 times, and the mean value of the resulting 100 fit parameters was calculated. Since this mean k value is the result of many fits using equally many data points with $(\alpha/\beta)_x$ values in each of the four considered parts of the range $[0,20)$, it is obtained with a more balanced database. By the same means, mean k values were obtained using only monoenergetic data and only SOBP data.

The same procedure was used to obtain fits to a balanced database with respect to LET^* values, using the same intervals, but now in terms of $keV/\mu m$. It was observed that the database contained only three SOBP data points with LET^* values in the interval $[15,20)$. In order to obtain at least five SOBP data points in each of the four subsets, all proton data with $LET^* < 21\ keV/\mu m$ was included in the procedure. Additionally, the procedure was carried out for proton data with $LET^* < 20\ keV/\mu m$, repeating the fitting to 12 randomly chosen data points 165 times so that the total number of data points used to obtain the mean fit parameter would be approximately the same as before.

3.2.3 Non-linear regression

Fitting of proton RBE data without database restrictions

To explore the appliance of non-linear RBE_{max}-LET^* relationships, the curve_fit() function was used to obtain fits for RBE_{max} as a function of LET^* and $(\alpha/\beta)_x$, using different fitting functions of varying complexity. The fitting functions included polynomials from 1st to 4th degree, as well as exponential functions. The RBE_{max} was assumed to be inversely proportional with $(\alpha/\beta)_x$, and the intercept as assumed to be equal to 1. The fitting functions can thus be written

$$RBE_{\max} = 1 + \frac{1}{(\alpha/\beta)_x} f(LET^*), \tag{3.3}$$

where the multiplicative function $f(LET^*)$ take the forms given in Table 4. Each of the resulting functions were fitted to all the proton data, only monoenergetic proton data and only SOBP proton data. To get an estimate of the goodness of the fits, the predictions on RBE_{max} and the real RBE_{max} values of the data points were used to calculate the root mean squared error (RMSE) for each of the resulting fits using the following equation:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y^{(i)} - f(\boldsymbol{x}^{(i)})\right)^2} \tag{3.4}$$

where n is the number of data points, $y^{(i)}$ is the real RBE_{max} value associated to the i^{th} set of $(\alpha/\beta)_x$ and LET^* values $\boldsymbol{x}^{(i)}$, and $f(\boldsymbol{x}^{(i)})$ is the predicted value for this set of $(\alpha/\beta)_x$ and LET^* values. In order to compare the results to the previously obtained fits where a linear RBE_{max}-LET^* relationship was applied, the RMSE of these fits were also calculated.

Table 4: Multiplicative functions inserted into (3.3) to obtain fits for RBE_{max} as a function of LET^* and $(\alpha/\beta)_x$, applying a non-linear relationship between RBE_{max} and LET^*. k_1-k_4 are fit parameters to be found in the regression procedure.

Name	Multiplicative function
f_{LQ}	$f(LET^*) = k_1 LET^* + k_2 LET^{*2}$
f_{LQC}	$f(LET^*) = k_1 LET^* + k_2 LET^{*2} + k_3 LET^{*3}$
f_{LQCq}	$f(LET^*) = k_1 LET^* + k_2 LET^{*2} + k_3 LET^{*3} + k_4 LET^{*4}$
f_{Q}	$f(LET^*) = k_1 LET^{*2}$
f_{QC}	$f(LET^*) = k_1 LET^{*2} + k_2 LET^{*3}$
f_{QCq}	$f(LET^*) = k_1 LET^{*2} + k_2 LET^{*3} + k_3 LET^{*4}$
f_{LE}	$f(LET^*) = (k_1 LET^*)\exp(-k_2 LET^*)$
f_{QE}	$f(LET^*) = \left(k_1 LET^{*2}\right)\exp(-k_2 LET^*)$
f_{LQE}	$f(LET^*) = \left(k_1 LET^* + k_2 LET^{*2}\right)\exp(-k_3 LET^*)$

f_{LE2}	$f(LET^*) = (k_1 LET^*)\exp(-k_2 LET^{*^2})$
f_{QE2}	$f(LET^*) = \left(k_1 LET^{*^2}\right)\exp(-k_2 LET^{*^2})$

Fitting on restricted databases

The non-linear fitting procedure was applied to two different restricted databases. The first database included proton data points with $(\alpha/\beta)_x < 25\,Gy$ and $LET^* < 20\,keV/\mu m$. The second database were obtained with the same restrictions on the upper limit of included $(\alpha/\beta)_x$ an LET^* values and an additional restriction on the lower limit of included LET^*, excluding data points with $LET^* < 5\,keV/\mu m$.

Fitting on balanced databases

Non-linear fitting on balanced databases in terms of $(\alpha/\beta)_x$ and LET^* was performed by appliance of the same procedure that was previously used to obtain linear fits to balanced databases. The procedure was repeated for each of the non-linear fitting functions resulting from insertion of the multiplicative functions in Table 4 into Equation (3.3), using all proton data, only monoenergetic data and only SOBP data. The mean fit parameters from fitting on a balanced database with respect to $(\alpha/\beta)_x$ were obtained from 100 non-linear fits to random samples with 5 data points from each subset with $(\alpha/\beta)_x$ values in the intervals [0,5), [5,10), [10,15) and [15,20). Fit parameters from fitting on a balanced database with respect to LET^* were retrieved by the same procedure, using four subsets of data points with LET^* values in each of the intervals [0,5), [5,10), [10,15) and [15,21)

Comparison of non-linear fits with the Rørvik weighted model

The five non-linear fitting functions that were indicated best suited to model the proton data were used to plot RBE_{max} as a function of LET. In order to compare the non-linear fits to the similar fitting results from the Rørvik model obtained with a weighted database (RORW model), the same relationship was plotted using the RBE_{max} function from this model.

Application of different models

To finally compare not only the RBE_{max} functions, but also the RBE estimates of each model, dose, LET and $(\alpha/\beta)_x$ dependencies of RBE were quantified using Equation (2.10). A selection of the linear and non-linear fits from fitting of RBE_{max} on databases containing both monoenergetic and SOBP data were used to obtain RBE models by insertion into Equation (2.10), assuming that $RBE_{min} = 1$. The resulting models were used to plot estimated RBE values as a function of LET^*, dose and $(\alpha/\beta)_x$. For each of the considered variables, the other two were kept constant at clinically relevant levels. The biological doses obtained from three of the linear models, two of the non-linear models and the RORU, RORW and MCN models were also explored through an SOBP scenario. In both the RBE and the biological dose estimates, the considered LET values are not given by the normalized LET (LET^*) relative to the LET of Cobalt-60, as the data points from the present database is not considered here.

4. Results

4.1 Database

The database contains 944 data points, 413 of them obtained with proton radiation and 531 are from irradiation experiments using other ions than protons. The data originates from a total of 140 different articles, most of these via the Paganetti and PIDE databases. In the following, only the proton data is presented as RBE modelling in this work was done using proton data only.

The proton data included $(\alpha/\beta)_x$ values in the range $0.577 - 114.9\ Gy$, while the ranges for the calculated LET^* values and the RBE_{max} values were $0.251 - 33.2\ keV/\mu m$ and $0 - 16.5$, respectively. 148 of the proton data points are obtained with monoenergetic radiation. The remaining 265 data points are obtained using SOBP beams. The distributions of the $(\alpha/\beta)_x$, LET^*, and RBE_{max} values for the proton data points are visualized by the histograms in Figure 14, Figure 15 and Figure 16, with separate plots showing the distributions for data points obtained with monoenergetic radiation and SOBPs.

As seen in Figure 14(a), the majority of the proton data points have relatively low $(\alpha/\beta)_x$ values. There are 334 data points with $(\alpha/\beta)_x \leq 10\ Gy$, and 241 of the data points have $(\alpha/\beta)_x \leq 5\ Gy$. This indicates that the majority of cell irradiation experiments are more representative for late responding tissues which are associated with low $(\alpha/\beta)_x$ values. The histograms in Figure 14(b) and Figure 14(c) shows a similar distribution for both the monoenergetic and the SOBP data points, with the majority of them having $(\alpha/\beta)_x$ values in the lowest part of the range.

$(\alpha/\beta)_x$ values above $20\ Gy$ are unusual and less relevant as they indicate abnormal repair mechanisms in the cells. Data with such $(\alpha/\beta)_x$ values are often excluded in RBE modelling, as the model should represent cells with normal repair mechanisms. As the proton data included a broad range of $(\alpha/\beta)_x$ values, and only 14 data points

had values above $20\ Gy$, a split axis was used in the histograms in Figure 14 to emphasize the distribution of $(\alpha/\beta)_x$ values for the data points with $(\alpha/\beta)_x \leq 20\ Gy$.

The mean LET^* value of the monoenergetic proton data is $11.3\ keV/\mu m$, while the SOBP data has a mean LET^* of $3.7\ keV/\mu m$. The distribution of LET^* values in Figure 15(a) shows a concentration of data points with values between 0 and 4 Gy, with a peak between $1-2\ Gy$. 328 of the data points have $LET^* \leq 10\ keV/\mu m$, and 271 of them are in the interval $0-5\ keV/\mu m$. The distribution of LET^* values for SOBP proton data in Figure 15(c) shows a very similar distribution to the one in Figure 15(a), with the largest portion of the data points having values below $5\ keV/\mu m$. However, Figure 15(b) shows a different distribution for the monoenergetic data points, where the LET^* values are more evenly distributed across the whole range, with a slow decreasing trend for increasing LET^*.

From the histograms in Figure 16, we see that almost all the proton data points have RBE_{max} values below 5. There are 20 data points with $RBE_{max} > 5$, and 17 of them are monoenergetic data points. One of the monoenergetic data points stand out from the rest, with a high value for RBE_{max} of 16.5. This data point is a potential outlier, and this should be kept in mind when using the database for further work.

In Figure 17, LET^* is plotted against $(\alpha/\beta)_x$ for all proton data points having $(\alpha/\beta)_x \leq 20\ Gy$, with separate plots for the monoenergetic and SOBP proton data. In Figure 17(a), there are three regions with high density of data points, restricted by $LET^* < 5\ keV/\mu m$ and $(\alpha/\beta)_x$ values in the intervals $1-2.1\ Gy$, $4-5\ Gy$ and $6-7\ Gy$. As seen in Figure 17(c), most of the data points in these regions are SOBP data. The majority of the SOBP data points have LET^* values in the lower range. Although this is seen for all values of $(\alpha/\beta)_x$, the variation in LET^* is higher for the SOBP data with $(\alpha/\beta)_x$ values in the lower range, compared to the data points with higher $(\alpha/\beta)_x$ values. On the other hand, looking at Figure 17(b), the monoenergetic data points are more spread, and there are no regions with distinct groups in the plot. There seems to

be a tendency of higher LET^* values for higher $(\alpha/\beta)_x$, as only one of the 24 monoenergetic data points with $(\alpha/\beta)_x > 10\ Gy$ has $LET^* < 5\ keV/\mu m$.

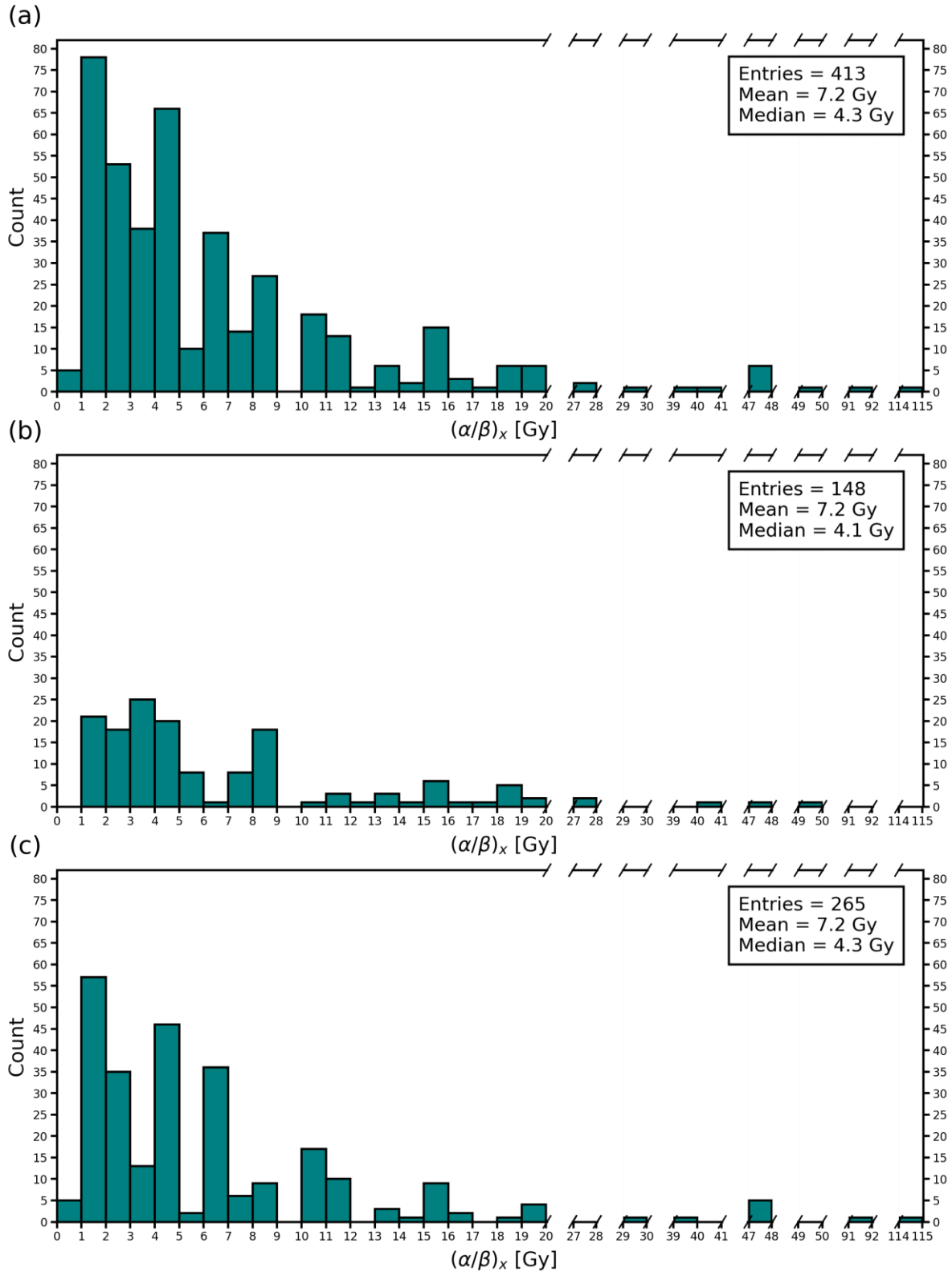

Figure 14: Histograms showing the distribution of $(\alpha/\beta)_x$ values for (a) all proton data points, (b) proton data points that are obtained with monoenergetic radiation and (c) proton data points that are obtained using an SOBP.

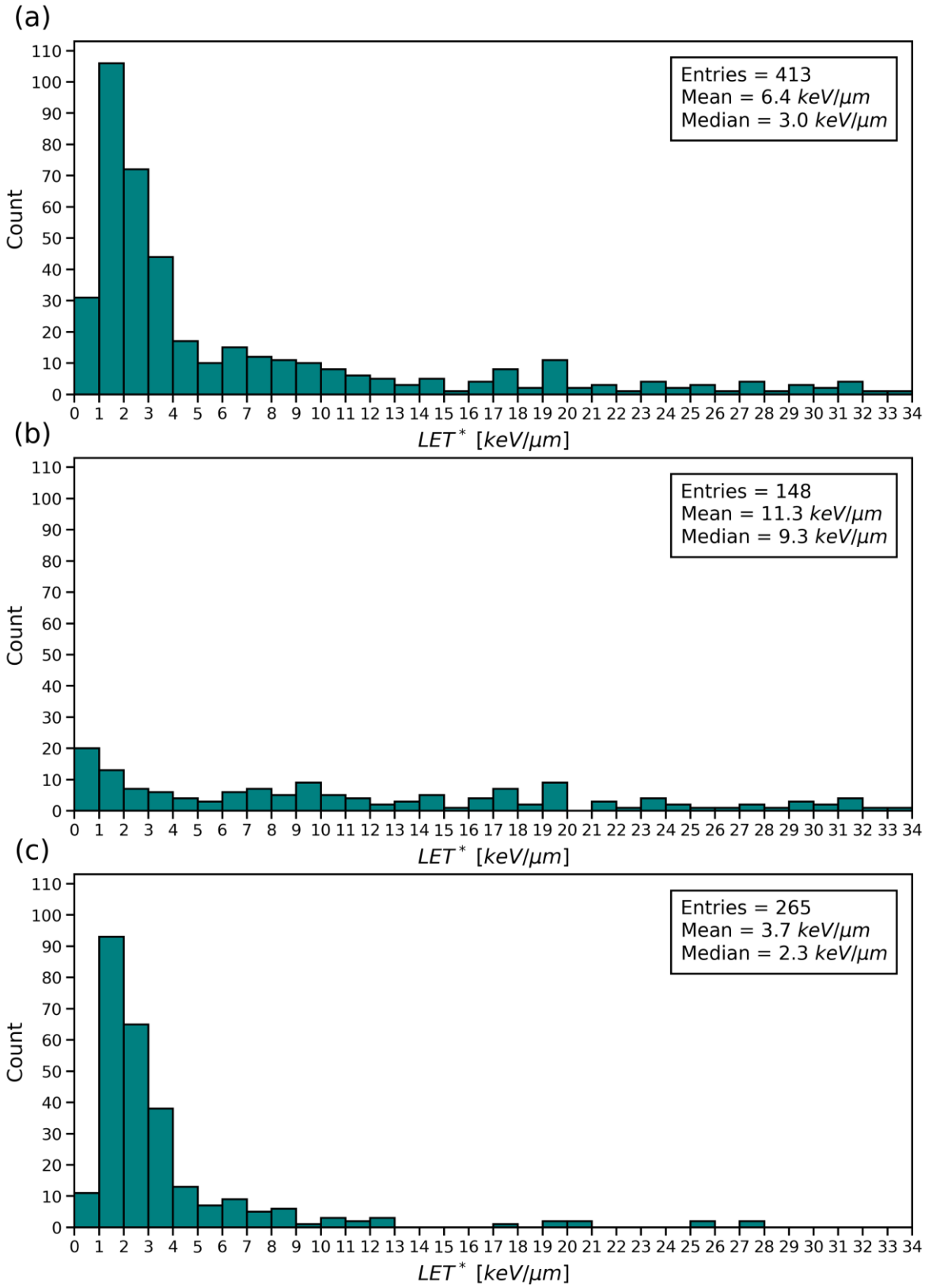

Figure 15: Histograms showing the distribution of LET^* values for (a) all proton data, (b) proton data obtained with monoenergetic radiation and (c) proton data obtained using an SOBP.

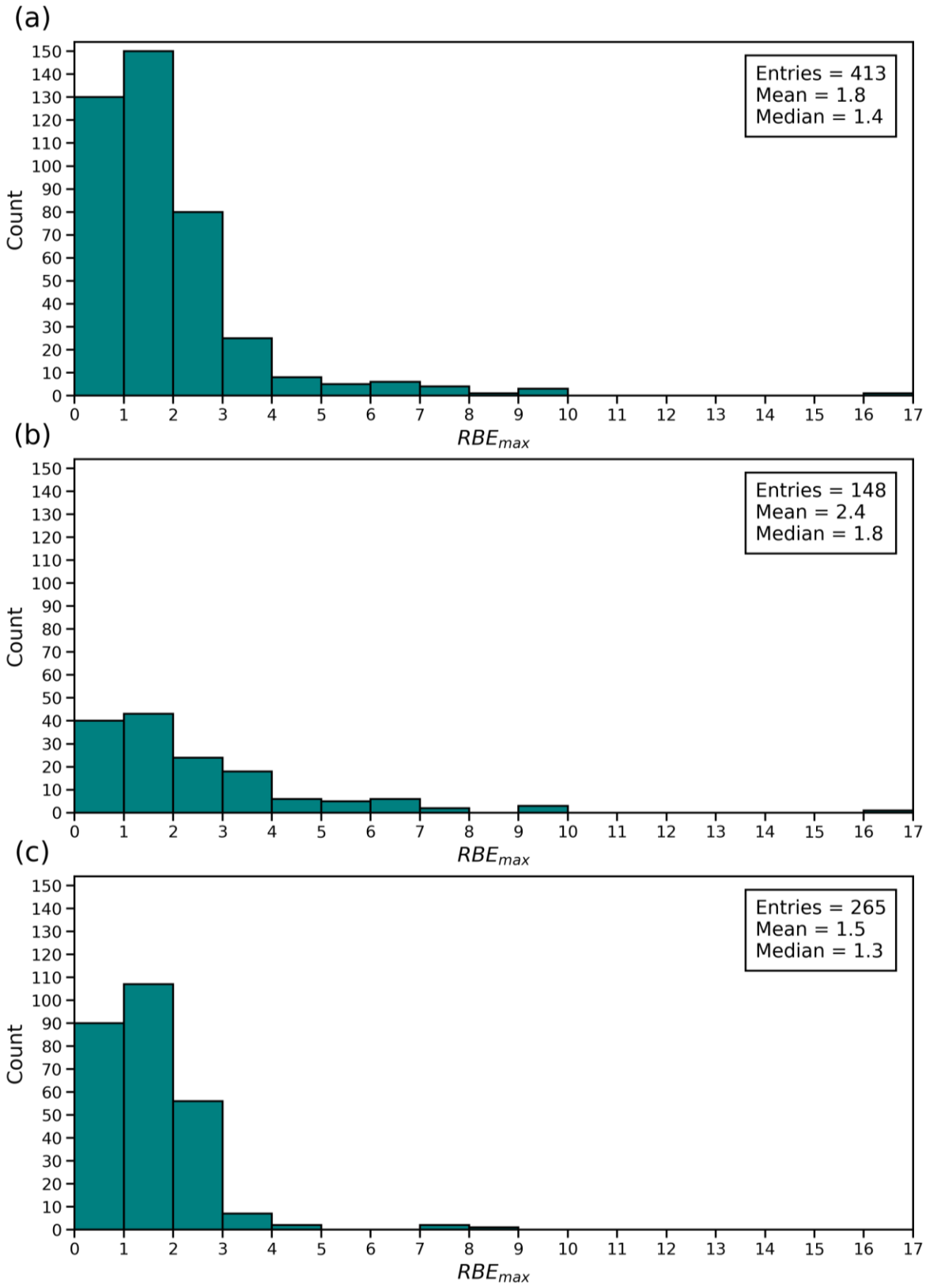

Figure 16: Histograms showing the distribution of RBE_{max} values for (a) all proton data, (b) only monoenergetic proton data and (c) only SOBP proton data.

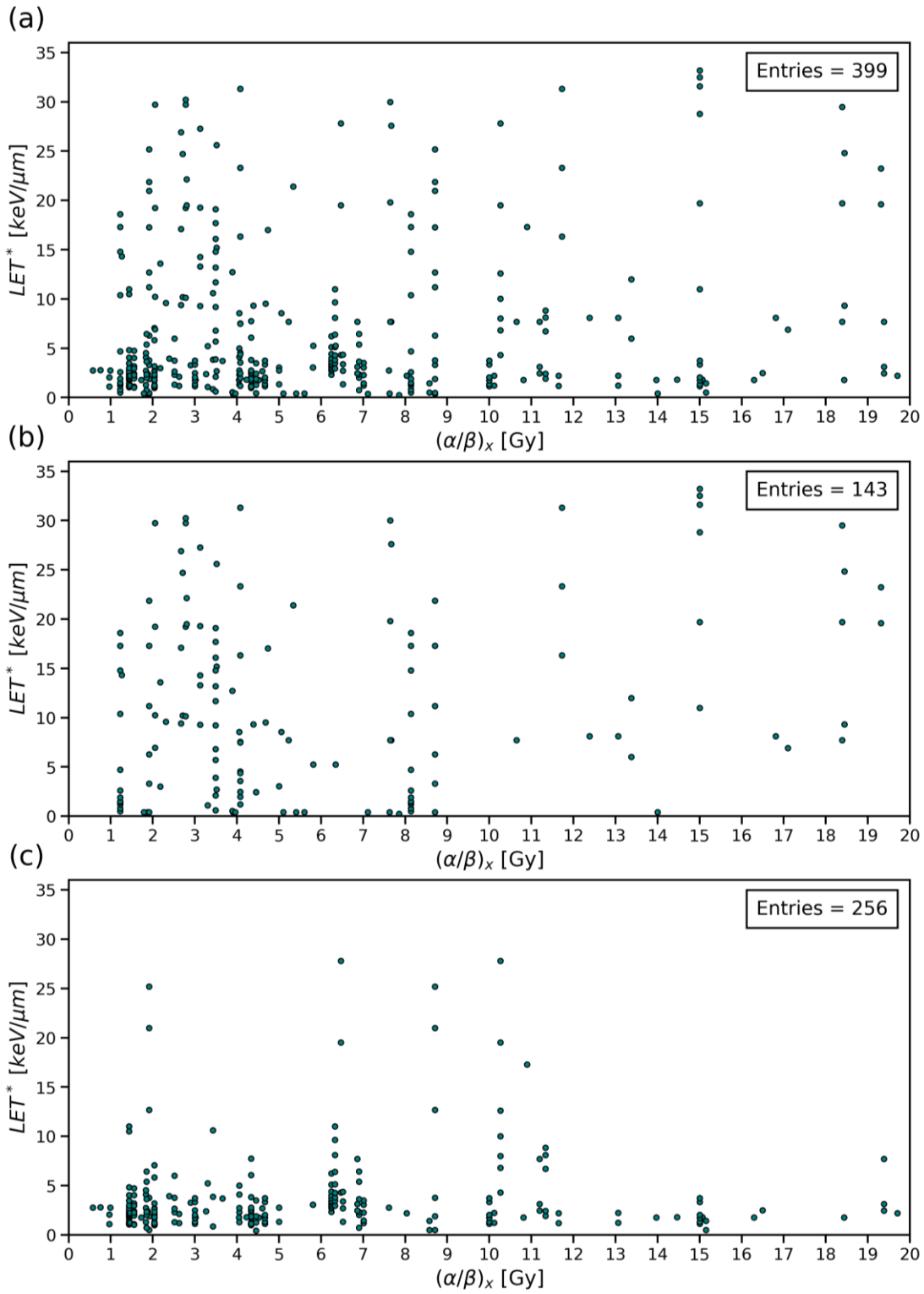

Figure 17: Proton LET^* plotted against $(\alpha/\beta)_x$ values. In (a), all proton data points with $(\alpha/\beta)_x \leq 20\,Gy$ are shown. Data points obtained using monoenergetic radiation and using an SOBP are shown in (b) and (c) respectively, with the same restriction for the $(\alpha/\beta)_x$ values as in (a). Similar plots without the restriction on $(\alpha/\beta)_x$ can be found in Figure A1 in the Appendix.

Key results

- The proton data covers a wide range of $(\alpha/\beta)_x$ values, but most of the data points have values indicating cells with normal repair mechanisms. The data is not evenly distributed across the $(\alpha/\beta)_x$ range, and the database is dominated by data points with $(\alpha/\beta)_x$ values between $1-9\ Gy$.
- The SOBP data is not evenly distributed over the range of LET^* values, and the majority of these data points have LET^* values between $1-4\ keV/\mu m$. The monoenergetic data have LET^* values that are more spread across the whole range, especially in the clinical relevant interval below $20\ keV/\mu m$. The different LET^* distributions for monoenergetic and SOBP data are also reflected by the mean LET^* values, which are $11.3\ keV/\mu m$ and $3.7\ keV/\mu m$, respectively.
- The database is dominated by proton data points with RBE_{max} values below 4.

4.2 Linear regression

4.2.1 Fitting of proton RBE data without database restrictions

Fits for RBE_{max} as a function of LET^* and $(\alpha/\beta)_x$ obtained by successively fitting Equation (3.2) to all the proton data, the monoenergetic proton data and the SOBP proton data are shown in Figure 18(a), (b) and (c), respectively. The general trend was similar for all fits, showing increasing RBE_{max} with increasing LET^* and decreasing RBE_{max} with increasing $(\alpha/\beta)_x$. It was observed in section 4.1 that the SOBP data is not evenly distributed over the range of LET^* and $(\alpha/\beta)_x$ values. This is also seen in Figure 18(c), which clearly shows that the database contain very few data points with higher values of LET^* in the considered $(\alpha/\beta)_x$ interval in the plot. Consequently, there are relatively large regions where the fit to SOBP data is based on very few data points, which can result in poor modelling in these regions. The resulting model from the SOBP data might thus be a good model for the lower region of LET^* values while

for higher LET^* the model might be less accurate and it should therefore be used with caution here.

By inserting the obtained fit parameters into Equation (3.2), the RBE_{max}-LET^* relationships are plotted for six different $(\alpha/\beta)_x$ values in Figure 19(a-c). We see that the slopes of the lines from the fits to all data ($k = 0.45 \pm 0.02$) and monoenergetic data ($k = 0.45 \pm 0.03$) are almost equal, while the slope of the line from the fit to SOBP data ($k = 0.48 \pm 0.03$) points is a little steeper. The fit parameters used to obtain the plots were not rounded to two decimals, thus the slopes of the lines from the fits to all data and monoenergetic data are not equal to each other as implied by the k values given above.

The fit parameters are very similar with overlapping values considering the estimated error. Thus applying a linear relationship between RBE_{max} and LET^* with a slope inversely dependent on $(\alpha/\beta)_x$ show no clear dependency between RBE_{max} and irradiation condition.

The fit parameters of the RORU model and the MCN model are $k = 0.65$ and $k = 0.36$, respectively. It should be noted that the prior model is derived only from monoenergetic proton data, while both monoenergetic and SOBP data was used to develop the latter model. The obtained fit parameters lies between the fit parameters of the RORU and MCN models, although the difference is larger when comparing to the prior model. In Figure 19(d-f), the fit to all proton data is used to plot the RBE_{max} as a function of LET^* for six different $(\alpha/\beta)_x$ values, and the same is done using the RORU model and the MCN model. For a given LET^* value, the span of RBE_{max} values between two curves plotted with different $(\alpha/\beta)_x$ values are widest for the RORU model, followed by the fit to all data and the MCN model, respectively. Thus, the RORU model shows the strongest dependency of RBE_{max} on $(\alpha/\beta)_x$ out of the three.

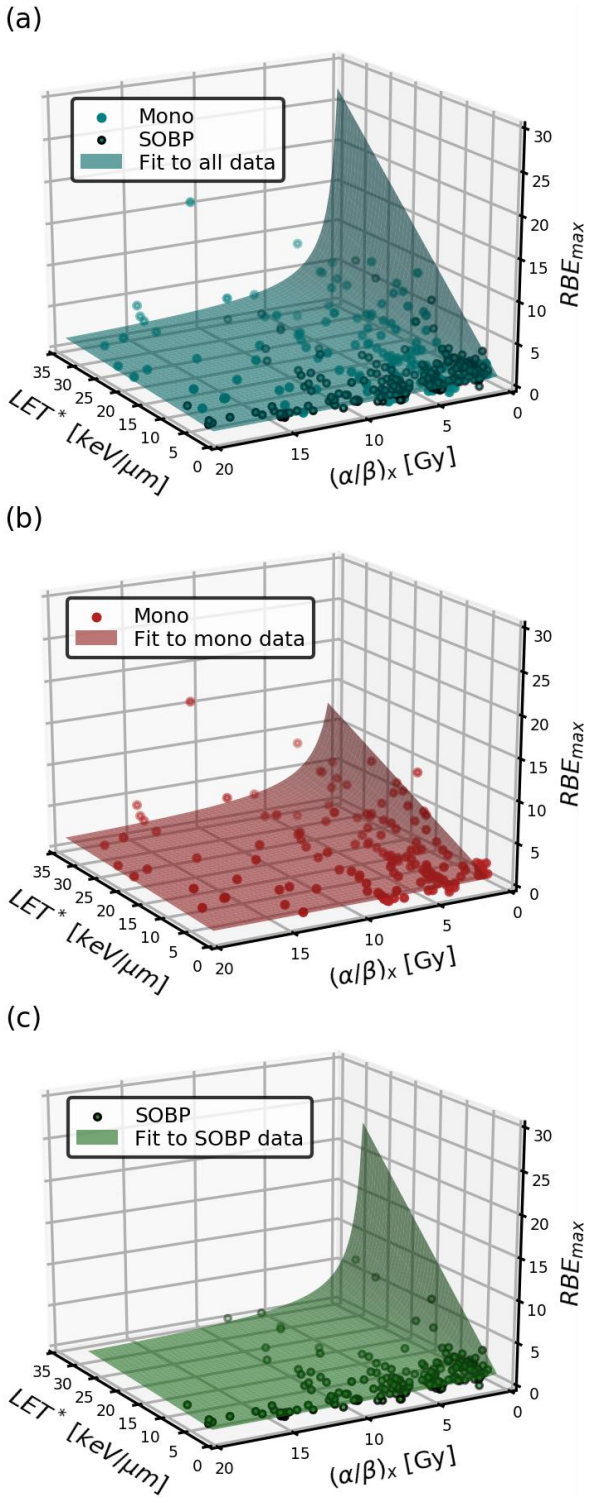

Figure 18: Fits for RBE_{max} as a function of LET^* and $(\alpha/\beta)_x$, assuming a linear relationship between RBE_{max} and LET^*, intercept equal to 1 and that RBE_{max} is inversely proportional with $(\alpha/\beta)_x$. The planes are obtained by fitting Equation (3.2) to (a) all proton data points, (b) only the monoenergetic proton data and (c) the SOBP proton data points. Although all values of $(\alpha/\beta)_x$ are included in the fitting, the planes are plotted for $(\alpha/\beta)_x \leq 20\ Gy$.

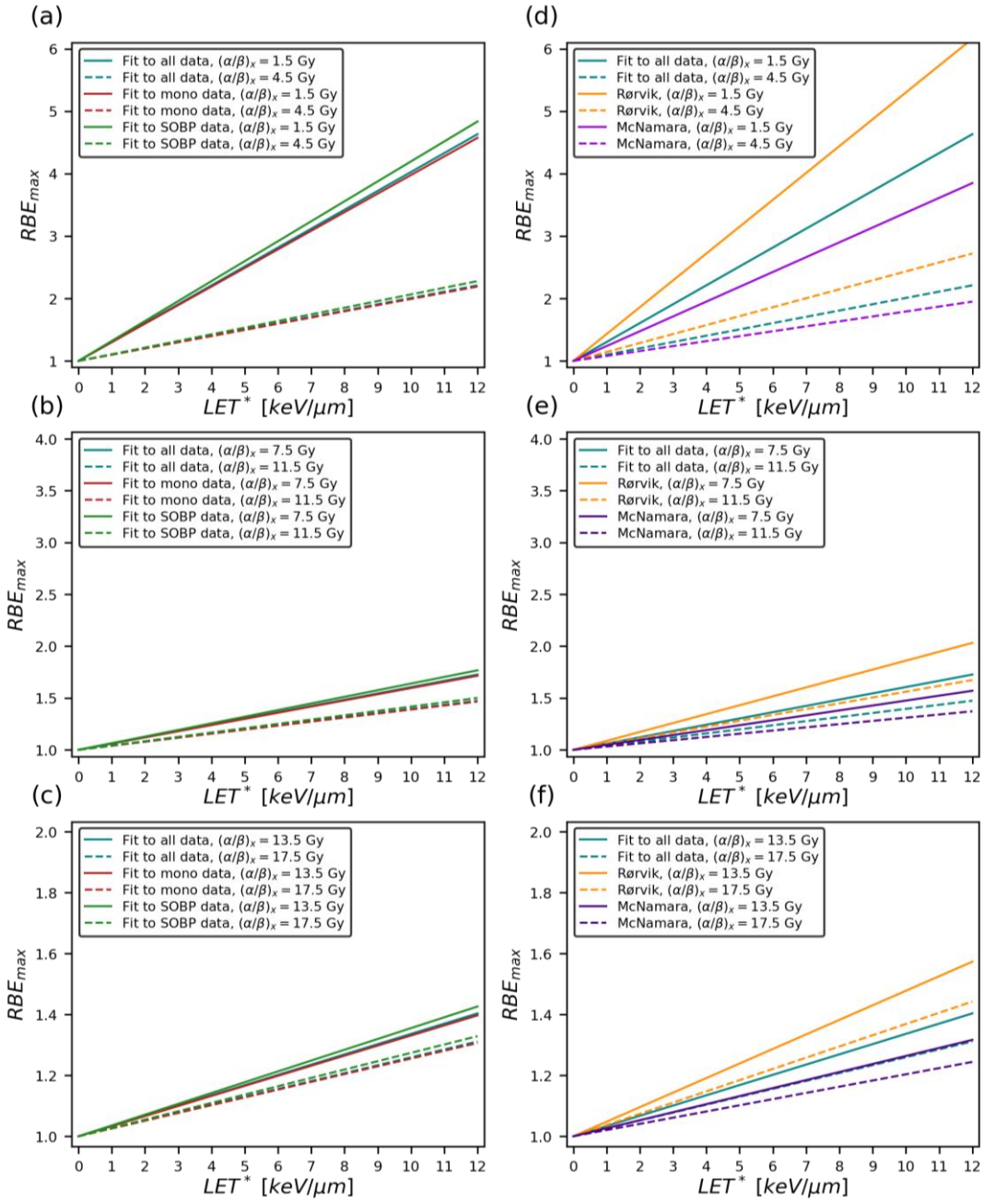

Figure 19: (a-c): RBE_{max} as a function of LET^* for different values of $(\alpha/\beta)_x$, according to the fits to all proton data (blue), monoenergetic proton data (red) and SOBP proton data (green) shown in Figure 18(a), (b) and (c), respectively. The lines are plotted by inserting the obtained fit parameters and the different $(\alpha/\beta)_x$ values into Equation (3.2). (d-f): RBE_{max} as a function of LET^* for different $(\alpha/\beta)_x$ values, according to the fit to all proton data shown in Figure 18(a) (blue), the RORU model (Equation (2.15)) (orange) and the MCN model (Equation (2.11)) (indigo). The blue lines are obtained by inserting the obtained fit parameter and the different $(\alpha/\beta)_x$ values into Equation (3.2), and the orange and indigo lines are obtained by inserting the different $(\alpha/\beta)_x$ values into Equations (2.15) and (2.11), respectively.

Key results

- RBE_{max} from all, monoenergetic and SOBP data was similar when fitting to Equation (3.2) without database restrictions.
- Although small, a trend could be seen that the fit to SOBP data yielded a steeper slope for the RBE_{max}-LET^* relationship compared to the fit to monoenergetic data. This could indicate a dependency on the irradiation condition used, and should be further investigated. However, the SOBP data is not evenly distributed and modelling to these data is likely to be subject to relatively large uncertainties for higher LET^* values. To take this into consideration, a method to obtain fits to balanced databases are explored in section 4.2.4.
- Compared to the RORU model and the MCN model, the fit to all data yielded a RBE_{max}-LET^* relationship for specific values of $(\alpha/\beta)_x$ with a smaller slope than the prior and a steeper slope than the latter.

4.2.2 Analysis of RBE dependence on $(\alpha/\beta)_x$

The $k/(\alpha/\beta)_x$ relation that is applied in the regression fits in Figure 18 is a widely used assumption for the dependency of RBE_{max} on $(\alpha/\beta)_x$ [22]. For each of the seven subsets described in section 3.2.2, an ordinary least square linear regression method was used to obtain a fit for the RBE_{max}-LET^* relationship, as shown in Figure 20. The procedure was repeated using only the monoenergetic data and only the SOBP data in each subset, as shown in Figure 21 and Figure 22, respectively.

Now, each of the seven fits in Figure 20, Figure 21 and Figure 22 represents the relationship between RBE_{max} and LET^* for a specific interval of $(\alpha/\beta)_x$ values, instead of using $(\alpha/\beta)_x$ as a variable. The assumption of a $k/(\alpha/\beta)_x$ relation for the dependency of RBE_{max} on $(\alpha/\beta)_x$ is thus not applied in the fits. According to this assumption, RBE_{max} is inversely proportional with $(\alpha/\beta)_x$. Thus, if the assumption is valid for the data, the regression fit for the subset with lowest $(\alpha/\beta)_x$ values should have the steepest slope, and the steepness should decrease successively for the subsets with higher $(\alpha/\beta)_x$ values.

In Figure 20, we see that the fit to data points with $(\alpha/\beta)_x$ values in the interval [0,3) has the steepest slope (red line), followed by the fits to proton data having $(\alpha/\beta)_x$ values in [9,12) (light green line), > 20 (pink line), [3,6) (orange line), [12,15) (blue line), [6,9) (dark green line), and [15,20) (purple line), respectively. Thus, the data do not confirm the assumption that RBE_{max} is inversely proportional to $(\alpha/\beta)_x$. However, the red line has the steepest slope, while the smallest slopes belong to the purple line. This indicates an overall trend of decreasing slope with increasing $(\alpha/\beta)_x$, even though the steepness of the fitted lines do not decrease successively for each of the following subsets containing data points with higher $(\alpha/\beta)_x$ values. The slope of the pink line is larger than expected from the assumption. However, the data points used to obtain this fit have $(\alpha/\beta)_x$ values typical for cells with abnormal repair mechanisms and are less relevant in a clinical setting.

As seen in Figure 20(d), the subset used to obtain the light green line contain one datapoint with an RBE_{max} value that stand out from the other. This data point can explain the unexpected steep slope of the fitted line for this subset, as it will affect the slope in a positive direction. The pink, orange, blue, dark green and purple lines have very similar slopes, all of them within a range of 0.0280. By comparison, the difference between the slope of the pink and light green lines is almost four times this range. As the experimental uncertainties of the data are not taken into consideration, the fitted curves are also associated with these uncertainties. Consequently, the five fits with very similar slopes could show more agreement with the assumption of a steeper slope for decreasing $(\alpha/\beta)_x$ if the experimental uncertainties were taken into account.

The fits to monoenergetic data in Figure 21 show some of the same properties as the fits to all proton data. Again, the fits to the $(\alpha/\beta)_x$ intervals [0,3) and [9,12) have the steepest slopes, although the latter is the steepest of them. This might be explained by the one data point with RBE_{max} value around 17, which will affect the light green line even more this time since there are only four monoenergetic data points with $(\alpha/\beta)_x$ values in [9,12). The slopes of the remaining fits are very similar, and the above

argument about experimental uncertainties will also apply to these fits. Looking away from the light green line, the overall trend of decreasing slope with increasing $(\alpha/\beta)_x$ can also be seen for the monoenergetic fits. The light green, blue and pink lines are fitted to respectively 4, 5 and 5 data points. This is not a satisfying amount of data, and the regression lines will be very sensitive to uncertainties and fluctuations in the small subsets they are fitted to.

In Figure 22, the fitted line to SOBP data with $(\alpha/\beta)_x$ values in the interval [0,3) has the steepest slope, followed by the fits to SOBP data having $(\alpha/\beta)_x$ values in the intervals [3,6), > 20, [6,9), [9,12), [15,20) and [12,15), respectively. As seen in Figure 22(e), the last-mentioned fit have a negative slope for RBE_{max} as a function of LET^*. However, it is based on a subset of only 4 data points and can therefore be disregarded as an unreliable result. Among the remaining fits, the slope of the pink line is the only one contradicting the assumption of a decreasing slope with increasing $(\alpha/\beta)_x$, as the slope of the other 5 fits are decreasing for each subset with higher $(\alpha/\beta)_x$ values. However, the slopes of the pink, dark green, light green and purple lines are very similar to each other, so both indications and contradictions of the $(\alpha/\beta)_x$ dependency at issue might be due to experimental uncertainties and fluctuations in the data.

The monoenergetic data point with RBE_{max} value around 17 which is a potential outlier originated from Prise et al 1990 [55]. To investigate how much it influences the slopes of the fits in Figure 20(d) and Figure 21(d), this data point was excluded and the regression procedure was repeated for the relevant subsets. The resulting fits are shown in Figure 23(a) for all proton data and in Figure 23(c) for monoenergetic proton data. In Figure 23(b) and Figure 23(d) these new fits for data with $(\alpha/\beta)_x$ values in [9,12) are plotted together with the previously obtained fits for the other subsets. The slopes of the new fits are more in line with the assumption of a decreasing slope with increasing $(\alpha/\beta)_x$, especially for the new line fitted to the monoenergetic data, which has a smaller slope than the fitted line to monoenergetic data with $(\alpha/\beta)_x$ values in

$[0,3)$ (red line), as seen in Figure 23(d). The results can thus be interpreted to strengthen the suspicion that this data point could be considered an outlier. However, it should be kept in mind that the new fit to monoenergetic data with $(\alpha/\beta)_x$ values in the interval $[9,12)$ is based on only 3 data points.

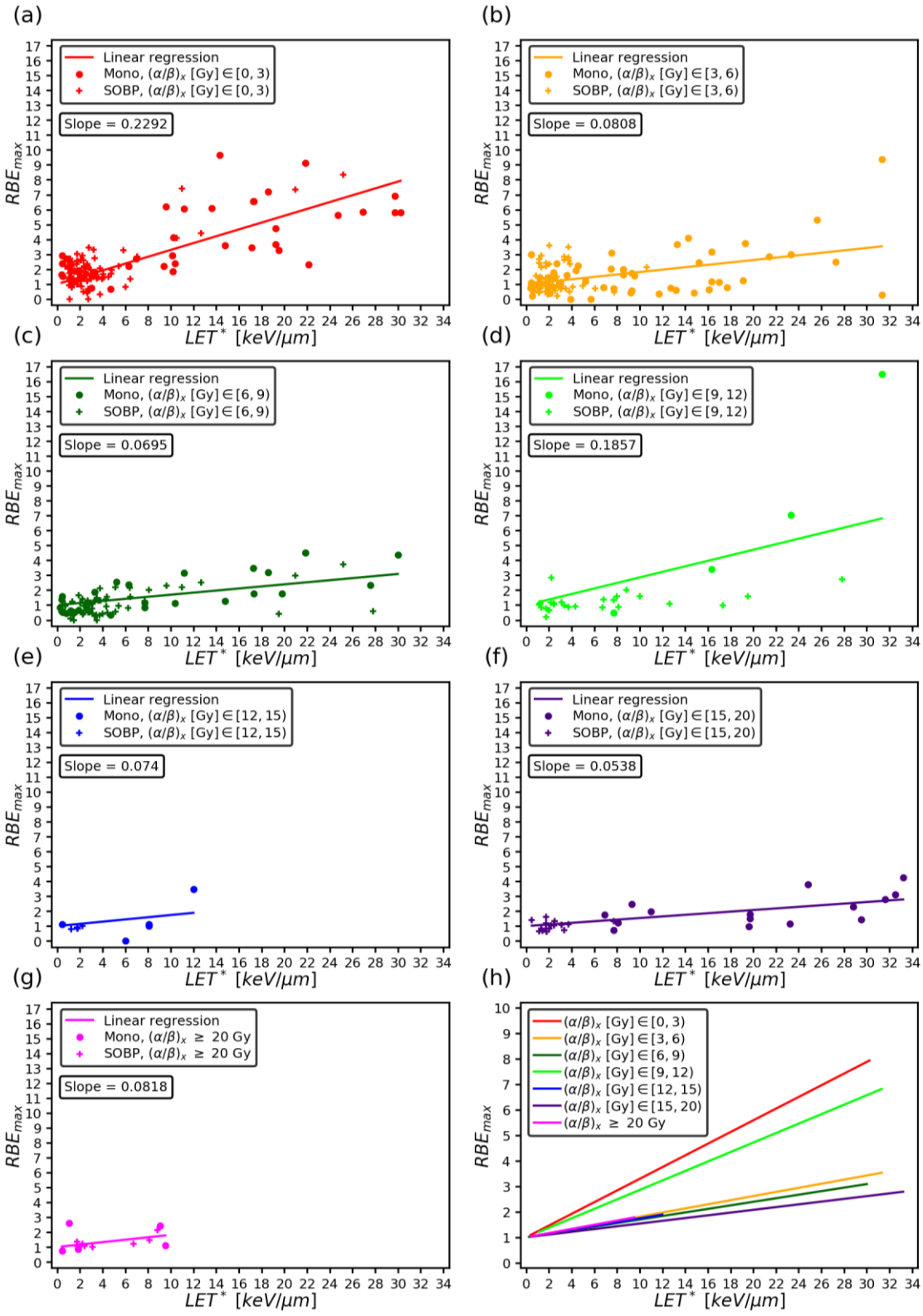

Figure 20: Linear regression fits for RBE_{max} as a function of LET^* for all proton data divided into seven subsets, each containing data points with $(\alpha/\beta)_x$ values in different intervals. The fitted curves for each subset are shown in (a)-(g), and all the regression lines are plotted together in (h).

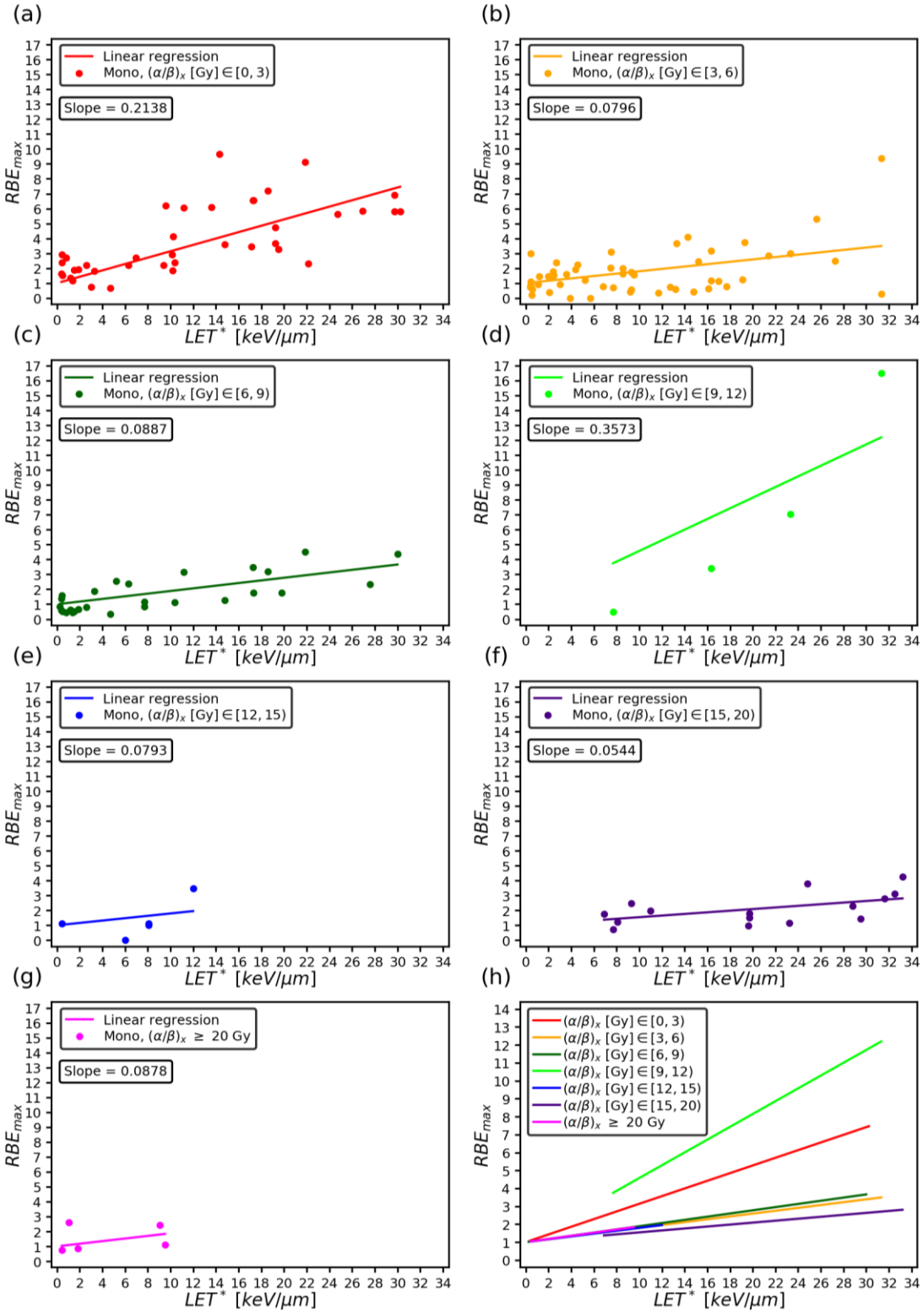

Figure 21: Linear regression fits for RBE_{max} as a function of LET^* for monoenergetic proton data divided into seven subsets, each containing data points with $(\alpha/\beta)_x$ values in different intervals. The fitted curves for each subset are shown in (a)-(g), and all the regression lines are plotted together in (h).

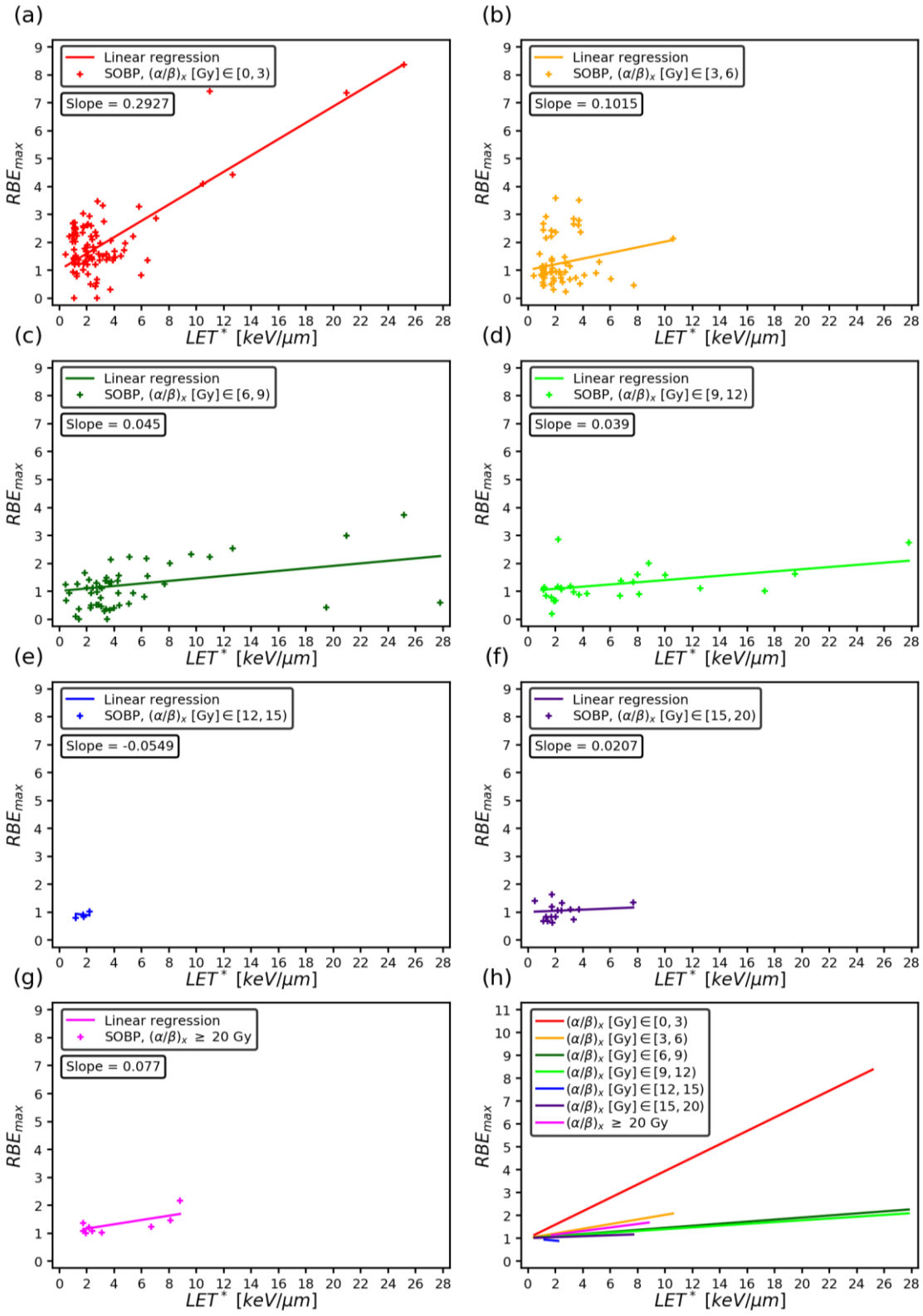

Figure 22: Linear regression fits for RBE_{max} as a function of LET^* for SOBP proton data divided into seven subsets, each containing data points with $(\alpha/\beta)_x$ values in different intervals. The fitted curves for each subset are shown in (a)-(g), and all the regression lines are plotted together in (h).

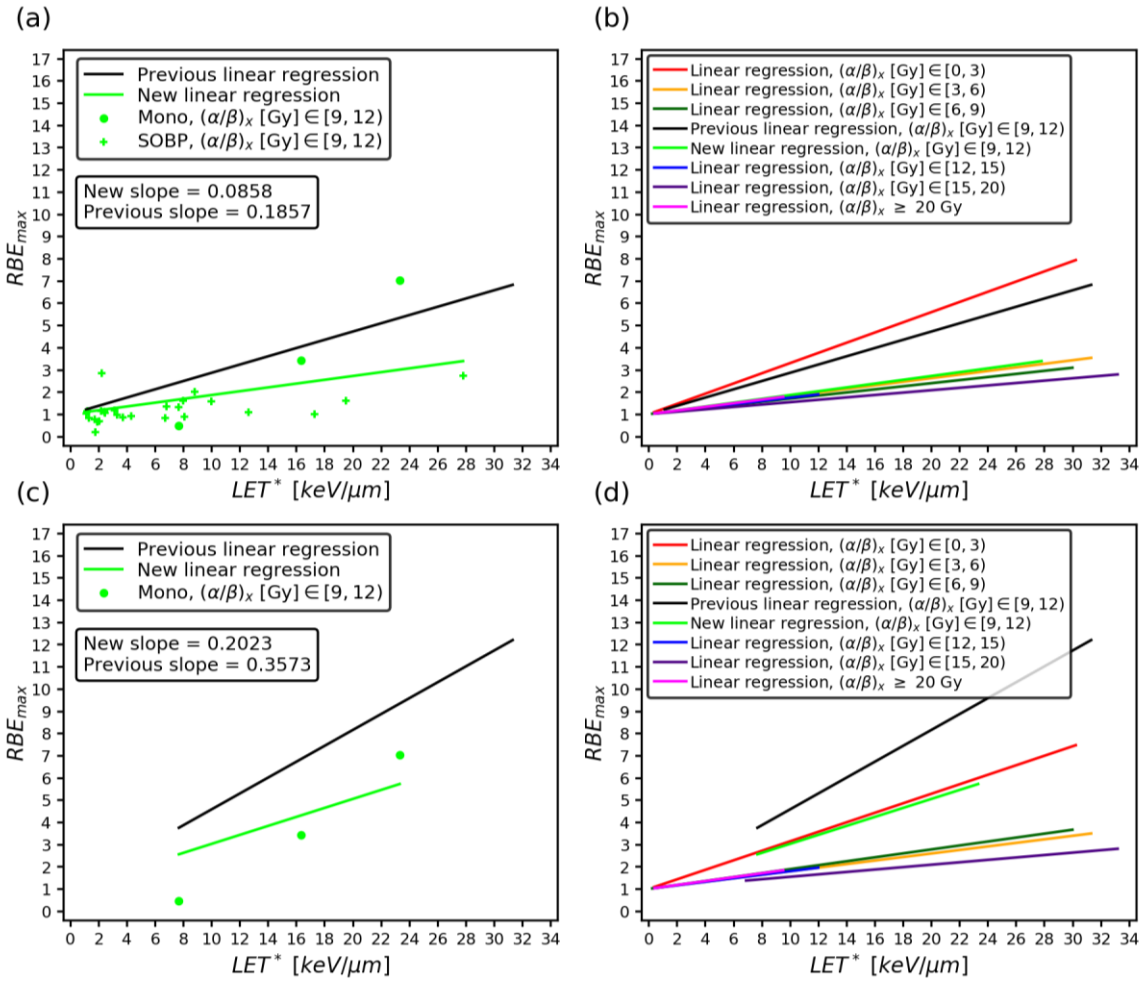

Figure 23: Linear regression fits for RBE_{max} as a function of LET^* for proton data with $9\ Gy \leq (\alpha/\beta)_x < 12\ Gy$, excluding the one data point with higher RBE_{max} value. (a): new fit to all proton data points, (b): the new fit to all proton data is plotted together with the previously obtained fits for the other six subsets, (c): new fit to monoenergetic proton data, (d): the new fit to monoenergetic proton data is plotted together with the previously obtained fits for the other six subsets.

The regression fits to all proton data points in Figure 20 are compared to the RORU model in Figure 24. The RBE_{max}-LET^* relationship according to the RORU model were plotted for each interval of $(\alpha/\beta)_x$ values by putting the mean $(\alpha/\beta)_x$ value of the data points in each subset into the RORU model obtained with an unweighted dataset (Equation (2.15)). Using the same equation, the area between the RORU lines corresponding to the upper and lower limits of $(\alpha/\beta)_x$ in each of the subsets were colored out. In order to do this for the subset with $(\alpha/\beta)_x$ values in $[0,3)$, the lower limit of $(\alpha/\beta)_x$ was set equal to $1\ Gy$, assuming that values below this limit are not of

clinical relevance. The fits to monoenergetic and SOBP data in Figure 21 and Figure 22 were also compared to the RORU model in the same way. The figures showing these comparisons are given in the Appendix (Figure A2 and Figure A3).

In Figure 24, we see that only two of the seven linear fits to all proton data lies between the upper and lower RORU lines; the fits to data points with $(\alpha/\beta)_x$ values in the intervals [0,3), and [6,9). For the three subsets with lowest $(\alpha/\beta)_x$ values (Figure 24(a-c)), the fitted lines have smaller slopes than the corresponding RORU lines, while for the last four intervals (Figure 24(d-g)) the opposite is true. The slopes of the RORU lines thus cover a wider range than the slopes of the fitted lines. All of the RORU lines for the mean $(\alpha/\beta)_x$ values in each subset are placed with approximately equal distance to the RORU lines for the upper and lower $(\alpha/\beta)_x$ values in each $(\alpha/\beta)_x$ interval. This indicates that the $(\alpha/\beta)_x$ values of the proton data are comparable to the $(\alpha/\beta)_x$ values of the data points used to develop the RORU model.

From the comparison between fits to monoenergetic data and the RORU model (Figure A2), it was observed that the fit to data points with $(\alpha/\beta)_x$ values in the interval [6,9) is the only one placed within the colored area. While the fits to monoenergetic data with $(\alpha/\beta)_x$ values in the intervals [0,3) and [3,6) had smaller slopes than the corresponding RORU lines, the rest of the fitted lines to monoenergetic data had steeper slopes than the corresponding RORU lines. The comparison between SOBP fits and the RORU model (Figure A3) showed that the fit to SOBP data with $(\alpha/\beta)_x$ values in the interval [0,3) is the only one of the seven fits to SOBP data that lies within the colored area defined by the upper and lower RORU lines. The fit to SOBP data points with $(\alpha/\beta)_x \geq 20\ Gy$ is the only line that has a steeper slope than the corresponding RORU line.

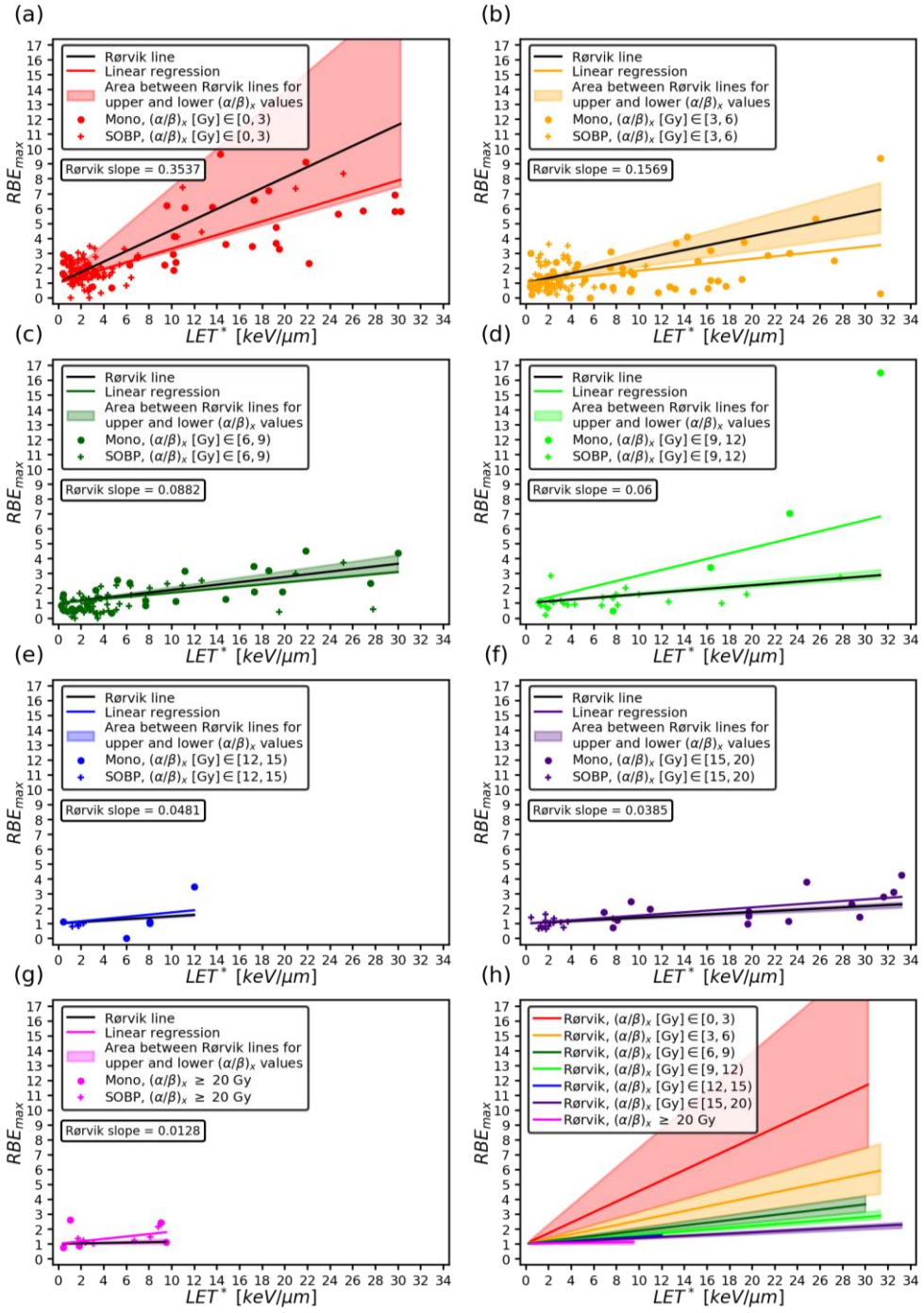

Figure 24: Comparison of the linear fits to all proton data (shown in Figure 20) and the RORU model. (a)-(g) shows the regression fits (colored lines), the RORU lines obtained by putting the mean of the $(\alpha/\beta)_x$ values for the data points in each interval of $(\alpha/\beta)_x$ into Equation (2.15) (black lines) and the area spanned by the curves obtained with Equation (2.15) using the upper and lower limits for the $(\alpha/\beta)_x$ value in each interval (colored area). In (a), the area is shown for a lower limit of $(\alpha/\beta)_x = 1\ Gy$, assuming that clinical use will not involve $(\alpha/\beta)_x$ values below this. The slope of the RORU lines are also given in (a)-(g). The RORU lines and limit areas for each $(\alpha/\beta)_x$ interval are potted together in (h).

In Figure 25, the fitted lines to all proton data in Figure 20 are compared to the MCN model in the same manner as explained for the comparison to the RORU model above. The fits to monoenergetic data and SOBP data were also compared to the MCN model, and similar figures showing these comparisons are given in the Appendix (Figure A4 and Figure A5).

The linear fits to all proton data with $(\alpha/\beta)_x$ values in the intervals [0,3) and [3,6) lie between the MCN lines for the upper and lower limits in the $(\alpha/\beta)_x$ intervals. All the fitted lines, except from the fit to data points with $(\alpha/\beta)_x$ values in [3,6), have a steeper slope than the corresponding MCN lines plotted with the mean $(\alpha/\beta)_x$ value in each subset.

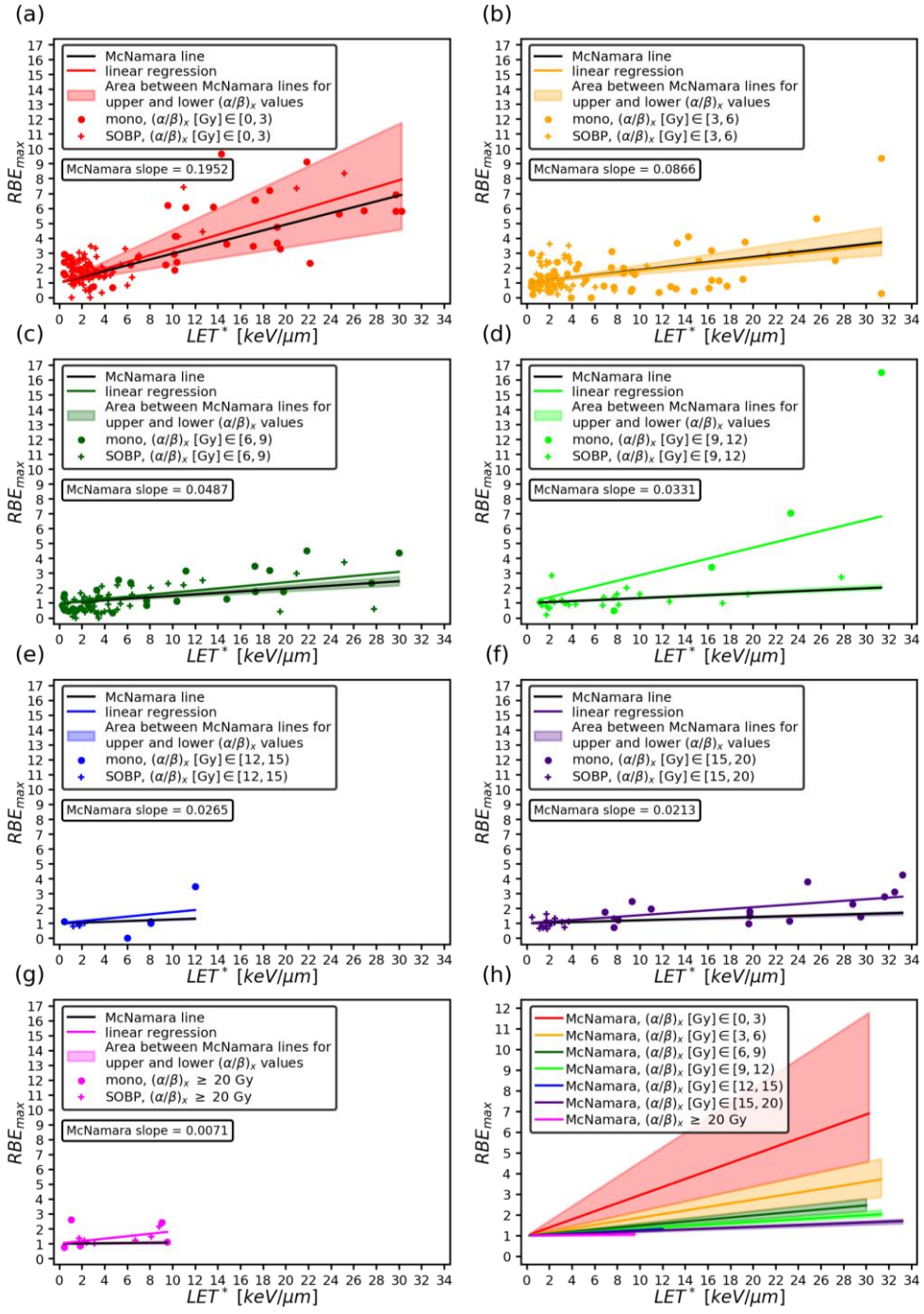

Figure 25: (a)-(g): The linear fits in Figure 20 for all proton data divided into seven subsets are shown (colored lines) together with the MCN lines obtained by using the mean $(\alpha/\beta)_x$ value for the data points in each subset as an input in Equation (2.11) (black lines), The slope of the MCN lines are given in each plot. The area between the MCN lines obtained by using the upper and lower limits of each $(\alpha/\beta)_x$ interval as an input in Equation (2.11) are also shown (colored areas). A lower limit of $(\alpha/\beta)_x = 1\ Gy$ is used in (a), assuming that clinical use will not involve $(\alpha/\beta)_x$ values below this. The MCN lines and limit areas for each $(\alpha/\beta)_x$ interval are potted together in (h).

The linear fits for RBE_{max} as a function of LET^* for each subset of data points with $(\alpha/\beta)_x$ values in different intervals shown in Figure 20, Figure 21 and Figure 22 were finally compared to the 3D fits for RBE_{max} as a function of LET^* and $(\alpha/\beta)_x$ shown in Figure 18. Equation (3.2) was used to plot the RBE_{max}-LET^* relationship according to the 3D fit to all data.

The linear fits to proton data with $(\alpha/\beta)_x$ values in the intervals [0,3), [3,6) and [6,9) are inside the respective areas defined by the lines from the 3D fit to all data for the upper and lower boundaries in each $(\alpha/\beta)_x$ interval, as seen in Figure 26. The linear fits to proton data points with $(\alpha/\beta)_x$ values in intervals [0,3) and [3,6) have smaller slopes than the corresponding lines from the 3D fit to all data. This indicates that the RBE_{max} might be overestimated for low $(\alpha/\beta)_x$ values and under-estimated for higher $(\alpha/\beta)_x$ values when RBE_{max} is assumed to have an inverse proportionality on $(\alpha/\beta)_x$, compared to the estimated RBE_{max} values when this assumption is not applied.

Figures showing similar comparisons of linear fits and 3D fits to monoenergetic data and SOBP data can be found in the Appendix (Figure A6 and Figure A7).

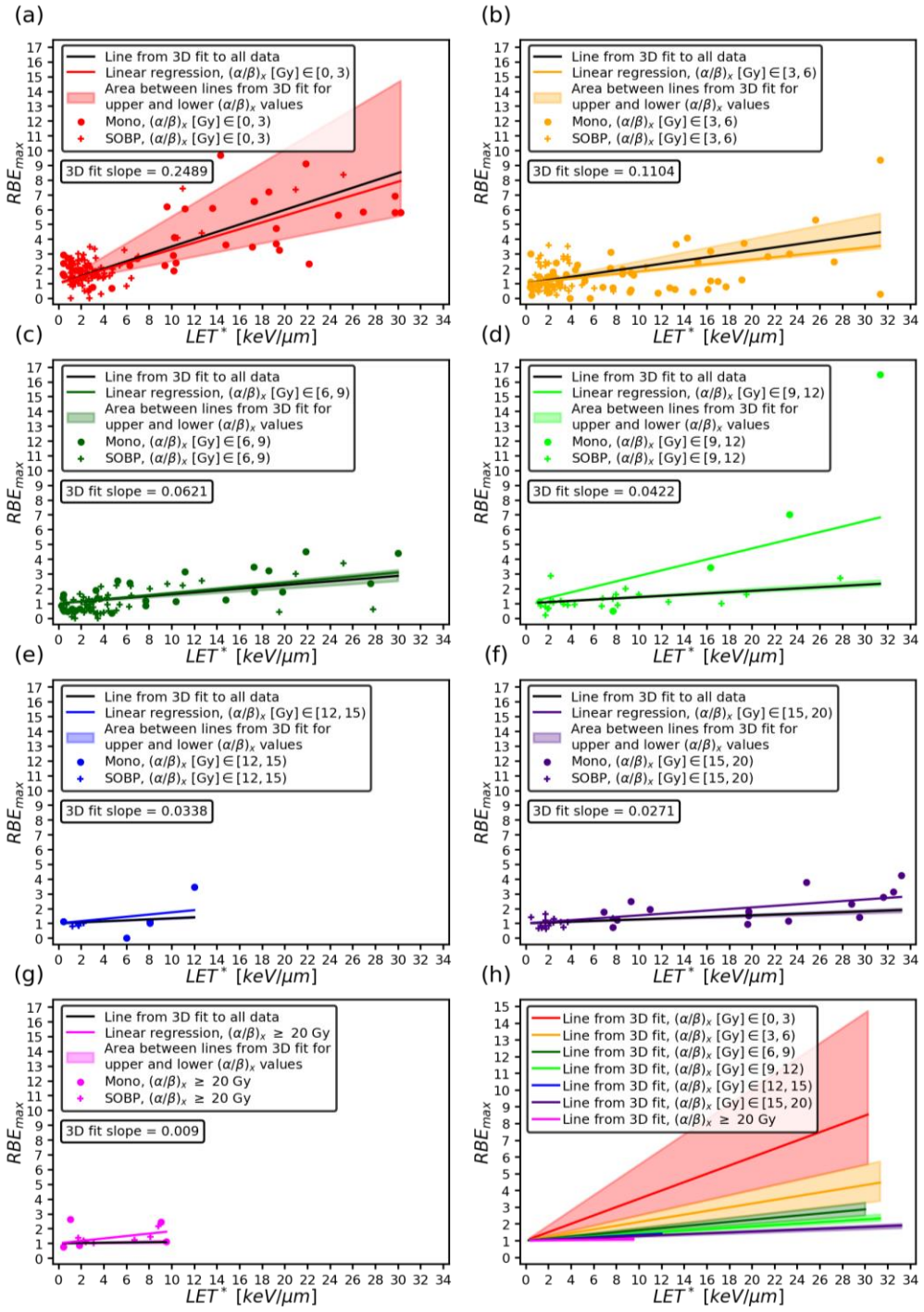

Figure 26: The linear regression fits to all proton data points from Figure 20 are shown in (a-g) together with lines obtained from the fitted plane to all data from Figure 18(a). These lines are plotted by inserting the obtained fit parameter $k = 0.45$ into Equation (3.2) and using the mean $(\alpha/\beta)_x$ value of the data points in each of the subsets as an input for $(\alpha/\beta)_x$ in the equation. The slopes of the lines from the 3D fit are given in each plot. The colored area in each plot is obtained by using the same equation, and the upper and lower boundaries in each interval of $(\alpha/\beta)_x$ values. In (h), the lines from the 3D fit and the limit areas for each $(\alpha/\beta)_x$ interval are plotted together.

Key results

- Regarding the RBE_{max}-LET^* relationships from linear regression fits to subsets containing proton data with $(\alpha/\beta)_x$ values in different intervals, an overall agreement with the assumption of an inverse proportionality of RBE_{max} on $(\alpha/\beta)_x$ were observed both for the fits to all proton data, and the fits considering monoenergetic proton data and SOBP proton data separately.
- The amount of monoenergetic data used in the analysis is clearly too small to get reliable results from the procedure that was applied to investigate the $(\alpha/\beta)_x$ assumption.
- The comparison between the fitted lines for the RBE_{max}-LET^* relationship and the 3D fit to all proton data indicated that applying the assumption that RBE_{max} is inversely proportional with $(\alpha/\beta)_x$ might lead to overestimated RBE_{max} values for low $(\alpha/\beta)_x$ ($\lesssim 6\ Gy$) and underestimated RBE_{max} values for higher $(\alpha/\beta)_x$.

4.2.3 Fitting on restricted databases

Equation (3.2) was used to obtain fits for RBE_{max} as a function of LET^* and $(\alpha/\beta)_x$ using different database restrictions in terms of reduced ranges for LET^* and $(\alpha/\beta)_x$. For each combination of database restrictions, the resulting fit parameters from successive fitting to all the proton data points in question, only the monoenergetic data points and only the SOBP data points are given in Figure 27(a), Figure 28(a) and Figure 29(a). The corresponding standard deviation errors (SDEs) of the fit parameters are given in Figure 27(b), Figure 28(b) and Figure 29(b), and the number of data points used in each fit are given in Figure 27(c), Figure 28(c) and Figure 29(c).

The fit parameters in the heatmap in Figure 27(a) are obtained using different restrictions on the upper limit of the range of included $(\alpha/\beta)_x$ and LET^* values. The previous obtained fit parameters from fitting to the unrestricted proton database are given in the first row for comparison.

The heatmap in Figure 27(a) indicates that the restrictions on $(\alpha/\beta)_x$ have a relatively small impact on the outcome of the fitting. Reducing the upper limit of included $(\alpha/\beta)_x$ values to $20\,Gy$ and $10\,Gy$ generally resulted in marginal deviations in the obtained fit parameters, as most of the k values obtained with these $(\alpha/\beta)_x$ restrictions (row 5-12 in the heatmap) were equal to the ones obtained with no restriction on $(\alpha/\beta)_x$ and coinciding restrictions on LET^* (row 1-4 in the heatmap). However, when the upper limit of included $(\alpha/\beta)_x$ values is reduced to $5\,Gy$, the resulting fit parameters (row 13-16 in the heatmap) are generally a little higher than the corresponding k values obtained with coinciding restrictions on LET^* and higher/no restriction on the upper limit of included $(\alpha/\beta)_x$ values (row 1-12 in the heatmap).

While the $(\alpha/\beta)_x$ restrictions seem to have a relatively small impact on the outcome of the fitting, the heatmap in Figure 27(a) indicates that the restrictions on LET^* affect the fit parameters in a more pronounced way. There is a clear trend of decreasing k values for each restriction reducing the range of included LET^* values. This trend is observed for all restrictions on $(\alpha/\beta)_x$ and both for the fit parameters from the fits to all proton data and the fit parameters from the fits to monoenergetic data and SOBP data. However, an exception from this trend is observed for the k values from fits to SOBP data with $LET^* < 5\,keV/\mu m$, which are higher than the k values from fits to SOBP data with $LET^* < 10\,keV/\mu m$ for all the applied restrictions on $(\alpha/\beta)_x$.

As the SOBP proton data results from irradiation with a series of overlapping Bragg peaks, the LET^* value of each data point actually represent a broad range of LET^* values. Consequently, the monoenergetic data are generally more suitable to provide a better understanding of the LET dependency.

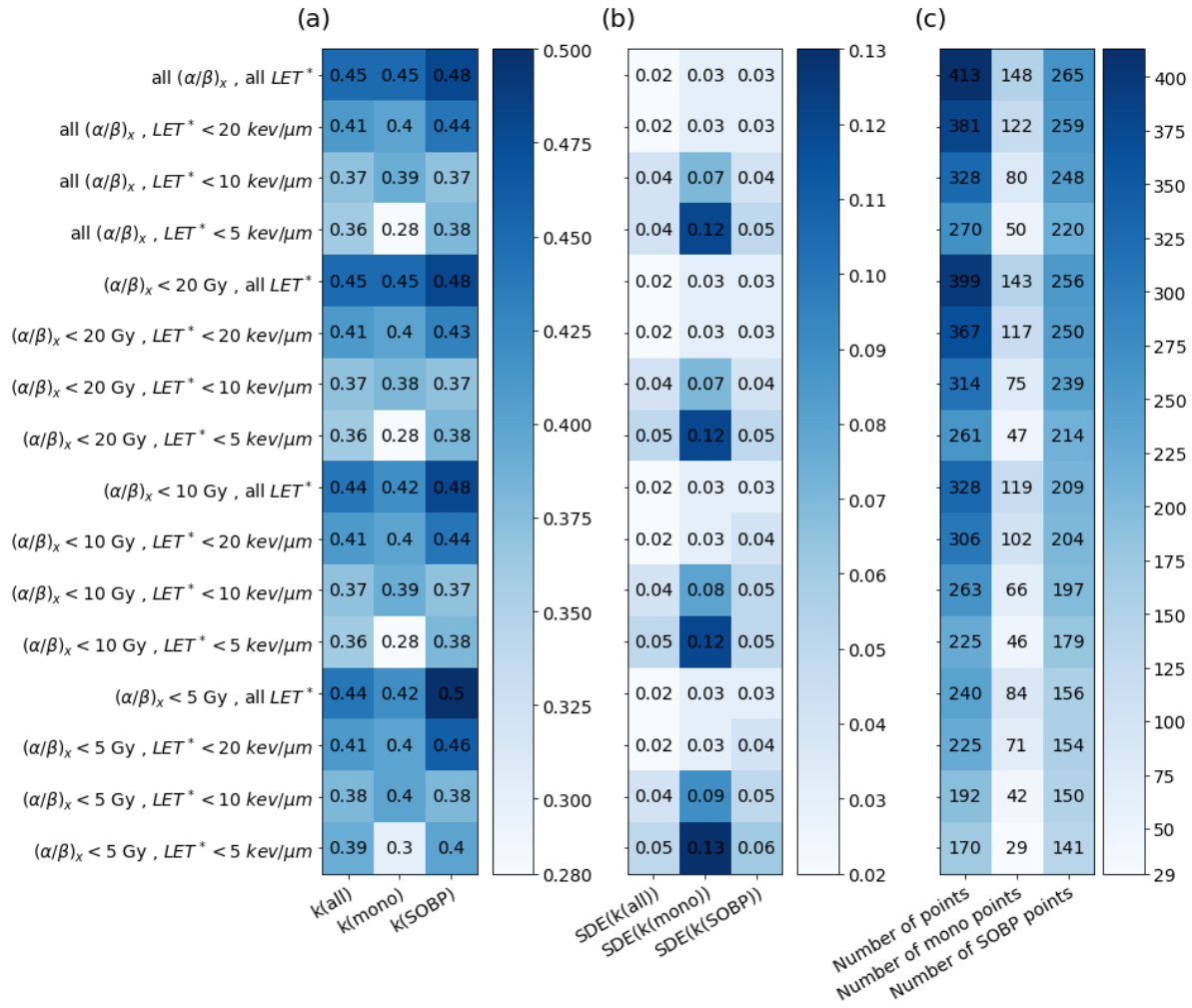

Figure 27: Resulting fit parameters from fitting on restricted databases according to Equation (3.2) are visualized by a heatmap (a). For each combination of reduced LET^* and $(\alpha/\beta)_x$ ranges, the fitting function was applied to all the relevant proton data, and separate fits were done for the relevant monoenergetic and SOBP data points. Corresponding SDEs calculated for each fit value and the number of data points in each of the restricted databases are given in the heatmaps in (b) and (c), respectively.

To obtain the fit parameters in the heatmap in Figure 28(a), a restriction was introduced on the lower limit of the range of included $(\alpha/\beta)_x$ values. As seen in the figure, proton data points with $(\alpha/\beta)_x < 5\ Gy$ are excluded in all the fits. To investigate the effect of this restriction, a comparison of the fit parameters in Figure 28(a) and the fit parameters in Figure 27(a) that was obtained using coinciding restrictions on the upper limit of the range of included $(\alpha/\beta)_x$ and LET^* values was carried out. The comparison yielded the following observations:

1) The fit parameters in Figure 28(a) cover a wider range than the ones in Figure 27(a).
2) The fit parameters in Figure 28(a) from fits to all data and monoenergetic data including all LET^* values and $LET^* < 20\ keV/\mu m$ are generally higher than the corresponding k values in Figure 27(a) that was obtained with coinciding restrictions on the upper limit of included $(\alpha/\beta)_x$ and LET^* values.
3) All of the fit parameters in Figure 28(a) that are obtained using all data points and monoenergetic data points with $LET^* < 10\ keV/\mu m$ and $LET^* < 5\ keV/\mu m$ are lower than the corresponding k values in Figure 27(a) that was obtained with coinciding restrictions on the upper limit of included $(\alpha/\beta)_x$ and LET^* values.
4) All fit parameters from fits to SOBP data in Figure 28(a) are lower than the corresponding fit parameters from fits to SOBP data using coinciding restrictions on the upper limit of the range of included $(\alpha/\beta)_x$ and LET^* values given in Figure 27(a).

The last mentioned observation shows a clear trend that the fits to SOBP data results in lower k values when the restriction on the lower limit of the range of included $(\alpha/\beta)_x$ values is applied. This indicates that an RBE model based on fitting with Equation (3.2) on data points having all values of $(\alpha/\beta)_x$ might overestimate the RBE for high $(\alpha/\beta)_x$ values. The outcome of fitting on SOBP data is largely affected by the restriction on the lower limit of $(\alpha/\beta)_x$, showing that the large amount of SOBP data with $(\alpha/\beta)_x$ values in this region is dominating in the fitting. To deal with this, a method of balancing the database should be applied to prevent the low $(\alpha/\beta)_x$ SOBP data from playing a too significant role in the fitting.

The fit parameters in Figure 28(a) indicates that lowering the upper limit of included $(\alpha/\beta)_x$ values result in a lower k value. A clear trend is observed that lowering the upper limit of included LET^* values result in lower k values. An exception from this trend is seen in the fit parameters from fits to SOBP data, as the k values from fits to

SOBP data with $LET^* < 20\ keV/\mu m$ are lower than the ones obtained from fits to SOBP data with $LET^* < 10\ keV/\mu m$ for all restrictions on the upper limit of included $(\alpha/\beta)_x$ values.

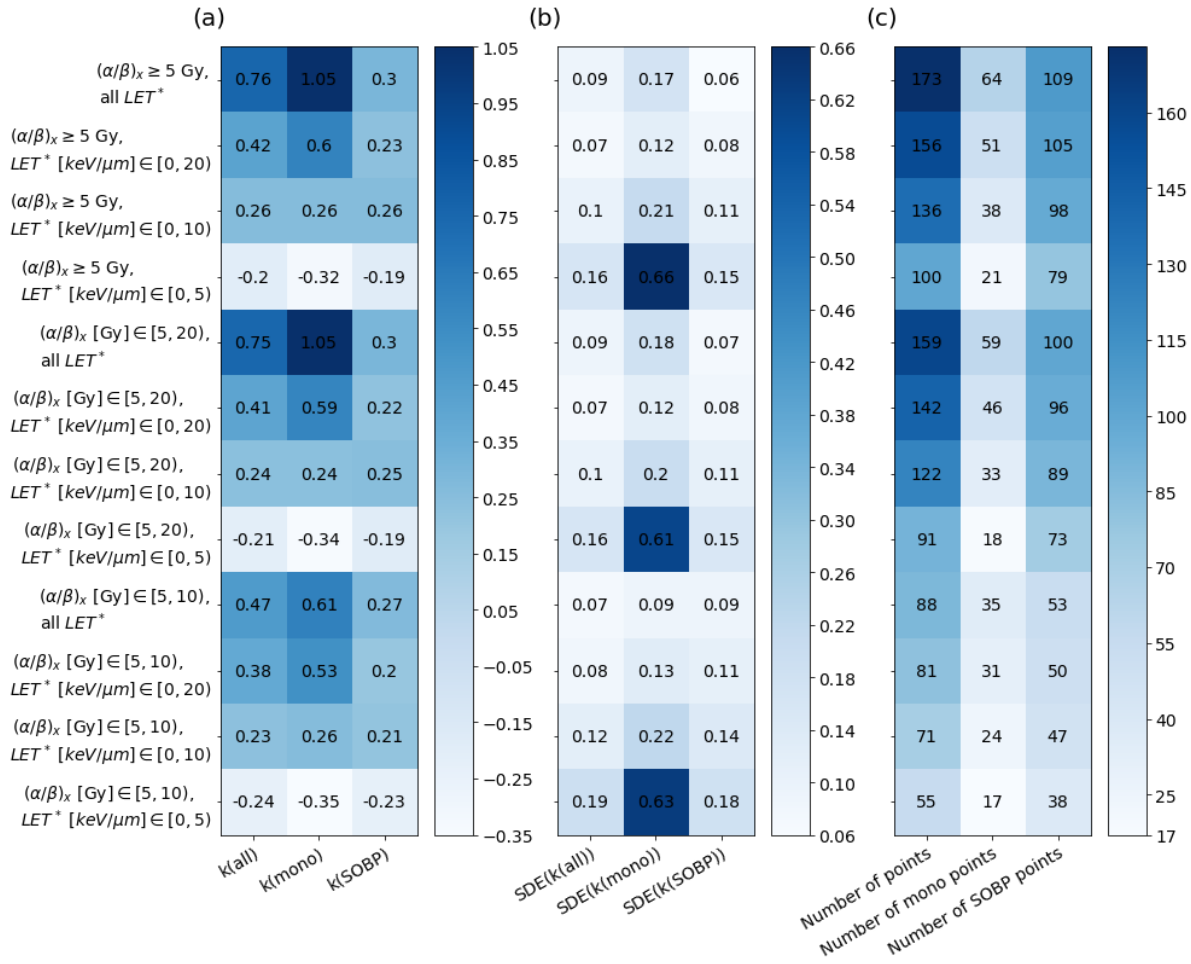

Figure 28: Fit parameters from fitting on restricted databases in terms of $(\alpha/\beta)_x$ and LET^* values are given in the heatmap in (a). The SDEs of the fit parameters are given in (b) and the number of data points in the restricted databases are given in (c). All the restricted databases are obtained with the same restriction on the lower limit in the range of included $(\alpha/\beta)_x$ values, as data points with $(\alpha/\beta)_x < 5\ Gy$ are excluded in all fits.

The fit parameters in the heatmap in Figure 29(a) are all obtained with the same restriction on the lower limit of the range of included LET^* values, excluding data points with $LET^* < 5\ keV/\mu m$.

In contrast to the fit parameters in the heatmap in Figure 28(a), the range of the fit parameters obtained with the restriction on the lower limit of included LET^* values are similar to the range of the fit parameters in Figure 27(a). Comparing the k values in Figure 29(a) to the corresponding k values obtained with the same restrictions on the upper limit of included $(\alpha/\beta)_x$ and LET^* values in Figure 27(a), three observations were done:

1) Fits to all data using all values of $(\alpha/\beta)_x$ and $(\alpha/\beta)_x < 20\ Gy$ resulted in larger k values when data points with $LET^* < 5\ Gy$ was excluded (although the differences are relatively small).
2) The remaining fit parameters from fits to all data points and monoenergetic data were either larger or equal to the corresponding k values in Figure 27(a) when the restriction on the lower limit of included LET^* values was applied.
3) A trend is observed for the fit parameters from fits to SOBP data in Figure 29(a) that the ones obtained with no restriction on the upper limit of included LET^* values and $LET^* < 20\ keV/\mu m$ are larger than the corresponding k values in Figure 27(a), while the fit parameters obtained with $LET^* < 10\ keV/\mu m$ are smaller than the corresponding ones in Figure 27(a).

The observed trend of increased k values in the fits to data points with $LET^* \geq 5\ keV/\mu m$ indicates that the slope of the applied linear RBE_{max}-LET^* relationship is steeper for higher LET^* values. This might imply a non linear trend for this relationship.

From the fit parameters in Figure 29(a), reducing the upper limit of included $(\alpha/\beta)_x$ values seem to have little impact on the outcome of the fitting. Although most of the k values obtained with the same restriction on LET^* are not equal to each other, the differences are small, and no clear trend is observed.

The trend that lowering the upper limit of the range of included LET^* values result in lower k values are also observed for the fit parameters from fits to all proton data and

SOBP data given in Figure 29(a). However, this trend is not seen for the fit parameters from fits to monoenergetic data with $LET^* \geq 5\ keV/\mu m$. These k values show several indications that no such trend is present, but the most prominent one is the fact that the fit parameter obtained with $LET^* < 10\ keV/\mu m$ is larger than the one obtained with $LET^* < 20\ keV/\mu m$ for all restrictions on $(\alpha/\beta)_x$.

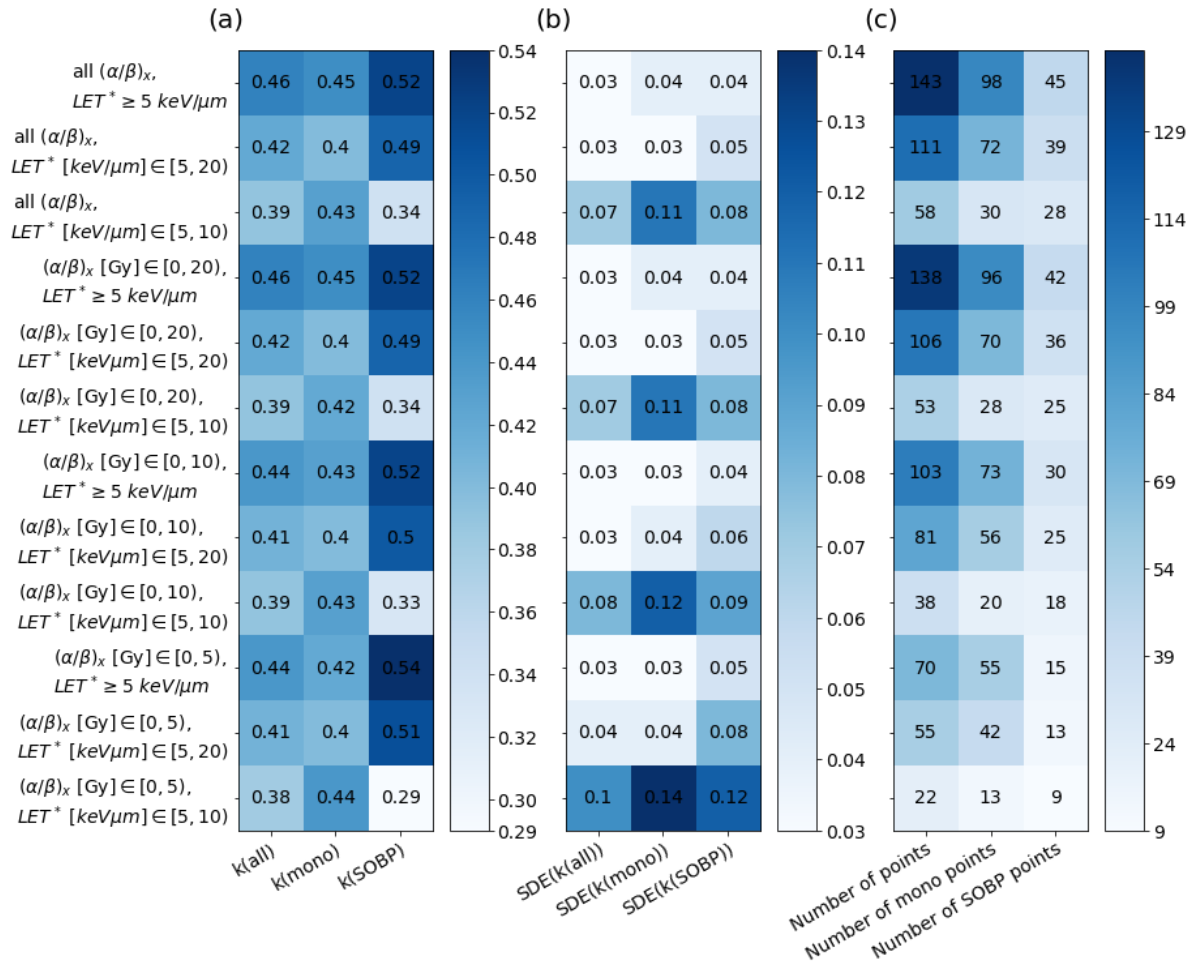

Figure 29: Fit parameters from fitting on restricted databases in terms of $(\alpha/\beta)_x$ and LET^* values are given in the heatmap in (a). The SDEs of the fit parameters are given in (b) and the number of data points in the restricted databases are given in (c). All the restricted databases are obtained with the same restriction on the lower limit in the range of included LET^* values, as data points with $LET^* < 5\ keV/\mu m$ are excluded in all fits.

In some cases, comparison of fit parameters obtained with coinciding restrictions on LET^* and different restrictions on the upper limit of included $(\alpha/\beta)_x$ values showed no clear dependency between the k values and the upper limits, while in other cases a trend was observed that lowering the upper limit of the range of included $(\alpha/\beta)_x$ values result in lower k values. Although no clear overall trend is observed, this might indicate that the assumed relationship between RBE_{max} and $(\alpha/\beta)_x$ in the fitting function (Equation (3.2)) is not a valid assumption for all values of $(\alpha/\beta)_x$.

As pointed out earlier, the fit parameters in Figure 28(a) obtained with data points having $(\alpha/\beta)_x \geq 5\ Gy$ show large deviations from the corresponding fit parameters in Figure 27(a) obtained with coinciding restrictions on the upper limits of included $(\alpha/\beta)_x$ and LET^* values. This indicates that the RBE_{max}-$(\alpha/\beta)_x$ relationship for the proton data points with $(\alpha/\beta)_x < 5\ Gy$ is different from the relationship for the remaining data points. The applied assumption that RBE_{max} is inversely proportional with $(\alpha/\beta)_x$ is not in line with this indication, and the need for exploration of other assumptions for the RBE_{max}-$(\alpha/\beta)_x$ relationship is thus implied.

Key results

- Restrictions on $(\alpha/\beta)_x$ in terms of reducing the upper limit of the range of included $(\alpha/\beta)_x$ values had relatively small impact on the outcome of the fitting.
- Reducing the upper limit of the range of included LET^* values shows a clear trend of lower k values. This trend is repeatedly observed in all the results from fitting on restricted databases, indicating a non-linear RBE_{max}-LET^* relationship.
- Excluding data points with $(\alpha/\beta)_x < 5\ Gy$, a new trend was observed that lowering the upper limit of included $(\alpha/\beta)_x$ values result in lower k values. It was observed that the appliance of this restriction affected the resulting fit parameters in a pronounced way, indicating that the RBE_{max}-$(\alpha/\beta)_x$ relationship is different for lower and higher $(\alpha/\beta)_x$ values.

4.2.4 Fitting on balanced databases

As seen from the histograms in Figure 14 and Figure 15 in section 4.1, the proton data is not evenly distributed over the range of LET^* and $(\alpha/\beta)_x$ values, especially the SOBP data. This could lead to poor RBE estimation in the ranges of LET^* and $(\alpha/\beta)_x$ with few data points, leading to inaccurate predictions as the fits might over- or underestimate the RBE_{max} in these regions. Results investigating this effect are present in the following. The resulting mean fit parameters from fitting to a balanced database with respect to $(\alpha/\beta)_x$ values and balanced databases with respect to LET^* values are given in the heatmap in Figure 30(a). The fit parameters obtained with imbalanced databases containing the data points in the same $(\alpha/\beta)_x$ and LET^* ranges are given for comparison in Figure 30(b).

The fit parameters from fits to a balanced database in terms of $(\alpha/\beta)_x$ shows that the large amount of proton data with $(\alpha/\beta)_x$ values in the lower range affects the outcome of the fitting drastically. Compared to the fit parameters from fitting to imbalanced data, the k values obtained with fitting to a balanced database using all proton data, monoenergetic data and SOBP data are approximately 27% larger, 47% larger and 17% smaller, respectively.

Using a balanced database in terms of LET^*, only minor differences were observed for the fit parameters from fitting to all data points and monoenergetic data compared to the k values obtained with imbalanced databases. This is also seen for the fit parameter obtained with balanced SOBP data with $LET^* < 20\ keV/\mu m$, which is almost equal to the corresponding fit parameter obtained with imbalanced data. However, the k value obtained with balanced SOBP data having $LET^* < 21\ keV/\mu m$ is larger than the corresponding fit parameter obtained with imbalanced SOBP data.

From the histograms in Figure 15, we see that the SOBP data is imbalanced in terms of LET^*, while the monoenergetic data have a more uniform distribution of LET^* values below $20\ keV/\mu m$. It was thus expected that the fit parameters from fitting to

balanced SOBP data in terms of LET^* would deviate more from the ones obtained with imbalanced SOBP data, while only small differences were expected in the k values obtained with balanced and imbalanced monoenergetic data. This was observed from the fits to data points with $LET^* < 21\ keV/\mu m$, but the fits to data points with $LET^* < 20\ keV/\mu m$ did not show this. Given that the database consists of 265 SOBP data points and only 148 monoenergetic data points, the imbalanced SOBP data was expected to affect the outcome of fitting to all proton data to some extent. Thus, comparing the fit parameters from fits to balanced proton data in terms of LET^* and the fit parameters obtained with imbalanced proton data, a larger difference than the one observed was expected.

Considering all proton data points and monoenergetic data points, the results from fitting on a balanced database in terms of $(\alpha/\beta)_x$ indicates that an RBE model based on the imbalanced database is likely to underestimate the RBE. However, the results form fitting on a balanced SOBP database in terms of $(\alpha/\beta)_x$ indicates that the RBE might be overestimated if the imbalanced database is used in the modelling. Using a balanced database in terms of LET^* seems less critical for the outcome of the modelling.

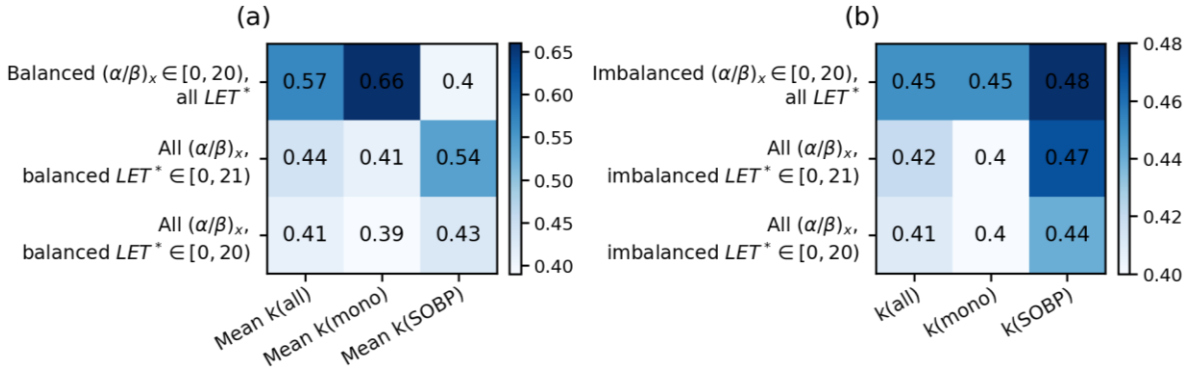

Figure 30: (a): Heatmap showing the fit parameters from fitting on balanced databases using Equation (3.2). The k values in the first row are mean values of the fit parameters from 100 fits on random samples of 20 data points; 5 from each subset of data points with $(\alpha/\beta)_x$ values in the intervals $[0,5)$, $[5,10)$, $[10,15)$ and $[15,20)$. The mean k values in the second row are from 100 fits on random samples of 20 data points; 5 from each subset of data points with LET^* values in the intervals $[0,5)$, $[5,10)$, $[10,15)$ and $[15,21)$, while the mean

fit parameters in the third row are from 165 fits on random samples of 12 data points; 3 from each subset of data points with LET^* values in the intervals $[0,5)$, $[5,10)$, $[10,15)$ and $[15,20)$. Fit parameters from fitting on imbalanced databases with coinciding restrictions on $(\alpha/\beta)_x$ and LET^* are given in (b) for comparison.

Key results

- Using a balanced database with respect to $(\alpha/\beta)_x$ indicated that the large amount of proton data with low $(\alpha/\beta)_x$ heavily affects the outcome of the fitting. In particular, the results imply that an RBE model based on the imbalanced monoenergetic database might underestimate the RBE, while the predicted RBE for irradiation with an SOBP might be slightly overestimated.
- The results from fitting on a balanced database with respect to LET^* showed weak indications that an RBE model based on imbalanced data in terms of LET^* might predict underestimated RBE values for irradiation with an SOBP.

4.3 Non-linear regression

4.3.1 Fitting of proton RBE data without database restrictions

The multiplicative functions in Table 4 were successively inserted into Equation (3.3) to obtain 11 different fitting functions applying a non-linear relationship between RBE_{max} and LET^*. These fitting functions were used to obtain fits for RBE_{max} as a function of LET^* and $(\alpha/\beta)_x$ using all proton data points, only monoenergetic proton data and only SOBP proton data. The resulting fit parameters are given in the heatmaps in Figure 31(a), Figure 32(a) and Figure 33(a), respectively. The fit parameter obtained with a linear RBE_{max}-LET^* relationship are given in the first row in each of the heatmaps, denoted $f_L(LET^*)$. In order to compare the non-linear fits to each other and to the linear fit, the RMSE was calculated for each of the obtained fits. The calculated RMSE values are given in Figure 31(b), Figure 32(b) and Figure 33(b).

The fitting results are further visualized in Figure 34, Figure 35 and Figure 36 by plots showing RBE_{max} as a function of LET^* for 6 different $(\alpha/\beta)_x$ values. In each figure,

the results from fitting with a linear RBE_{max}-LET^* relationship are shown in (a) for comparison (f_L). Proton data points with $(\alpha/\beta)_x$ values inside the intervals [0,3), [3,6), [6,9), [9,12), [12,15) and [15,20) are plotted in different colours. The non-linear fits were used to plot RBE_{max} as a function of LET^* using the $(\alpha/\beta)_x$ values in the centre of each $(\alpha/\beta)_x$ interval, that is 1.5 Gy, 4.5 Gy, 7.5 Gy, 10.5 Gy, 13.5 Gy and 17.5 Gy. The curves are plotted in the same colour as the data points in the corresponding interval of $(\alpha/\beta)_x$ values.

As seen in Figure 31(b), most of the RMSE values for fits to all proton data are very similar to each other, and the lowest value is shared by six of the fits. Most RMSE values for the fits to monoenergetic data and SOBP data are also similar to each other, as shown in Figure 32(b) and Figure 33(b) respectively. However, it is observed from these heatmaps that the fit obtained with f_{LQCq} have lowest RMSE value both for the monoenergetic data and the SOBP data. This fitting function was also used to obtain one of the six fits to all data that had lowest RMSE. Therefore, an overall indication that this function yields better fits to the proton data is observed, although the evidence supporting this are very weak. This is in agreement with the conclusions from Rørviks investigation of the RBE-LET relationship [10]. As seen in Figure 34(d) and Figure 35(d), the fits to all data and monoenergetic data obtained with f_{LQCq} show similar qualities when the RBE_{max} is plotted as a function of LET^*, although the latter has more distinct fluctuations in the slope which is most clearly seen from the curve plotted with $(\alpha/\beta)_x = 1.5\ Gy$ in Figure 35(d). From Figure 36(d) it is observed that the curves for RBE_{max} as a function of LET^* from the corresponding fit to SOBP data has a somewhat different shape, as the RBE_{max} is initially increasing with increasing LET^* and starts to decrease with increasing LET^* at $LET^* \approx 22\ keV/\mu m$. The observed differences are also reflected in the fit parameters from fitting with f_{LQCq}, as the ones obtained with all data and monoenergetic data in Figure 31(a) and Figure 32(a) resemble each other, while the fit parameters obtained with SOBP data in Figure 33(a) show larger deviations from these.

It was observed that four of the fitting functions that resulted in lowest RMSE values for all proton data also yielded low RMSE values when they were used to obtain fits to monoenergetic data and SOBP data. This was the case for f_{LQ}, f_{LQC}, f_{LE} and f_{LQE}, which resulted in fits with second lowest RMSE values both in Figure 32(b) and Figure 33(b). These fitting functions are therefore indicated to be approximately equally good as the 4th degree polynomial function. Given the marginally small deviations in RMSE values, the results can be interpreted to weakly imply that these five functions might generally be better suited to model the data than the other considered functions. As these five non-linear functions overall yielded lower RMSE values than the linear fits, it is implied that a non-linear RBE_{max}-LET^* relationship might be favoured.

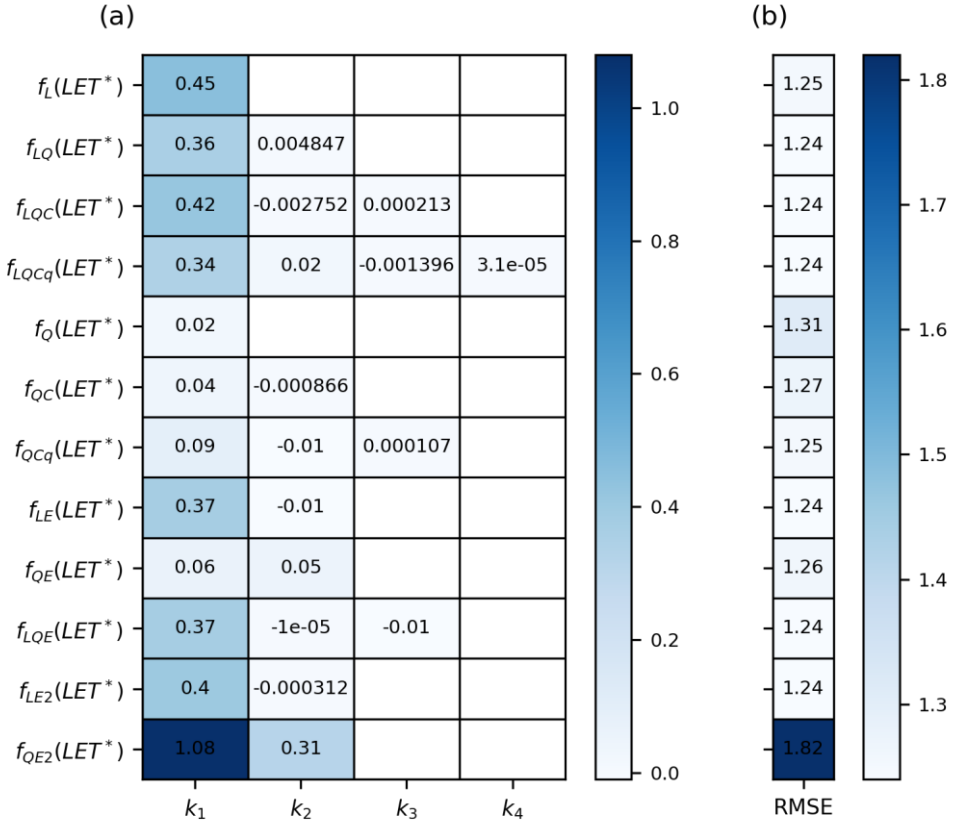

Figure 31: (a): Fit parameters from fitting on all proton data points, using fitting functions resulting form insertion of different multiplicative functions into Equation (3.3) to apply different non-linear relationships between RBE_{max} and LET^*. The previously obtained fit parameter from fitting on all proton data applying a linear RBE_{max}-LET^* relationship is given in the first row in the heatmap. The corresponding RMSEs for each of the fits are given in (b).

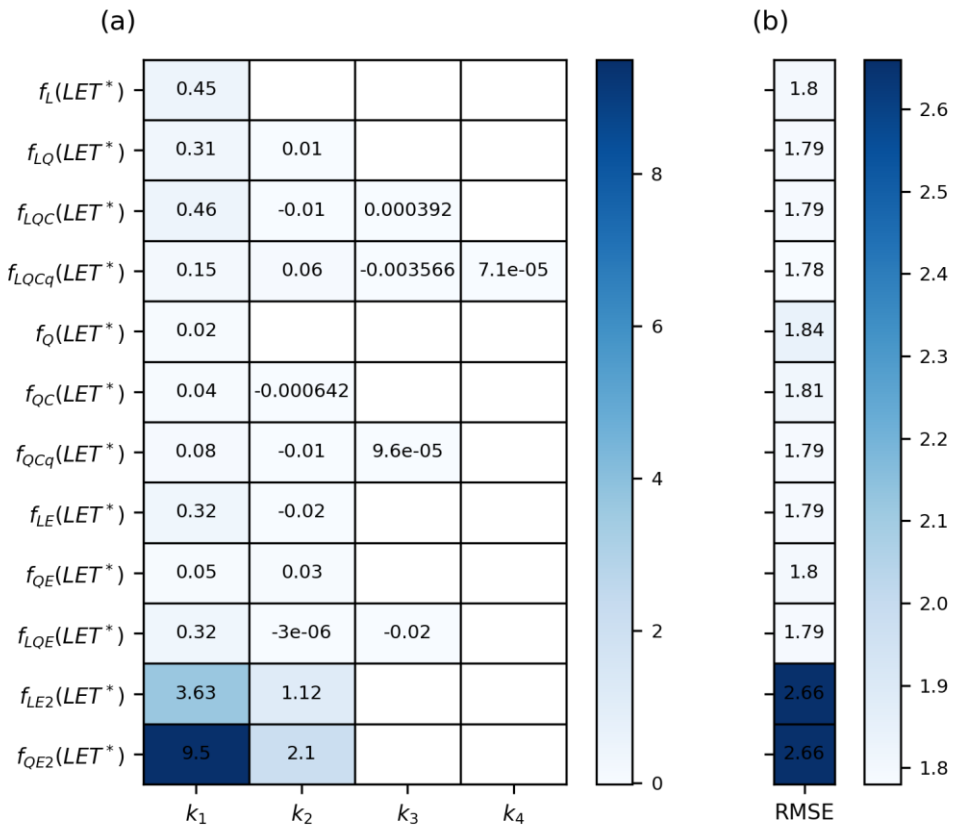

Figure 32: (a): Fit parameters from fitting on monoenergetic proton data, using fitting functions resulting form insertion of different multiplicative functions into Equation (3.3) to apply different non-linear relationships between RBE_{max} and LET^*. The previously obtained fit parameter from fitting on monoenergetic data applying a linear RBE_{max}-LET^* relationship is given in the first row in the heatmap. The corresponding RMSEs for each of the fits are given in (b).

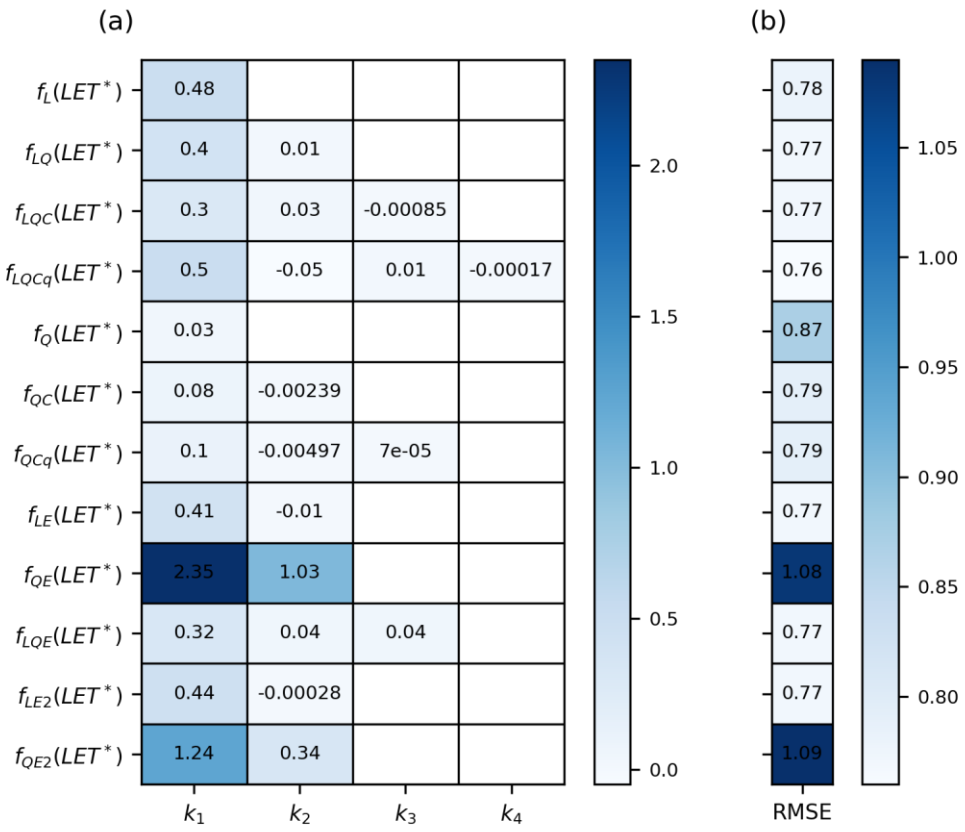

Figure 33: (a): Fit parameters from fitting on SOBP proton data, using fitting functions resulting form insertion of different multiplicative functions into Equation (3.3) to apply different non-linear relationships between RBE_{max} and LET^*. The previously obtained fit parameter from fitting on SOBP data applying a linear RBE_{max}-LET^* relationship is given in the first row in the heatmap. The corresponding RMSEs for each of the fits are given in (b).

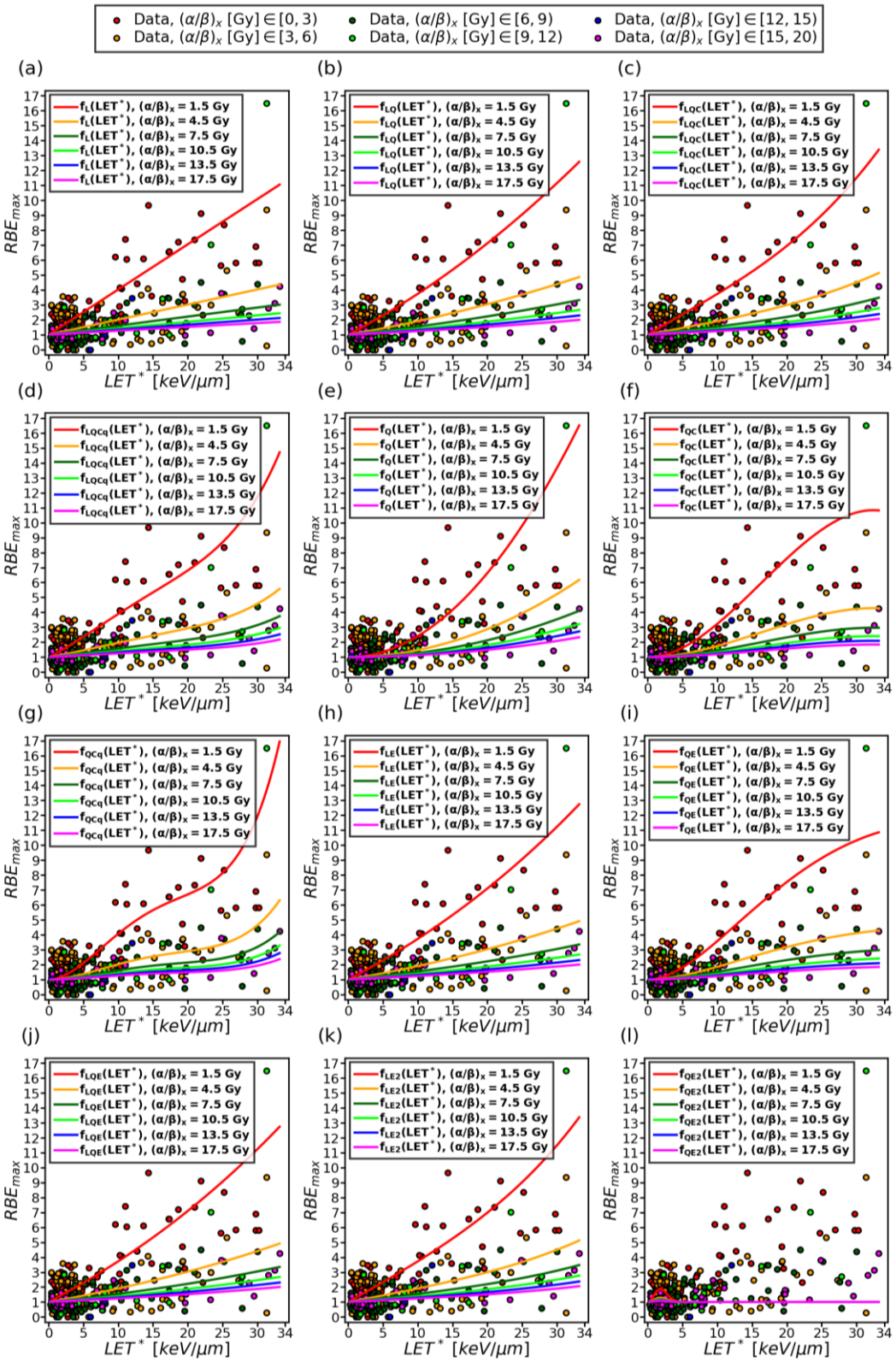

Figure 34: (b)-(l): RBE_{max} plotted as a function of LET^* for specific values of $(\alpha/\beta)_x$ using the results from fitting on all proton data points applying different non-linear RBE_{max}-LET^* relationships. The result from appliance of a linear relationship between RBE_{max} and LET^* is shown in (a).

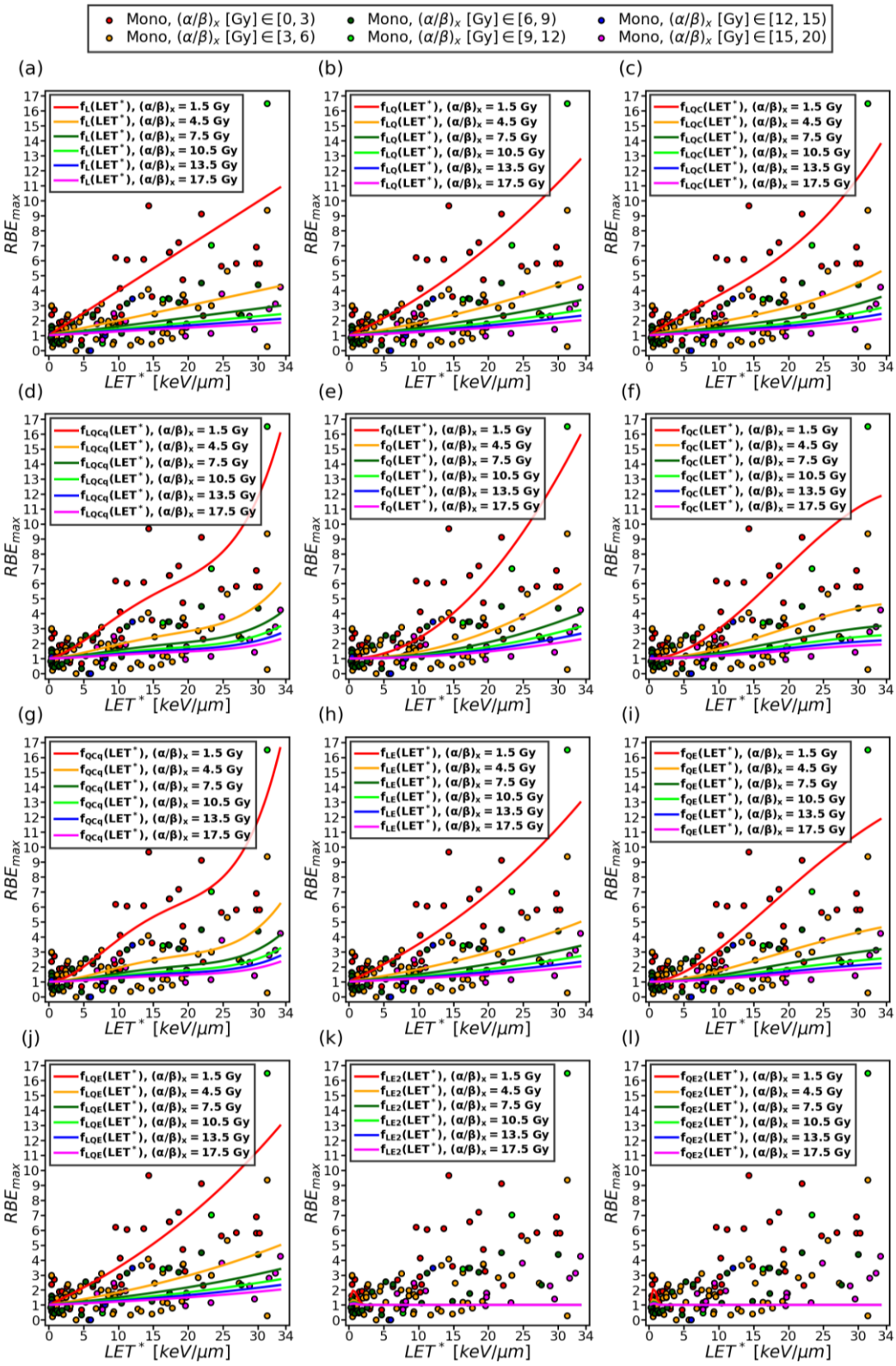

Figure 35: (b)-(l): RBE_{max} plotted as a function of LET^* for specific values of $(\alpha/\beta)_x$ using the results from fitting on monoenergetic proton data points applying different non-linear RBE_{max}-LET^* relationships. The result from appliance of a linear relationship between RBE_{max} and LET^* is shown in (a).

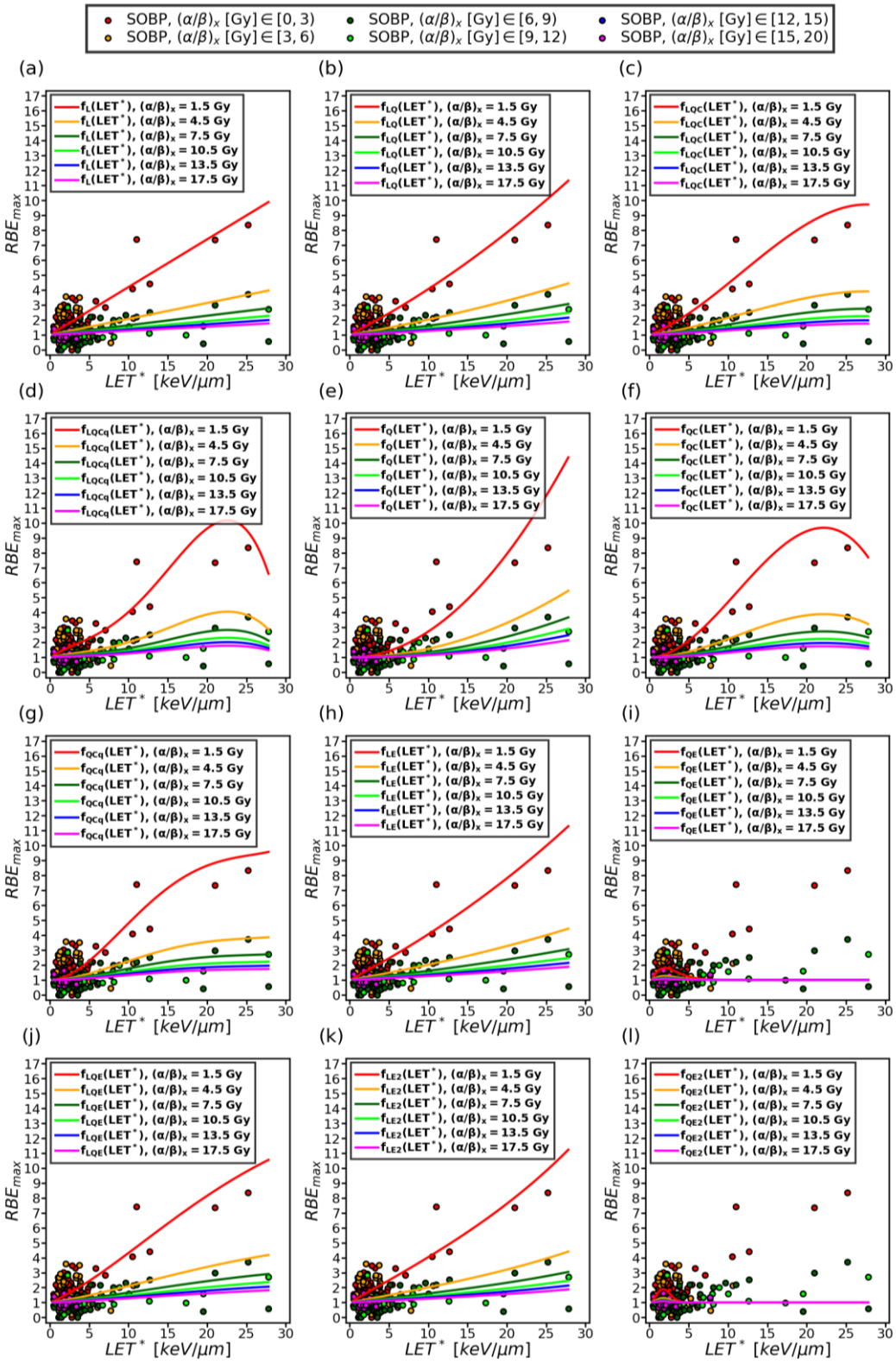

Figure 36: (b)-(l): RBE_{max} plotted as a function of LET^* for specific values of $(\alpha/\beta)_x$ using the results from fitting on SOBP proton data points applying different non-linear RBE_{max}-LET^* relationships. The result from appliance of a linear relationship between RBE_{max} and LET^* is shown in (a).

Key results

- Calculated RMSE values were very similar for most of the fits obtained with non-linear fitting functions. However, an overall slightly better performance was indicated for the fitting functions with the multiplicative functions f_{LQ}, f_{LQC}, f_{LQCq}, f_{LE} and f_{LQE}, indicating that non-linear fits may give better RBE estimation compared to linear fits.
- The fitting function f_{LQCq} gave the overall best performance I terms of RMSE. This is in agreement with the conclusions from Rørvik et al 2017 [10], using a monoenergetic database.

4.3.2 Fitting on restricted databases

The fitting procedure applied in section 4.3.1 was repeated for restricted databases in terms of reduced range of included $(\alpha/\beta)_x$ and LET^* values. The fit parameters in the heatmaps in Figure 37(a), Figure 38(a) and Figure 39(a) are obtained with the data points that was considered to be of most clinical relevance. According to Paganetti 2014 [4], LET_d values above $20\ keV/\mu m$ are rarely present in clinical proton therapy. Therefore, only data points with $LET^* < 20\ keV/\mu m$ were included in the fitting. The restriction on $(\alpha/\beta)_x$ was inspired by the one used to develop the Rørvik models. Only data points with $(\alpha/\beta)_x$ values indicating cells with normal repair mechanisms were included, i.e. $(\alpha/\beta)_x < 25\ Gy$. In Figure 37(b), Figure 38(b) and Figure 39(b), the calculated RMSE values for each of the obtained fits are given.

As seen in Figure 37(b) and Figure 39(b), the fits to all data and SOBP data obtained with the 4th degree polynomial fitting function (f_{LQCq}) have lowest RMSE values. Although the difference is not large for the fits to all data, the RMSE for the fit to SOBP data obtained with this function show a little larger deviation from the other RMSE values in Figure 39(b). The RMSE values for the fits to monoenergetic data weakly implies that the fits obtained with f_{LQC} and f_{LQCq} are better than the other fits. The fitting results on the restricted database therefore implies that the 4th degree fitting

function might yield the best fits to the proton data. This is the same indication as was observed from the non-linear fitting on the unrestricted database. However, it seems a little stronger in this case, as the RMSE values for the fits obtained with this fitting function generally stand more out from the other RMSEs in Figure 37(b), Figure 38(b) and Figure 39(b), compared to the RMSE values for fitting on the unrestricted database in Figure 31(b), Figure 32(b) and Figure 33(b). It was also observed that the fitting functions f_{LQ}, f_{LQC}, f_{LE} and f_{LQE} generally resulted in low RMSE values both for the fits to all data points and the fits to monoenergetic and SOBP data. This was also seen in the results form non-linear fitting on the unrestricted database, and again indicates that these functions are good candidates for non-linear modelling of the data.

In Figure 40, the fits on the restricted database obtained with the 4th degree polynomial fitting function are used to plot the RBE_{max}-LET^* relationships for the same $(\alpha/\beta)_x$ values as used earlier to visualize the fitting result on the unrestricted database. From the figure it is observed that all three fits have similar shapes, as the RBE_{max} increase more rapidly with increasing LET^* up to $LET^* \sim 16-18\ keV/\mu m$, where the curves have an extreme and the RBE_{max} starts decreasing with increasing LET^*. The same shape was observed for the RBE_{max}-LET^* relationship plotted with the fit to the unrestricted SOBP data obtained with the same fitting function shown in Figure 36(d). However, this is not observed for the corresponding fits to all proton data and monoenergetic data in Figure 34(d) and Figure 35(d).

The fit to SOBP data in Figure 40(c) is based on very few data points with LET^* values in the upper part of the considered range and a large amount of low LET^* data. The fit should therefore not be trusted, as it is likely to be considerably affected by the imbalanced distribution of LET^* values. The fact that the SOBP data is not evenly distributed over the range of LET^* values also affects the fit to all data points in Figure 40(a), as seen from comparison with the fit to monoenergetic data, which has a much more uniform distribution of LET^* values.

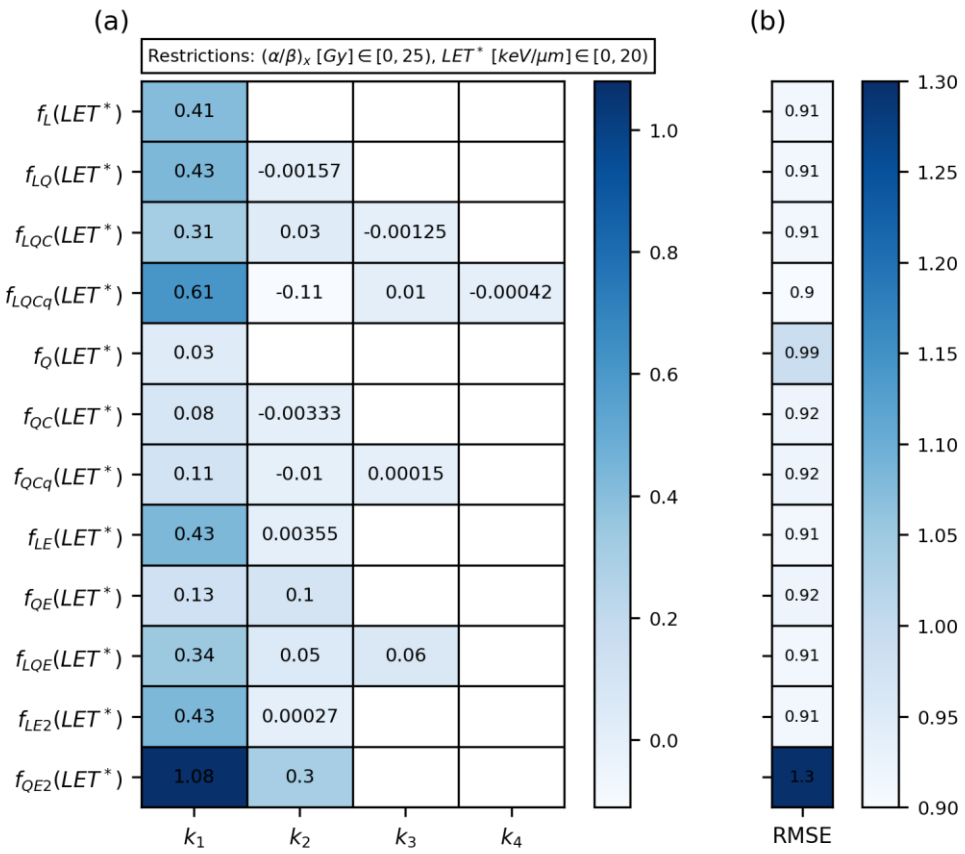

Figure 37: (a): Fit parameters from fitting on all proton data points in a restricted database including the most clinical relevant data. The fits are obtained with fitting functions resulting form insertion of different multiplicative functions into Equation (3.3) to apply different non-linear relationships between RBE_{max} and LET^*. A linear relationship is also applied, and the resulting fit parameters are given in the first row in the heatmap. The corresponding RMSEs for each of the fits are given in (b).

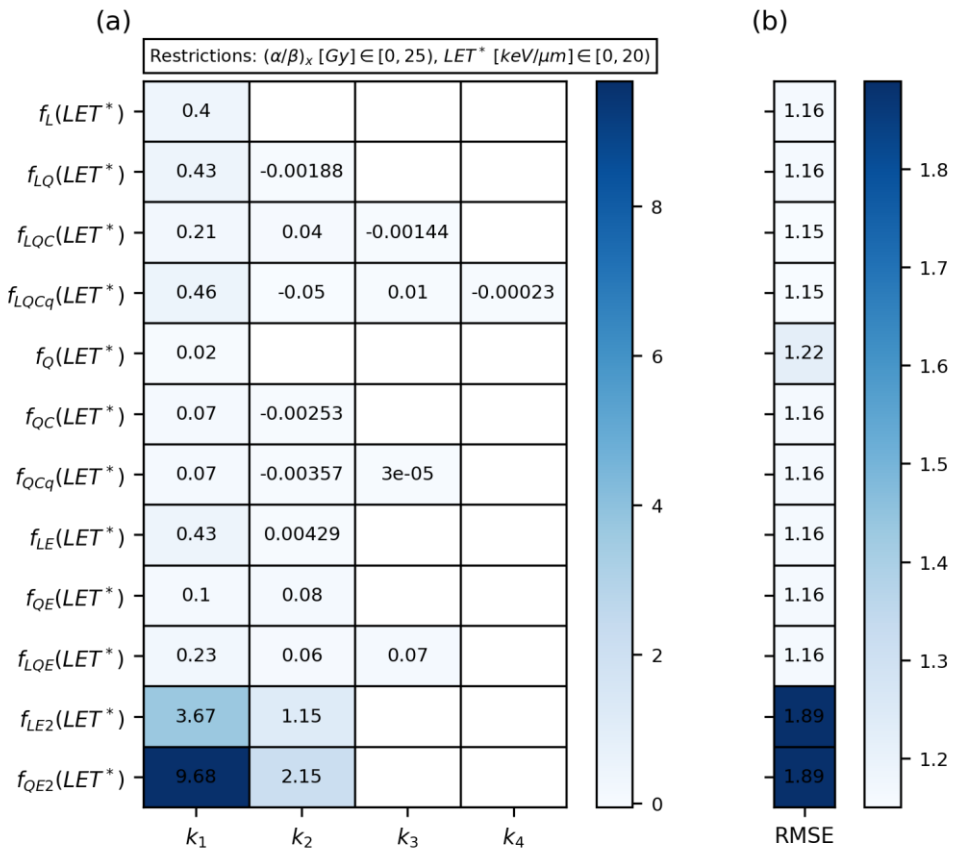

Figure 38: (a): Fit parameters from fitting on monoenergetic data points in a restricted database including the most clinical relevant data. The fits are obtained with fitting functions resulting form insertion of different multiplicative functions into Equation (3.3) to apply different non-linear relationships between RBE_{max} and LET^*. A linear relationship is also applied, and the resulting fit parameters are given in the first row in the heatmap. The corresponding RMSEs for each of the fits are given in (b).

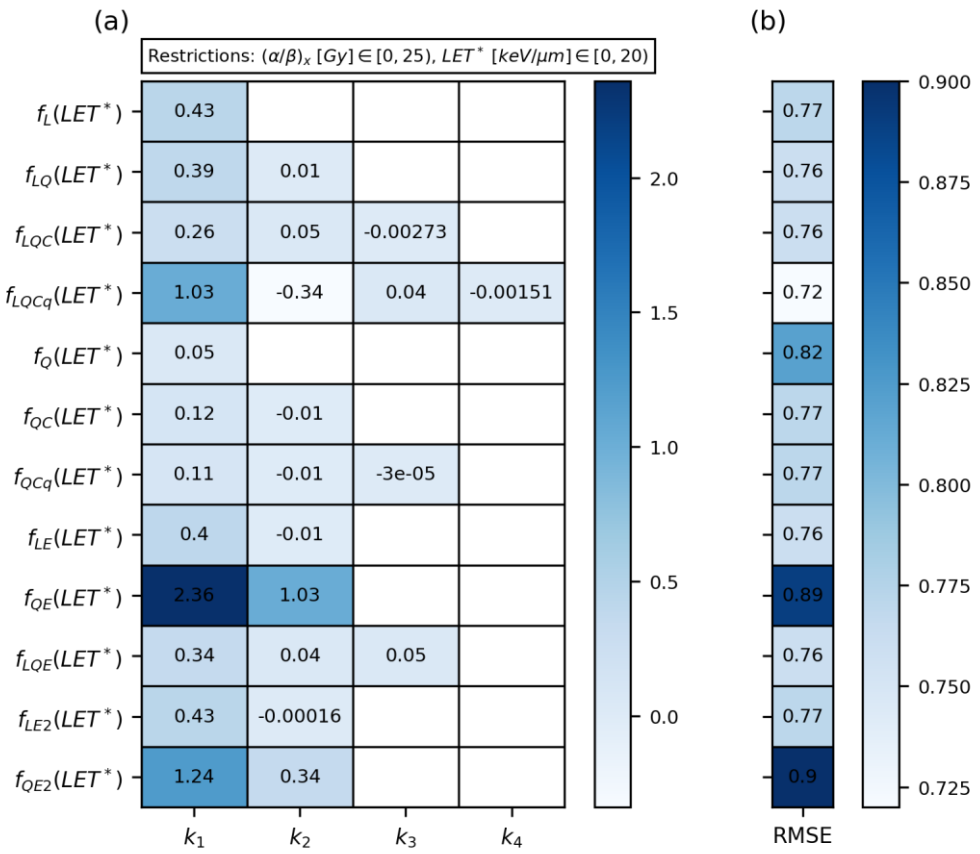

Figure 39: (a): Fit parameters from fitting on SOBP data points in a restricted database including the most clinical relevant data. The fits are obtained with fitting functions resulting form insertion of different multiplicative functions into Equation (3.3) to apply different non-linear relationships between RBE_{max} and LET^*. A linear relationship is also applied, and the resulting fit parameters are given in the first row in the heatmap. The corresponding RMSEs for each of the fits are given in (b).

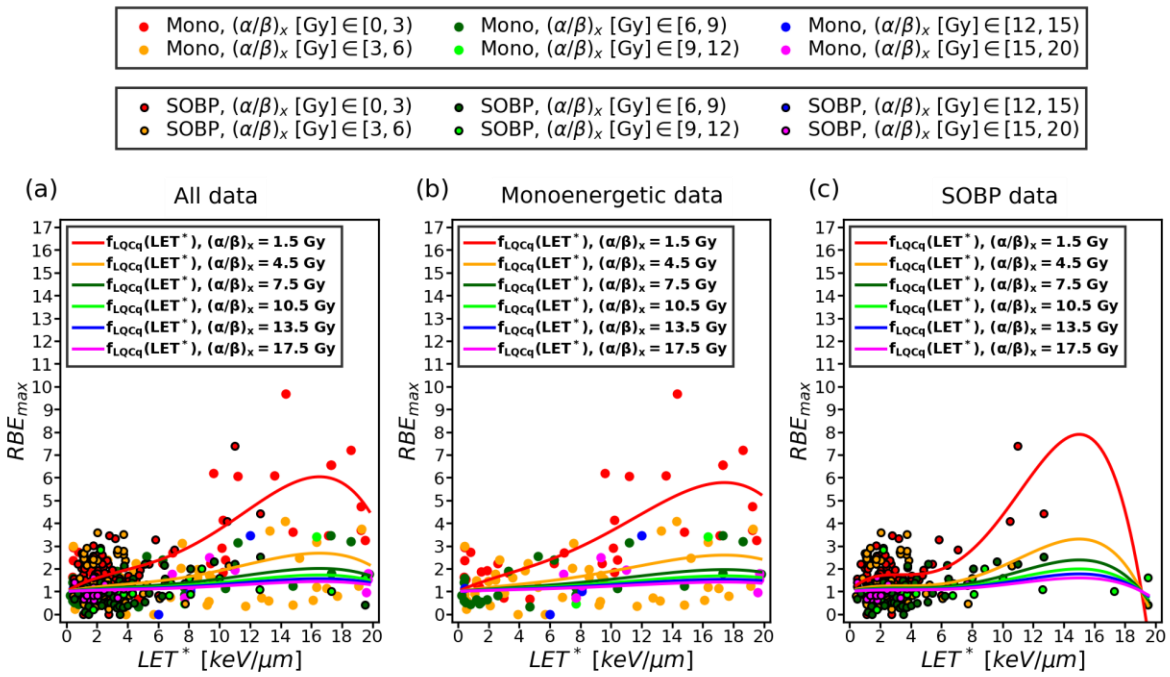

Figure 40: RBE_{max} plotted as a function of LET^* for specific values of $(\alpha/\beta)_x$ using the results from fitting on a restricted database applying the 4th degree polynomial fitting function. In (a), the fit to all proton data in the database restricted by $(\alpha/\beta)_x < 25\ Gy$ and $LET^* < 20\ keV/\mu m$ is used, while the fits to monoenergetic and SOBP data are used to plot the curves in (b) and (c), respectively.

Non-linear fitting was also performed on a database subject to a restriction on the lower limit of included LET^* values. Data points with $LET^* < 5\ keV/\mu m$ were excluded from the database, while the same restrictions as above was used on the upper limits of included $(\alpha/\beta)_x$ and LET^* values. Applying quadratic-exponential fitting function, f_{QE}, the curve_fit() function raised an error that the least-squares minimization failed. The appliance of the fitting functions f_{LE2} and f_{QE2} also resulted in errors, as the curve_fit() function was unable to estimate the covariance of the parameters. These three fitting functions were therefore excluded from the fitting procedure. The fit parameters and the calculated RMSE values for each of the non-linear fits to all proton data in the restricted database, only the monoenergetic data and only the SOBP data are given in Figure 41, Figure 42 and Figure 43, respectively.

As seen in the figures, the deviations in RMSE values are very small both for the fits to all data and the fits to monoenergetic and SOBP data. In Figure 41(b) the lowest RMSE value from fits to all data is shared by two of the fits, while five of the fits to monoenergetic data share the lowest RMSE in Figure 42(b), and three of the fits to SOBP data share the lowest RMSE value in Figure 43(b). However, the only fit that is among the ones having lowest RMSE in all three cases is from fitting with the 4th degree polynomial fitting function, which is another weak indication that this might be the best fitting function for the data. The fits obtained with f_{LQC}, f_{QC}, f_{QCq} and f_{LQE} had low RMSE values both for all proton data points and monoenergetic and SOBP data, indicating that these functions are reasonable candidates to be considered for non-linear fitting of proton RBE data.

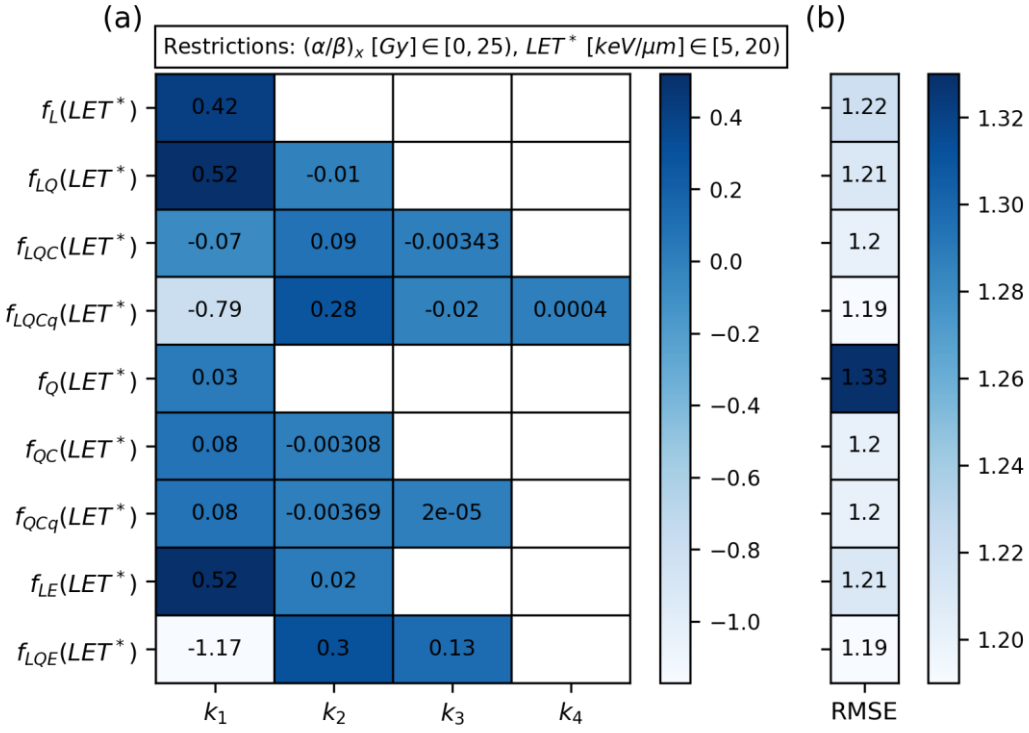

Figure 41: (a): Fit parameters from fitting on all proton data points in a restricted database including data with $(\alpha/\beta)_x < 25\,Gy$ and LET^* values in the interval [5,20) (given in $keV/\mu m$). The fits are obtained with fitting functions resulting form insertion of different multiplicative functions into Equation (3.3) to apply different non-linear relationships between RBE_{max} and LET^*. A linear relationship is also applied, and the resulting fit parameters are given in the first row in the heatmap. The corresponding RMSEs for each of the fits are given in (b).

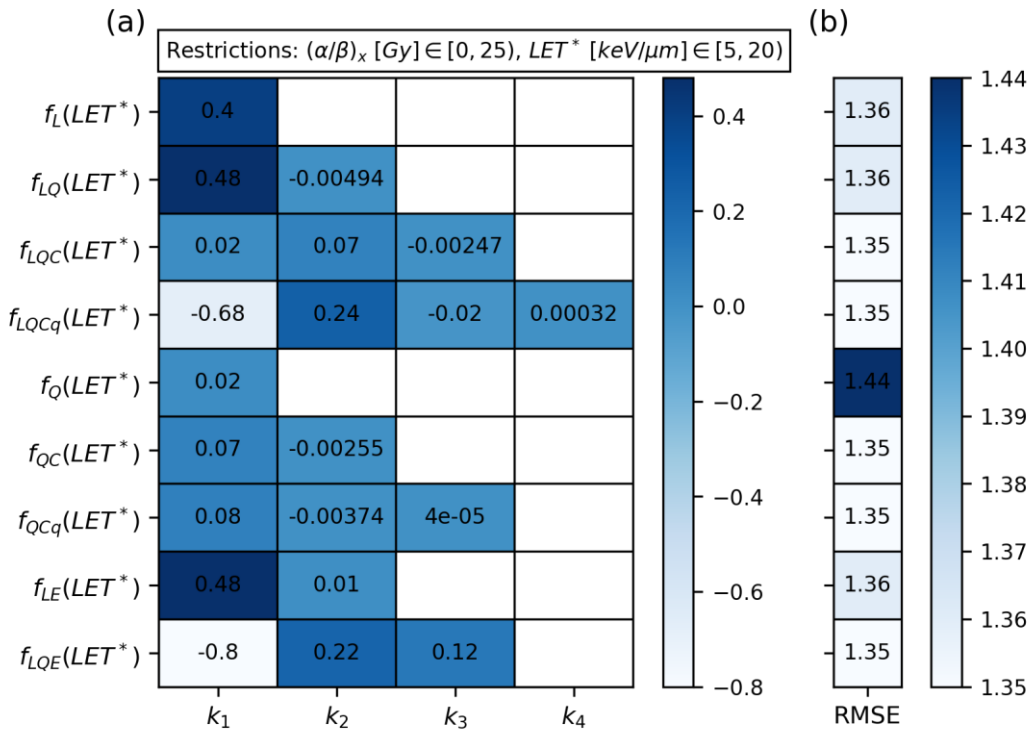

Figure 42: (a): Fit parameters from fitting on the monoenergetic data points in a restricted database including data with $(\alpha/\beta)_x < 25\ Gy$ and LET^* values in the interval $[5,20)$ (given in $keV/\mu m$). The fits are obtained with fitting functions resulting form insertion of different multiplicative functions into Equation (3.3) to apply different non-linear relationships between RBE_{max} and LET^*. A linear relationship is also applied, and the resulting fit parameters are given in the first row in the heatmap. The corresponding RMSEs for each of the fits are given in (b).

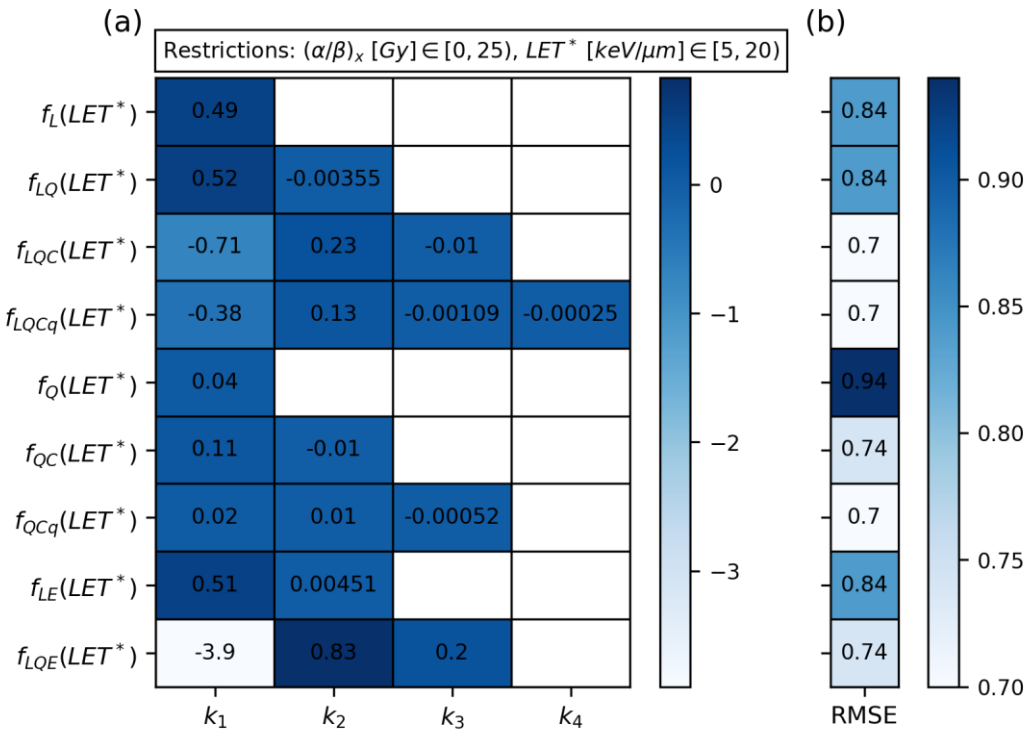

Figure 43: (a): Fit parameters from fitting on the SOBP data points in a restricted database including data with $(\alpha/\beta)_x < 25\,Gy$ and LET^* values in the interval [5,20) (given in $keV/\mu m$). The fits are obtained with fitting functions resulting form insertion of different multiplicative functions into Equation (3.3) to apply different non-linear relationships between RBE_{max} and LET^*. A linear relationship is also applied, and the resulting fit parameters are given in the first row in the heatmap. The corresponding RMSEs for each of the fits are given in (b).

Key results

- Fitting on restricted databases show the same results as fitting on the unrestricted database, indicating that the appliance of f_{LQCq} yielded the best fits when comparing the calculated RMSE values. Several non-linear fitting functions were indicated to be better suited to model the data than the linear fitting function, according to the RMSE values.
- The RMSE values for the fits to proton data with $(\alpha/\beta)_x < 25\,Gy$ and $LET^* < 20\,keV/\mu m$ indicated that f_{LQ}, f_{LQC}, f_{LE} and f_{LQE} are also good candidates to model the proton data.
- Applying an additional restriction on the lower limit of included LET^* values, excluding data points with $LET^* < 5\,keV/\mu m$, the RMSE values of the

resulting fits indicated that f_{LQC}, f_{QC}, f_{QCq} and f_{LQE} are the most promising model candidates, in addition to f_{LQCq}.

4.3.3 Fitting on balanced databases

The polynomial functions in the first seven rows in Table 4 was inserted into Equation (3.3) and the resulting fitting functions was used to obtain fits on balanced databases with respect to $(\alpha/\beta)_x$ and LET^*. The fitting was carried out using the same procedure and intervals as in section 3.2.2 to obtain mean fit parameters from 100 fits on random samples.

The resulting mean k values from appliance of this procedure to all data points with $(\alpha/\beta)_x < 20\ Gy$ are given in Figure 44(a), while only the monoenergetic data and only the SOBP data was used to obtain the mean k values in Figure 45(a) and Figure 46(a), respectively. The RMSE values of the resulting fits are given in (b) in each figure. The fits obtained with a corresponding imbalanced database are given in (c) in each figure, and the RMSE values of these fits are given in (d) in each figure.

It is clear from the figures that the appliance of a balanced database with respect to $(\alpha/\beta)_x$ largely affects the fitting outcome, as most of the fit parameters show relatively large deviations from the corresponding ones obtained with an imbalanced database. The RMSE values indicate that the linear fitting function is best suited for fitting on the balanced database both when all proton data are used in the fitting and when monoenergetic and SOBP data are treated separately. Fitting on the balanced database using the 4th degree polynomial fitting function, which was previously indicated to be the most promising candidate for non-linear fitting, result in fits having high RMSE values. The large deviations between the results from fitting on balanced- and imbalanced databases with respect to $(\alpha/\beta)_x$ indicates that the larger portion of data points with low $(\alpha/\beta)_x$ values have a pronounced effect on the outcome of non-linear fitting. However, when performing non-linear fitting, it might not be necessary to implement the procedure of balancing the data, as the non-linear fitting functions are

more suited to handle variable amounts of data points in different regions. With the applied procedure to obtain balanced databases, the data points with $(\alpha/\beta)_x$ values in the interval $[15,20)$ will be more weighted in the fitting, as there are fewer data points in this interval. Therefore, balancing the database might actually have a bad impact on the fitting outcome, as some datapoints can be assigned unreasonable large weights. The large deviations observed might thus be connected to this effect, and are not necessarily caused by the larger portion of data points with low $(\alpha/\beta)_x$ values alone.

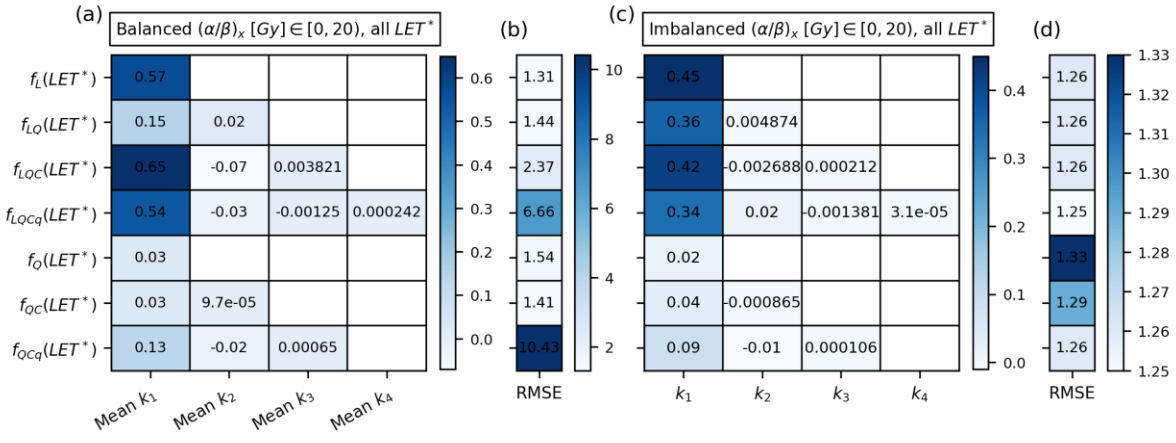

Figure 44: (a): Fit parameters from non-linear fitting on all proton data points in a balanced database with respect to $(\alpha/\beta)_x$. The corresponding RMSE values for each resulting fit is given in (b). In (c), fit parameters from fitting on all proton data points in the corresponding imbalanced database are given for comparison, and the RMSE values for these fits are given in (d).

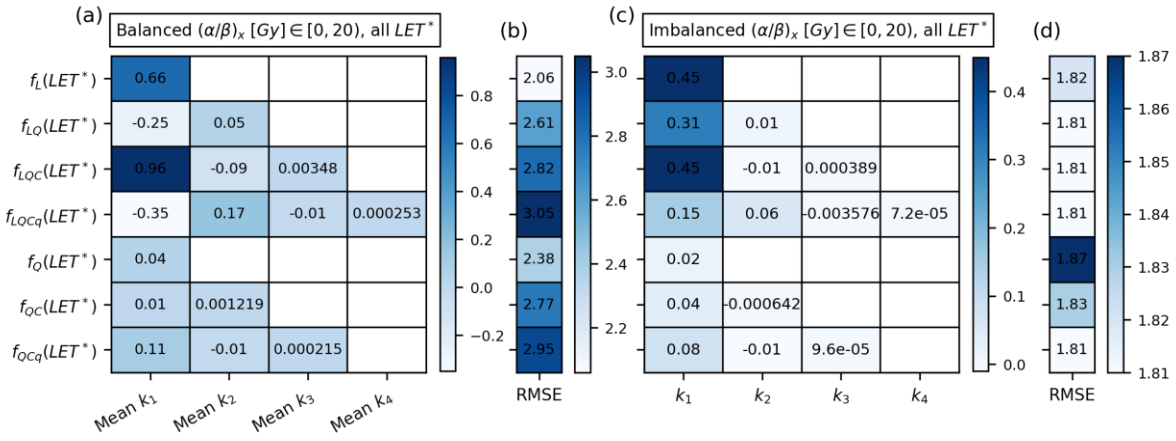

Figure 45: (a): Fit parameters from non-linear fitting on monoenergetic proton data points in a balanced database with respect to $(\alpha/\beta)_x$. The corresponding RMSE values for each resulting fit is given in (b). In (c), fit parameters from

fitting on monoenergetic proton data points in the corresponding imbalanced database are given for comparison, and the RMSE values for these fits are given in (d).

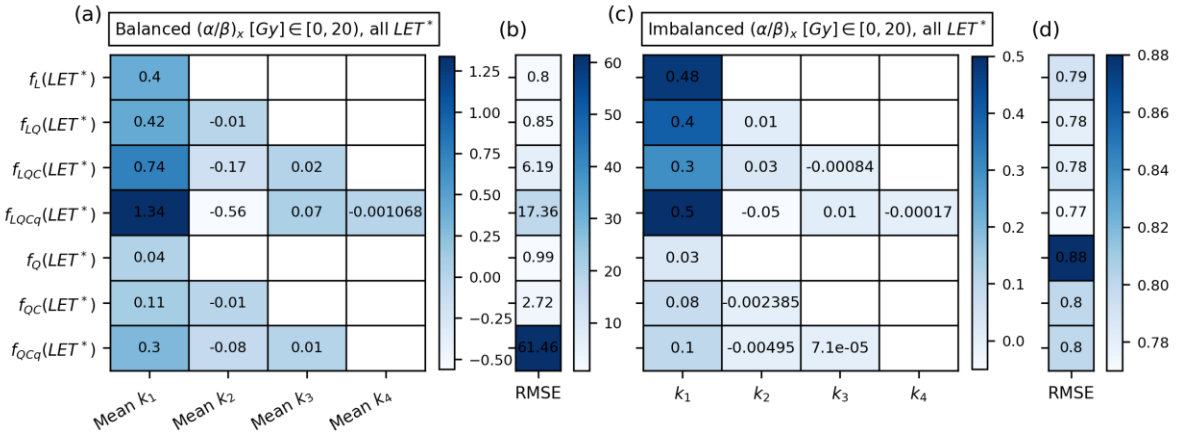

Figure 46: (a): Fit parameters from non-linear fitting on SOBP proton data points in a balanced database with respect to $(\alpha/\beta)_x$. The corresponding RMSE values for each resulting fit is given in (b). In (c), fit parameters from fitting on SOBP proton data points in the corresponding imbalanced database are given for comparison, and the RMSE values for these fits are given in (d).

The fit parameters from non-linear fitting on balanced databases with respect to LET^* using all proton data with $LET^* < 21\ keV/\mu m$ are given in Figure 47(a), and the fit parameters obtained using monoenergetic data and SOBP data are given in Figure 48(a) and Figure 49(a), respectively. To get an estimate of the goodness of the fits, RMSE values were calculated and are given in (b) in each figure. The fit parameters from fitting on the corresponding imbalanced database, and the RMSEs of these fits are given in respectively (c) and (d) in each figure.

From Figure 47(a) Figure 48(a) and Figure 49(a), it is observed that the fit parameters from non-linear fitting on the balanced database are similar or equal to the corresponding fit parameters obtained with an imbalanced database. Consequently, very small deviations are seen between the RMSE values for corresponding fits to the balanced and imbalanced databases. This indicates that the outcome of non-linear fitting on the proton data is insignificantly influenced by the fact that the SOBP proton data is not evenly distributed over the range of LET^* values. However, from Figure 49 it is observed that the fit to balanced SOBP data obtained with f_{LQCq} has a relatively

large RMSE compared to the corresponding fit to imbalanced SOBP data. This is worth noting, as it does not support the indications in previous results from non-linear fitting that f_{LQCq} might be the best candidate to model the data.

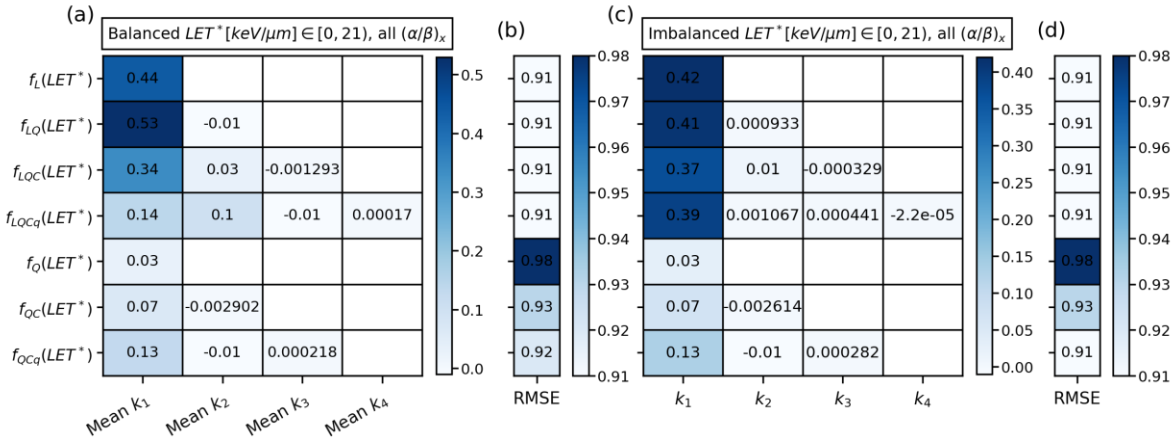

Figure 47: (a): Fit parameters from non-linear fitting on all proton data points in a balanced database with respect to LET^*. The corresponding RMSE values for each resulting fit is given in (b). In (c), fit parameters from fitting on all proton data points in the corresponding imbalanced database are given for comparison, and the RMSE values for these fits are given in (d).

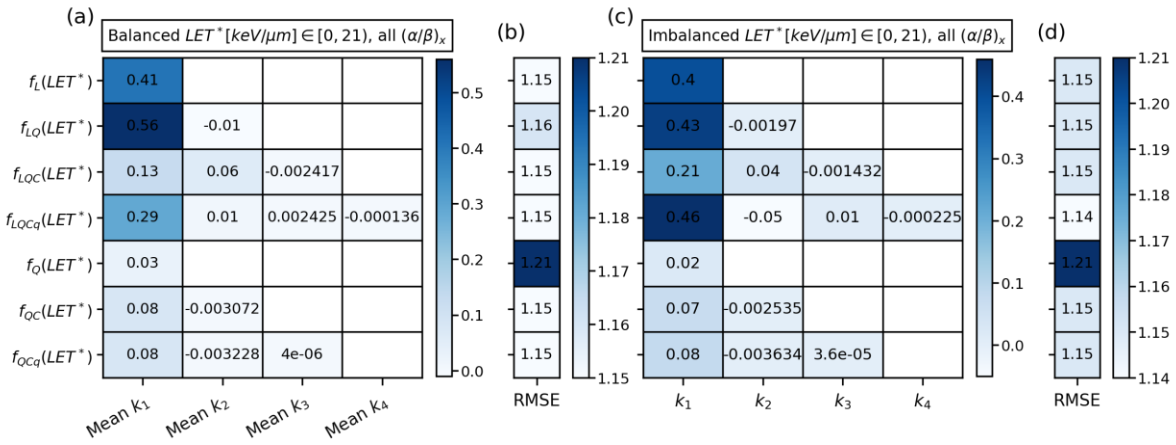

Figure 48: (a): Fit parameters from non-linear fitting on monoenergetic proton data points in a balanced database with respect to LET^*. The corresponding RMSE values for each resulting fit is given in (b). In (c), fit parameters from fitting on monoenergetic proton data points in the corresponding imbalanced database are given for comparison, and the RMSE values for these fits are given in (d).

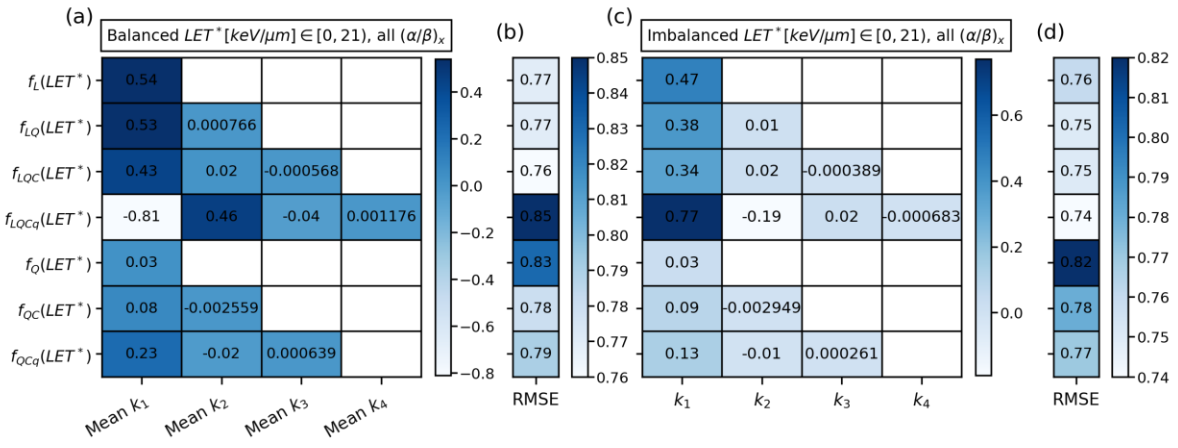

Figure 49: (a): Fit parameters from non-linear fitting on SOBP proton data points in a balanced database with respect to LET^*. The corresponding RMSE values for each resulting fit is given in (b). In (c), fit parameters from fitting on SOBP proton data points in the corresponding imbalanced database are given for comparison, and the RMSE values for these fits are given in (d).

Key results

- The fitting outcome is heavily affected when a balanced database with respect to $(\alpha/\beta)_x$ is used in the fitting, and the RMSE values of the fits indicate that the linear fitting function result in most precise modelling of the proton data.
- The outcome form fitting on a balanced database with respect to LET^* show only small deviations from the fits to the corresponding imbalanced database, and a slightly better performance was implied for non-linear fitting compared to the linear fits. Consequently, only minor differences are observed in the calculated RMSE values as well. However, an exception was observed for fit to balanced SOBP data obtained with f_{LQCq}, which has an increased RMSE.

4.3.4 Comparison with the Rørvik weighted model

The five fitting functions with f_{LQ}, f_{LQC}, f_{LQCq}, f_{LE} and f_{LQE} were indicated to yield best performance in the non-linear fitting both on the unrestricted database in section 4.3.1 and on the restricted database containing data points with $(\alpha/\beta)_x < 25\ Gy$ and $LET^* < 20\ keV/\mu m$ in section 4.3.2. These functions were used to obtain fits to proton data with $(\alpha/\beta)_x < 25\ Gy$, including all values of LET^*. The results were compared

with the RBE_{max} function of the Rørvik model obtained with a weighted dataset (RORW), which is derived using a 4th degree polynomial fitting function. In Figure 50, the RBE_{max}-LET^* relationships of the fits are plotted for different $(\alpha/\beta)_x$ values together with the RBE_{max}-LET^* relationship according to the RORW model plotted for the same values of $(\alpha/\beta)_x$.

It is clear from the figure that the RORW model applies the same assumption for the RBE_{max}-$(\alpha/\beta)_x$ relationship as the fitted curves. While the RORW curves starts to decrease with increasing LET^* values at $\sim 30\ keV/\mu m$, this feature is not observed for any of the fitted curves. As seen in Figure 50, the RBE_{max}-LET^* relationships of the fits obtained with f_{LQ}, f_{LQC}, f_{LE} and f_{LQE} have very similar characteristics, showing that the RBE_{max} increases with increasing LET^* with a gradually steeper slope for higher LET^* values. The curve from the fit obtained with f_{LQCq} has a slightly different shape with more distinct fluctuations in the slope. Although the RORW curve also show this characteristic, the fluctuations do not coincide for these two curves. From Figure 50(c), it is observed that in the ranges of let where the RORW curve is steepest, the slope of the f_{LQCq} curve is at its lowest. The deviation in the shapes of these curves might stem from differences in the applied fitting procedure, as the RORW model is derived from a weighted regression, taking the experimental uncertainties of the data into account.

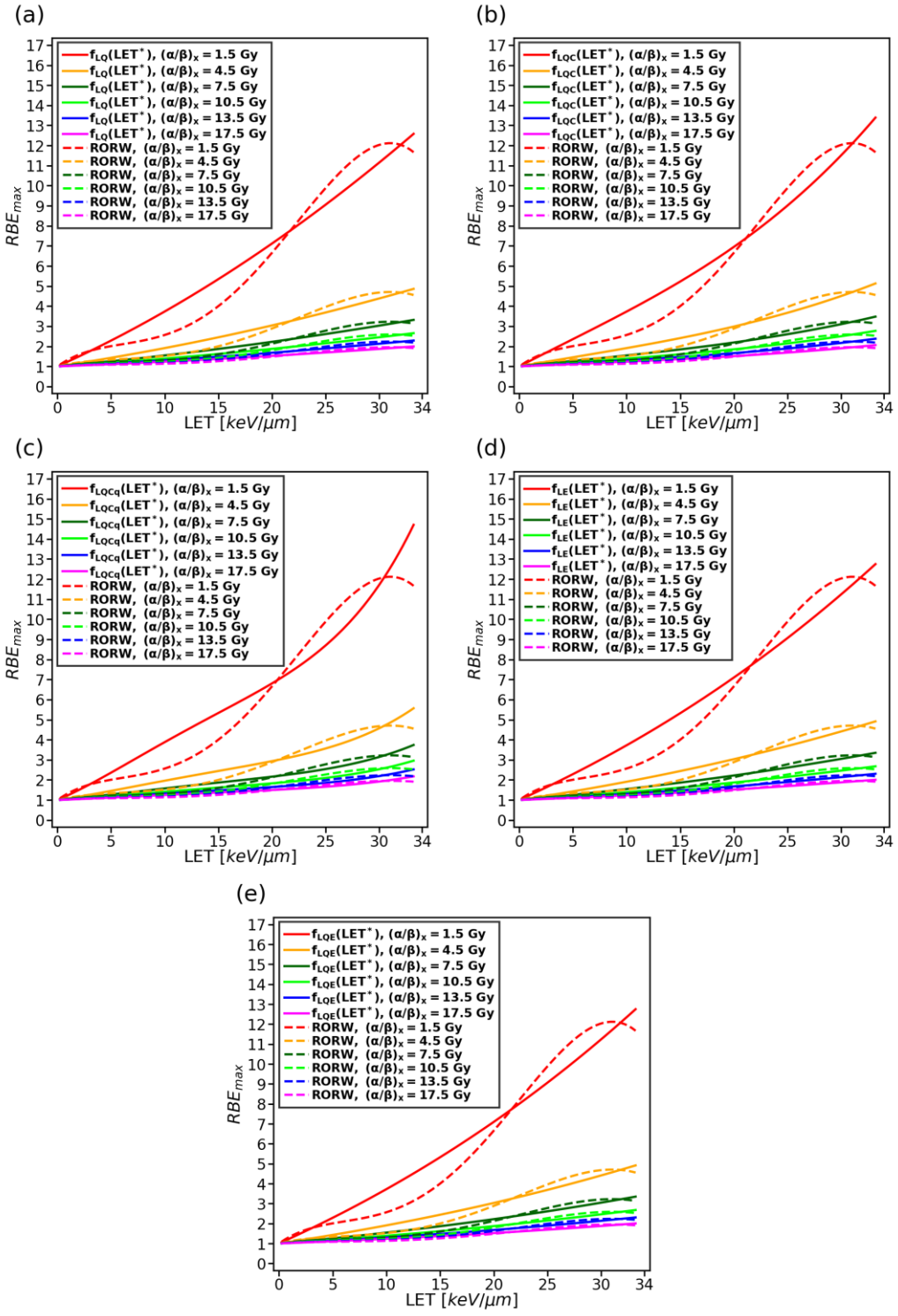

Figure 50: RBE_{max} plotted as a function of LET^* for six different $(\alpha/\beta)_x$ values. The results from fitting on proton data with $(\alpha/\beta)_x < 20\ Gy$, using the fitting functions from insertion of f_{LQ}, f_{LQC}, f_{LQCq}, f_{LE} and f_{LQE} into Equation (3.3) are used to plot the curves in (a), (b), (c), (d) and (e), respectively. In each plot, Equations (2.16) and (2.17) are used to plot the RBE_{max} as a function of let according to the RORW model for the same six $(\alpha/\beta)_x$ values.

4.4 Application of different models

RBE as a function of LET, dose and $(\alpha/\beta)_x$ is shown for a selection of models in Figure 51, Figure 52 and Figure 53, respectively. In each figure, the RBE estimates of the linear models are shown in (a) and (b), while the estimates of non-linear models are given in (c) and (d). The selection of linear models aimed at showing the effect of LET^* restrictions and showing the effect of fitting on a balanced database in terms of $(\alpha/\beta)_x$. The selected non-linear models are the ones that showed best performance on the unrestricted database and on the restricted database containing data points that were considered most clinically relevant, based on the RMSE values of the fits. In (c) and (d) in Figure 51, Figure 52 and Figure 53, the estimated RBE from two of the linear models and two of the non-linear models are compared to RBE estimates obtained with coinciding dose, LET^* and $(\alpha/\beta)_x$ values using the RORU model, the RORW model and the MCN model.

From Figure 51 it is clear that all the models contradict a constant RBE of 1.1. Compared to a constant RBE of 1.1, the models show relatively large agreement, as the RBE increases with LET similarly for all models, although panels (e) and (f) reveals differences which could have a clinical impact, in particular in the region around 5 $keV/\mu m$ where the RORW model deviates from the other. The two non-linear models from the present work show similarities with the RORW model in this region, as the RBE estimates of these models show a slower increase with LET compared to the linear models. All models from the present work are more or less in between the RORW and RORU models. In panels (a) and (b), the model obtained with a balanced database with respect to $(\alpha/\beta)_x$ stand out from the other linear models with a steeper increase in RBE with LET. The reason for this is not fully understood, and a more detailed analysis of the data is required. A possible reason could be that the data points with higher $(\alpha/\beta)_x$ values, which are given higher weights in the fitting procedure, deviates a lot from the low $(\alpha/\beta)_x$ data. However, this can not be verified at this point and requires further investigation.

As seen in Figure 52, the dose dependency is similar for all the models, both at high and low $(\alpha/\beta)_x$ values. In panels (e) and (f), wee see that the RBE estimates from different models deviate most from each other at doses below 3 Gy. As radiation treatment normally is delivered with a fraction doses of 2 Gy, organs at risk normally receives doses below 2 Gy and consequently potentially a high RBE. The differences are smaller for higher $(\alpha/\beta)_x$ values, as seen from comparison of the left and right panels in the figure. The dependency of RBE on dose deviates more between the linear models in panels (a) and (b) compared to the non-linear models in (c) and (d) which have almost identical dose dependencies.

Both the models from the present work and the three established RORU, RORW and MCN models are derived with the same assumption for the dependency of RBE on $(\alpha/\beta)_x$. This is reflected in Figure 53, as the decrease in RBE with increasing $(\alpha/\beta)_x$ is similar for all the considered models, and the curves mainly differ by their placement in the vertical direction. The RBE estimates deviate most between different models for low $(\alpha/\beta)_x$, where the RORW model also deviates to some degree from the other models, as seen clearly in panel (f). From the same panel, it is clear that all models from the present work are between the RORU and the MCN models. The approach used in section 4.2.2 could in principle be used to find a new $(\alpha/\beta)_x$ dependency. As a somewhat weaker $(\alpha/\beta)_x$ dependency was indicated by the analysis, it is likely that this would produce a model which estimates the RBE to decreases slower with increasing $(\alpha/\beta)_x$ compared to the RBE estimates from the considered models.

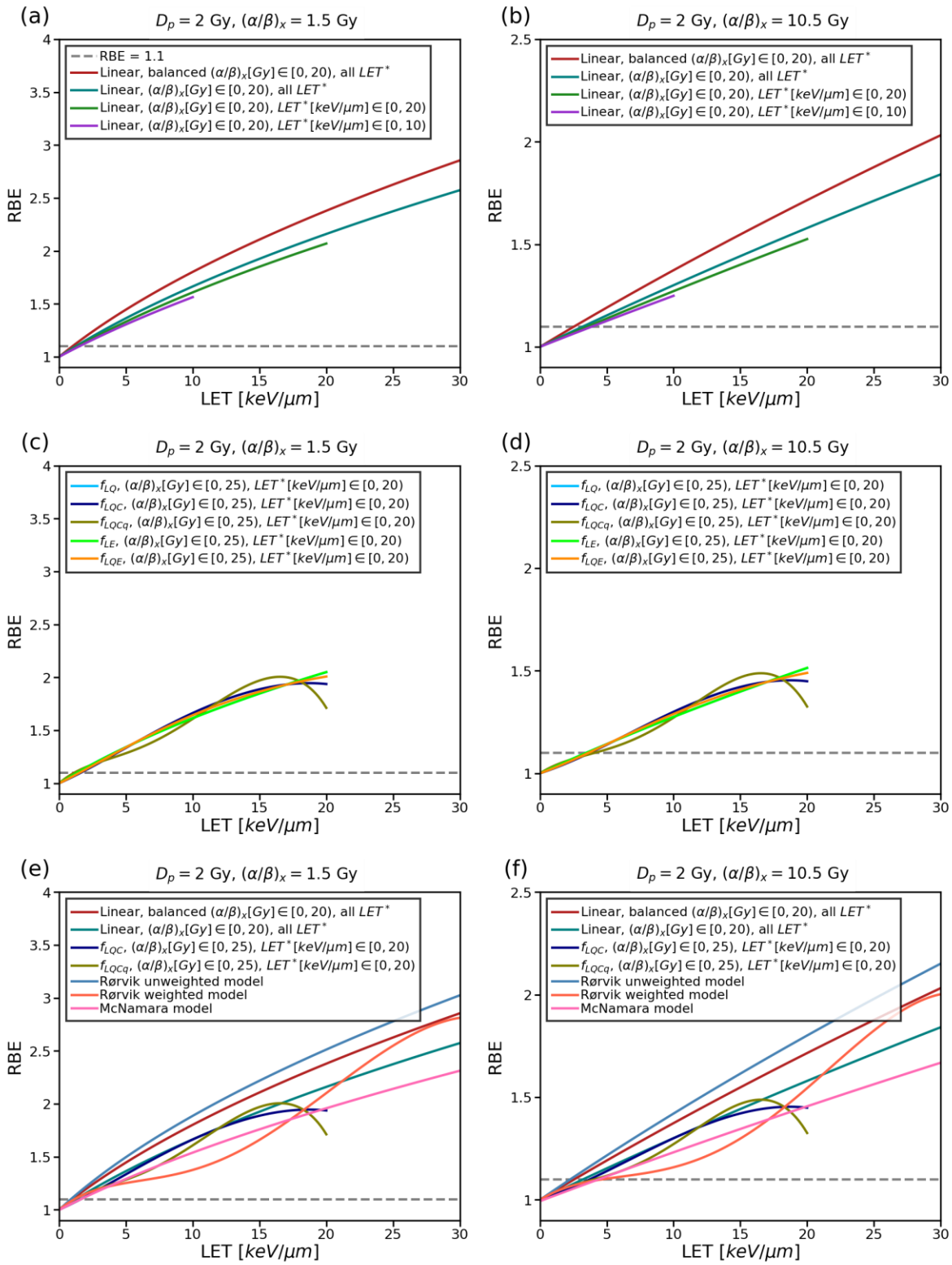

Figure 51: The RBE for different LET^* values for a monoenergetic beam with a physical proton dose of 2 Gy and two clinical relevant $(\alpha/\beta)_x$ values.

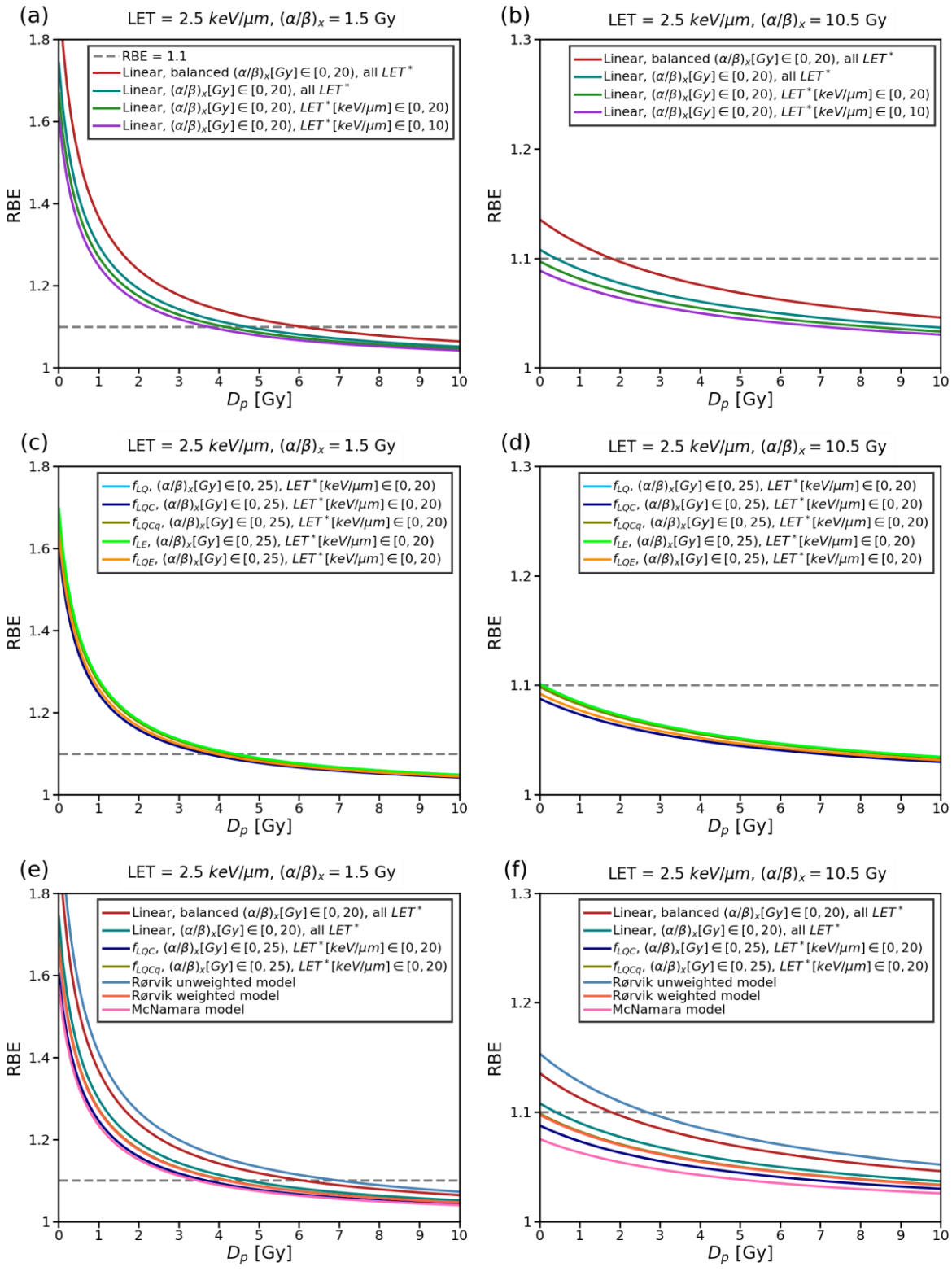

Figure 52: The RBE for different doses for a monoenergetic beam with a LET^* value of 2.5 $keV/\mu m$ and two clinical relevant $(\alpha/\beta)_x$ values.

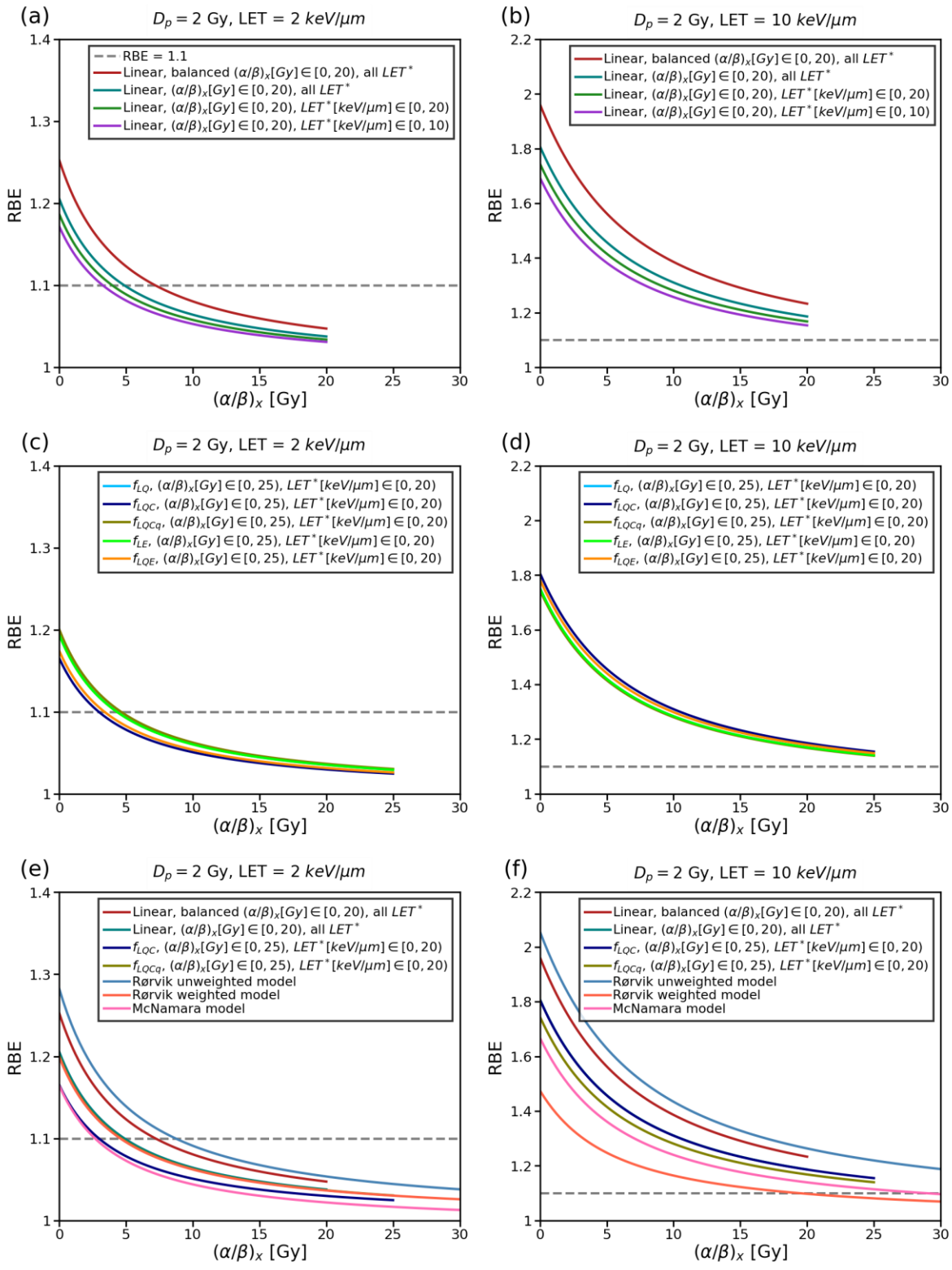

Figure 53: The RBE for different $(\alpha/\beta)_x$ values for a monoenergetic beam with a physical dose of 2 Gy and two clinical relevant LET^* values.

Figure 54 shows the biological (RBE-weighted) dose for a simulated SOBP scenario, calculated from three of the linear and two of the non-linear RBE models from the present work. Biological doses obtained with a constant RBE of 1.1 and from the RORU, RORW and MCN models are given for comparison. It is clear form the figure that all the considered variable RBE models largely disagrees with the constant RBE (1.1) model, especially at the distal half of the SOBP. Compared to a constant RBE of 1.1, all variable models, except from the RORU model show an overall agreement. For the entrance region between 0 and 2.5 cm, the estimated biological dose from all models were below or approximately at the same level as the constant RBE (1.1) model. At the proximal part of the SOBP, the RBE estimates deviate more between the models, and the RORU model estimates a relatively large biological dose. All the considered models estimated an increased RBE towards the distal end of the SOBP, with a steady increase in RBE across the SOBP for all models except RORW, which estimated a smaller increase in RBE towards the distal end of the SOBP. The LET in this region is between 5 and 10 $keV/\mu m$, which is the range of LET values where the RORW model RBE estimates deviated most from the other models in Figure 51(f).

The linear model obtained with all LET values estimates higher RBE than both the non-linear models across the SOBP. These three models are derived from relatively similar data, as the database contains few data points with LET above 20 $keV/\mu m$ and the highest LET value in the database is around 34 $keV/\mu m$. The quartic model is similar to the linear model at the proximal part of the SOBP, but the difference between these models increase across the SOBP, resulting in a much larger difference in RBE estimates at distal end of the SOBP. The opposite trend is seen for the cubic model, which is below the linear and quartic models at the proximal part of the SOBP, and with a larger increase in RBE with depth ends up giving a similar estimate as the linear model at the distal end of the SOBP. The linear models given by the green and purple curves are derived from databases with lower restrictions on the upper limit of LET and $(\alpha/\beta)_x$. Both these models give lower RBE estimates than the linear model obtained with all LET values across the SOBP. As most of the LET values in the SOBP

scenario are below 10 $keV/\mu m$, and an $(\alpha/\beta)_x$ value of 3 Gy was applied in the calculations, and the linear models given by green and purple lines are derived from restricted databases containing LET and $(\alpha/\beta)_x$ values similar to these, it is likely that these models provide good RBE estimates for this scenario. The RBE estimates given by the models from this work is generally higher than the estimated RBE from the MCN model. Compared to the RORW model, the models from this work generally gives similar or lower RBE estimates at the proximal part of the SOBP, and higher estimates at the distal end of the SOBP.

Considering all models including the constant RBE model, the deviations in biological dose estimates at the proximal part of the SOBP is approximately 0.2 Gy(RBE), while the deviations at the distal end of the SOBP is around 0.8 Gy(RBE). The large variations in model estimates at the distal end are significant in the context of a clinical setting. Thus, applying one of the considered RBE models in a treatment planning process, the outcome of treatment planning would be considerably affected by the choice of model.

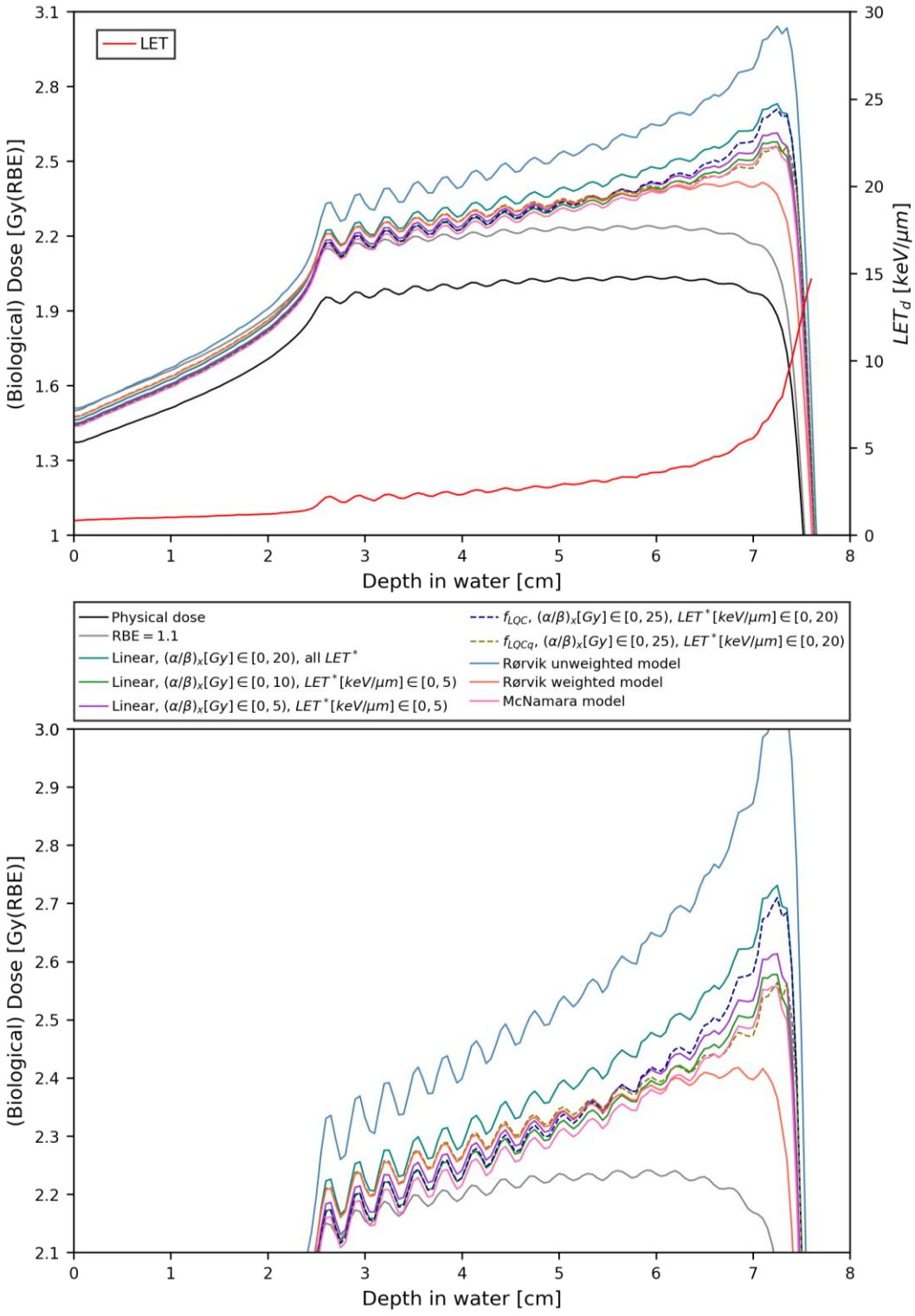

Figure 54: Depth dose distribution of an SOBP in water with model estimates obtained with linear and non-linear models from the present work and the RORW, RORU and MCN models. In the upper panel, the red curve represents the LET_d given by the right axis. The lower panel focus on the curves representing the biological dose given by the left axis in both panels. An $(\alpha/\beta)_x$ value of 3 Gy was applied in the calculations.

5. Discussion

A large database was collected and analyzed. The monoenergetic data was evenly distributed over the range of LET values, while neither monoenergetic nor SOBP data points was evenly distributed over the range of $(\alpha/\beta)_x$ values. Linear and non-linear regression was performed to obtain fits for RBE_{max} as a function of LET and $(\alpha/\beta)_x$. Overall, the analysis showed that the RBE increases with increasing LET and decreasing $(\alpha/\beta)_x$, which is in line with previous findings in published literature [4]. The fitting outcome was most clearly affected by restrictions on the upper range of included LET^* values and the lower range of included $(\alpha/\beta)_x$ values. Fitting on balanced databases showed that the distribution of data over the range of $(\alpha/\beta)_x$ values in the database can play a significant role for the outcome of fitting. Calculated RMSE values indicated similar performance for the linear fits and several of the non-linear fits.

Linear fitting of RBE_{max} as a function of LET and $(\alpha/\beta)_x$ without database restrictions resulted in similar fit parameters for all data, monoenergetic data and SOBP data. Although the fit parameter for the SOBP data is a little higher than the monoenergetic one, the difference is too small to be considered a clear indication of a dependency on irradiation condition, as it is likely to be due to experimental uncertainties and fluctuations in the data. However, a trend of higher k values for the SOBP data was observed from linear fitting on restricted databases with reduced upper limit of included LET and $(\alpha/\beta)_x$ values. These results indicate that the RBE_{max} is dependent on the radiation condition, and irradiation with a SOBP yields a steeper increase in RBE_{max} with increasing LET, compared to the RBE_{max}-LET relationship when monoenergetic radiation is used.

From the analysis of the assumption that RBE_{max} is inversely proportional with $(\alpha/\beta)_x$ in section 4.2.2 is clear that a larger database is needed to get more reliable results with this procedure. Several of the linear fits are based on too few data points to obtain conclusive results, especially in the analysis of monoenergetic proton data.

The results from the analysis indicated an overall trend of increasing RBE_{max} with decreasing $(\alpha/\beta)_x$. However, it could not be verified that the $1/(\alpha/\beta)_x$ dependency is a valid assumption for the RBE_{max}-$(\alpha/\beta)_x$ relationship of the considered proton data.

Comparison between the linear fits to data with $(\alpha/\beta)_x$ values in specific intervals and the RORU model showed that the fits to all data and monoenergetic data with low $(\alpha/\beta)_x$ ($\lesssim 6\ Gy$) had smaller slopes than the corresponding Rørvik lines, while the fits to higher $(\alpha/\beta)_x$ values had steeper slopes. The same was observed from the comparison with the obtained 3D fits applying the assumption that RBE_{max} is inversely proportional with $(\alpha/\beta)_x$.This indicates that the predicted RBE_{max} might be overestimated at low $(\alpha/\beta)_x$ and underestimated at high $(\alpha/\beta)_x$ when this assumption is applied. The analysis of the current database therefore implies that the inverse dependency of RBE_{max} on $(\alpha/\beta)_x$ is of a weaker character than the $1/(\alpha/\beta)_x$ relation. Appliance of a dependency more in line of $1/\sqrt{(\alpha/\beta)_x}$ might thus be more suited for precise modelling of the proton RBE. This is thus considered a good place to start in future investigation of the RBE_{max}-$(\alpha/\beta)_x$ relationship.

The results from fitting on balanced databases demonstrated the potential impact of using imbalanced data in the analysis. In Rørvik's exploration of different published phenomenological models from 2018 [22], the distributions of LET and $(\alpha/\beta)_x$ in the model databases were explored through a violin plot. It was shown that several of the considered models were based on imbalanced databases, both in terms of LET and $(\alpha/\beta)_x$. The Rørvik database, used to derive the RORU and RORW models was shown to be balanced with respect to LET. The fits obtained to balanced databases in the present work are based on repetitive fitting on randomly chosen samples of data points with $(\alpha/\beta)_x$ and LET values in different parts of the ranges. A weakness in this approach is that some data points will be given higher weight in the fitting procedure, i.e. data points in intervals with little data available. Uncertainties in the data may therefore lead to overall more uncertainty. Nevertheless, one could argue that the fits

are more likely to better reflect the dependencies in the considered range of $(\alpha/\beta)_x$ and LET values. A balanced database is generally more essential when a linear fitting procedure is applied, given that non-linear fitting functions are better suited to handle an imbalanced database. This can help explaining the large deviations in terms of RMSE in the outputs from non-linear fitting on balanced databases.

As seen in the present work, and previously published phenomenological RBE models, the data used in the regression fit plays a significant role. The present analysis of linear fitting on restricted databases showed that the applied restrictions on LET largely affected the fitting outcome. The significance of the applied upper limit of included LET values have previously been demonstrated by Rørvik, who refitted the previously published WED model with a lower LET_d cut-off to the associated database, resulting in a 13% lower fit parameter [22]. From linear fitting on restricted databases, indications of a non-linear RBE_{max}-LET^* relationship was observed. Lowering the upper limit of included LET^* values generally resulted in lower k values, meaning that the slope of the linear RBE_{max}-LET^* relationship is steeper when higher LET^* values are included in the fitting. A trend of increased slope was also observed when data points with $LET^* < 5\ keV/\mu m$ was excluded from the fitting, which strengthens the indication that a non-linear relationship involving larger increase in RBE_{max} per increase in LET^* for higher LET^* values is better suited to model the proton data. Some of the results from linear fitting on databases with reduced range of included $(\alpha/\beta)_x$ values indicated that the RBE_{max}-$(\alpha/\beta)_x$ relationship might be different for high and low $(\alpha/\beta)_x$. However, this observation was largely seen in the analysis concerning data points with $(\alpha/\beta)_x < 5\ Gy$, and the inclusion or exclusion of these data. The observed trends might therefore be connected to the fact that the data points are not evenly distributed across the range of $(\alpha/\beta)_x$ values. As the database contains a large amount of data with $(\alpha/\beta)_x < 5\ Gy$, the observed deviations from fits obtained with restrictions concerning these data points might be a result of the fact that the database will be more balanced when these data points are excluded.

The results from linear fitting on restricted databases thus implies the need for further investigation of a non-linear dependency of RBE_{max} on LET. Although this was done by Rørvik et al [10], investigating the RBE-LET relationship by fitting of polynomials from 1st to 5th degree to a database of 85 in vitro data points, further exploration using different regression techniques and larger databases is needed.

The RMSE values of the non-linear fits to the unrestricted database indicated that several of the non-linear fitting functions yields similar goodness of fit as the linear fitting function. However, it should be kept in mind that RMSE is only a measure of the mean distance between the predicted values and the actual values, and an overfitted model can therefore be falsely indicated to be a good fit by only considering the RMSE. The visualization of the RBE_{max}-LET relationships obtained with non-linear fitting showed that several of the fits had similar characteristics as the linear fit, as the slopes of the curves were similar for lower and higher values of LET^{*}. The present analysis of linear and non-linear fitting do not clearly favor one regression technique over the other. However, as the results from linear fitting on restricted databases indicated a non-linear trend, and a general trend of lower RMSE values for non-linear fits was observed, the analysis implies that a non-linear RBE_{max}-LET relationship might be preferred. A weak indication was observed that the 4th degree polynomial fitting function was best suited to model the data, which is in line with previous results from Rørviks investigation of a non-linear RBE-LET relationship [10]. From the SOBP scenario, it was observed that the RBE estimates from the quartic model is very similar to the estimates given by the linear models obtained with the strictest database restrictions (given by purple and green curves in Figure 54). It was pointed out that these two models are likely to be more applicable in the SOBP scenario, based on the $(\alpha/\beta)_x$ and LET values used to derive them. The fact that the RBE estimates of the quartic model resemble these two models estimates might thus indicate a good performance of this model in the SOBP scenario.

The appliance of models in an SOBP scenario clearly show the large deviations between the different RBE models.

Finally, it should be noted that as RBE models are based on *in vitro* data, the translation validity of the results to *in vivo* is debatable, and the uncertainties connected to the appliance of the results in a clinical setting is typically not known. It is therefore of great importance to compare *in vitro* based predictions to available *in vivo* and clinical data on RBE. The need for further *in vivo* studies at lower fractional doses have previously been demonstrated [56].

Future work

- The present analysis of a possible RBE dependence on the irradiation condition (monoenergetic vs SOBP) showed weak indications that such a dependence could be present. However, the results were not conclusive, and further investigation is needed to verify whether such a trend is present or not. It was seen in the present work that more work is needed to isolate this dependency in the analyses, as it could not be determined whether the observed deviations in the results from the present analyses was due to a dependency on irradiation condition or other effects. A good place to start is thus to collect a balanced database containing similar amount of data obtained from irradiation experiments using a monoenergetic beam and experiments using an SOBP beam. Analyses of such a database could involve fits to only monoenergetic data and only SOBP data, and investigate whether the predictions from these fits show different trends or not.
- The present work indicated that the dependency of RBE_{max} on $(\alpha/\beta)_x$ is somewhat weaker than $1/(\alpha/\beta)_x$. A $1/\sqrt{(\alpha/\beta)_x}$ was suggested, as it might be a good candidate. However, more work is needed to explore this dependency, and potentially find a more suitable one than the widely used assumption that RBE_{max} is inversely proportional with $(\alpha/\beta)_x$. The procedure applied in section 4.2.2 could in principle be used to find a new $(\alpha/\beta)_x$ dependency.

- A non-linear RBE-LET relationship should be further explored. Although several results from the present analysis imply a non-linear trend, further investigation is needed to determine if a non-linear relationship is preferred over a linear relation, or perhaps the other way around.
- Comparisons of *in vitro* based predictions to available *in vivo* data both from experimental and clinical settings is much needed to fully understand possible uncertainties in the appliance of RBE models derived from *in vitro* data in a clinical setting.
- Further exploration of the appliance of different regression techniques and methods to evaluate and compare regression outputs can contribute to reduced uncertainties in the RBE modelling. It was seen in the present work that the evaluation and comparison based only on RMSE did not provide conclusive results, and might not fully cover all perspectives and deviations in the results. As RMSE can only provide an estimate of the goodness of fit and do not detect overfitted models, other evaluation methods and statistical tests should be considered.
- Application of machine learning (ML) algorithms in RBE modelling could also be looked into. ML algorithms designed for data preprocessing can be a valuable tool for outlier detection and data selection, and regression analysis based on application of ML models should be further investigated.

6. Conclusion

In the present study, a database of *in vitro* proton data was analysed in the framework of a phenomenological proton RBE model, with the main goal of exploring how data selection, regression methods and model assumptions affect the RBE estimation. Investigating the RBE dependence on $(\alpha/\beta)_x$, implied a weaker dependency than the widely assumed $1/(\alpha/\beta)_x$ relation. Linear fitting on both the unrestricted and restricted databases showed a trend of higher k values for the SOBP data compared to the monoenergetic data, indicating a dependency of RBE on the radiation condition, with a steeper increase in RBE with increasing LET when an SOBP beam is used, compared to applying a monoenergetic beam. Restrictions on the range of included $(\alpha/\beta)_x$ values indicated a different RBE-$(\alpha/\beta)_x$ relationship for low and high $(\alpha/\beta)_x$, which is in line with the results from investigation of the $1/(\alpha/\beta)_x$ relation. Reducing the upper limit of included LET values, a clear trend of lower k values was observed, indicating that the slope of the applied linear RBE-LET relationship is smaller for lower LET and thus implying that non-linear models might provide a better representation of the RBE-LET relationship. Further investigation applying non-linear fitting on the proton data implied better performance for several of the non-linear fits compared to linear fits to the same data. The quartic fitting function showed overall best performance on the current database. The present analyses generally highlighted the pronounced effects of data selection on model output and RBE predictions. This was also demonstrated by obtaining RBE estimates from different models for a SOBP scenario, showing large differences between the models. Published RBE models show large variations in RBE estimates, and these variations should be reduced to be able to apply a variable RBE in clinical settings. Many different model assumptions are applied, and more work is needed to decide which assumptions result in most accurate RBE modelling. From the SOBP scenario in the present work, one can deduce that if one of the considered variable RBE models were to replace the assumption if a constant proton RBE=1.1, the RORW model and the quartic model from the present work could be good candidates, as these seems to give RBE estimates that are higher than

RBE=1.1, but not to the same extent as several other models, such as the RORU model. Both these models give increasing RBE estimates across the SOBP, which is in line with previous observations. As the models vary in RBE estimates at the distal end of the SOBP, further investigation of the RBE in this region could be the final step to decide between the two models.

Appendix

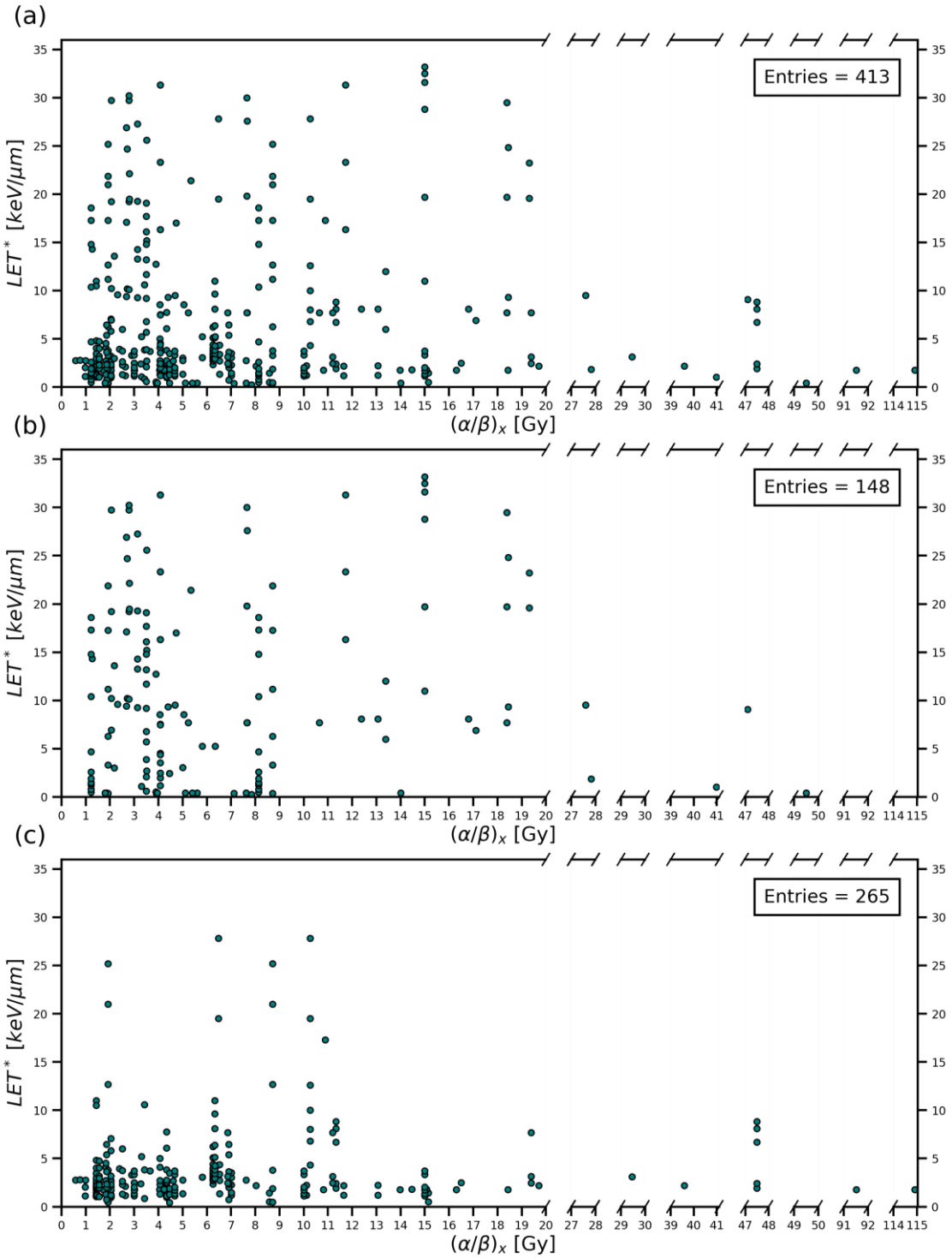

Figure A1: Proton LET^* plotted against $(\alpha/\beta)_x$ values, without the restriction on included $(\alpha/\beta)_x$ values that was applied in Figure 17. In (a), all proton data points are shown. Data points obtained using monoenergetic radiation and using an SOBP are shown in (b) and (c) respectively.

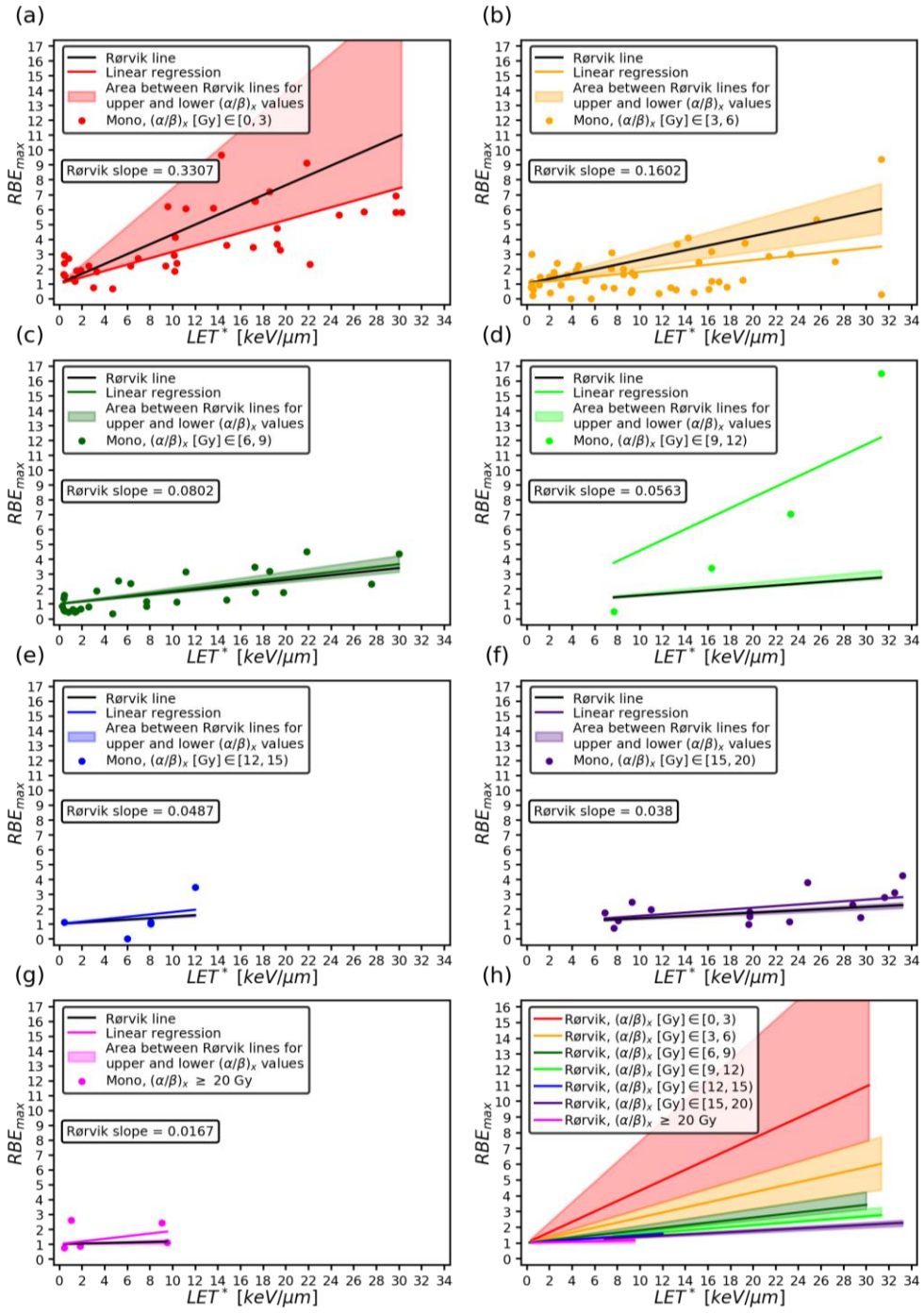

Figure A2: Comparison of the linear fits to the monoenergetic proton data (shown in Figure 21) and the RORU model. (a)-(g) shows the regression fits (colored lines), the RORU lines obtained by putting the mean of the $(\alpha/\beta)_x$ values for the data points in each interval of $(\alpha/\beta)_x$ into Equation (2.15) (black lines) and the area spanned by the curves obtained with Equation (2.15) using the upper and lower limits for the $(\alpha/\beta)_x$ value in each interval (colored area). In (a), the area is shown for a lower limit of $(\alpha/\beta)_x = 1\ Gy$, assuming that clinical use will not involve $(\alpha/\beta)_x$ values below this. The slope of the RORU lines are also given in (a)-(g). The RORU lines and limit areas for each $(\alpha/\beta)_x$ interval are potted together in (h).

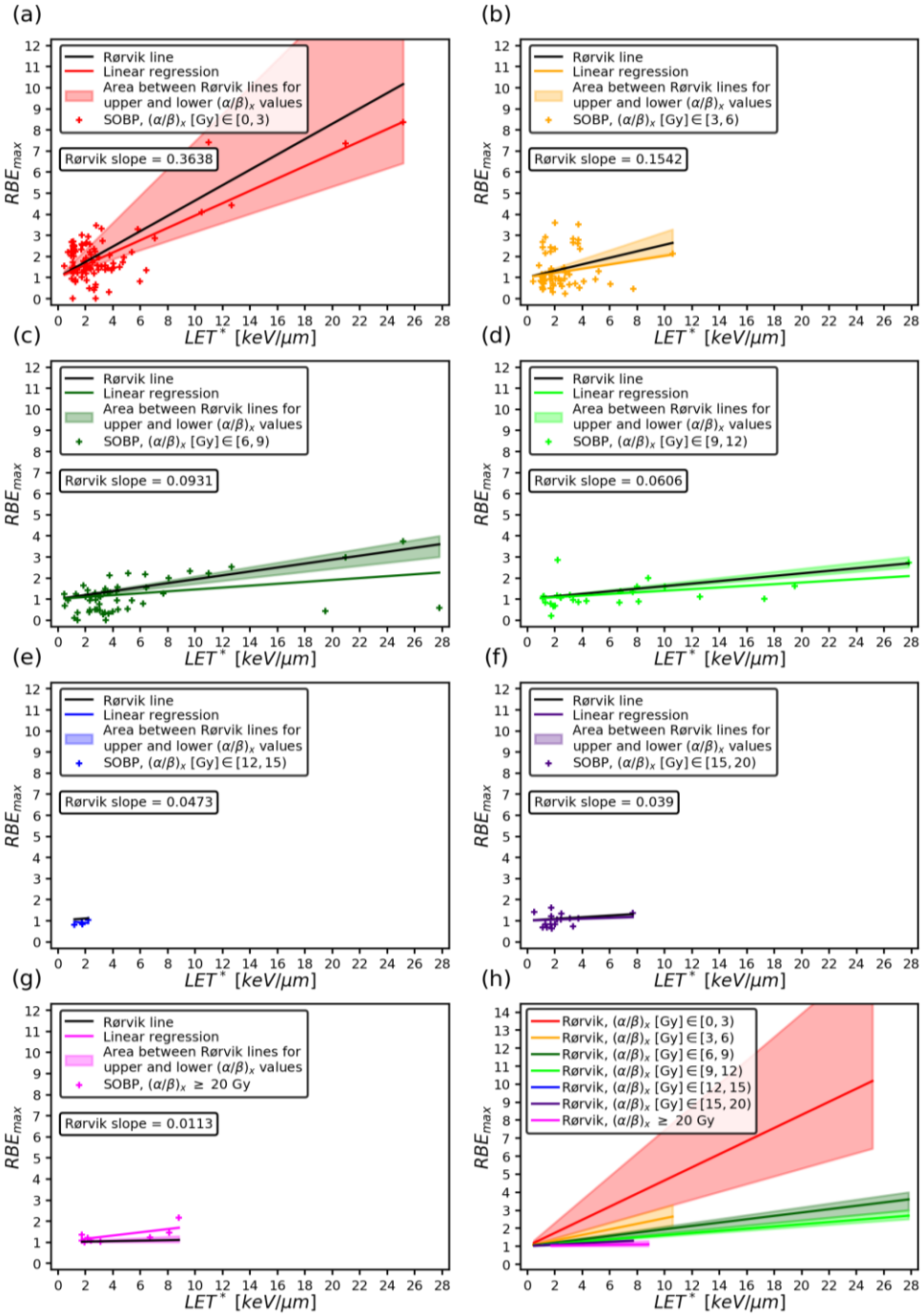

Figure A3: Comparison of the linear fits to the SOBP proton data (shown in Figure 22) and the RORU model. (a)-(g) shows the regression fits (colored lines), the RORU lines obtained by putting the mean of the $(\alpha/\beta)_x$ values for the data points in each interval of $(\alpha/\beta)_x$ into Equation (2.15) (black lines) and the area spanned by the curves obtained with Equation (2.15) using the upper and lower limits for the $(\alpha/\beta)_x$ value in each interval (colored area). In (a), the area is shown for a lower limit of $(\alpha/\beta)_x = 1\ Gy$, assuming that clinical use will not involve $(\alpha/\beta)_x$ values below this. The slope of the RORU lines are also given in (a)-(g). The RORU lines and limit areas for each $(\alpha/\beta)_x$ interval are potted together in (h).

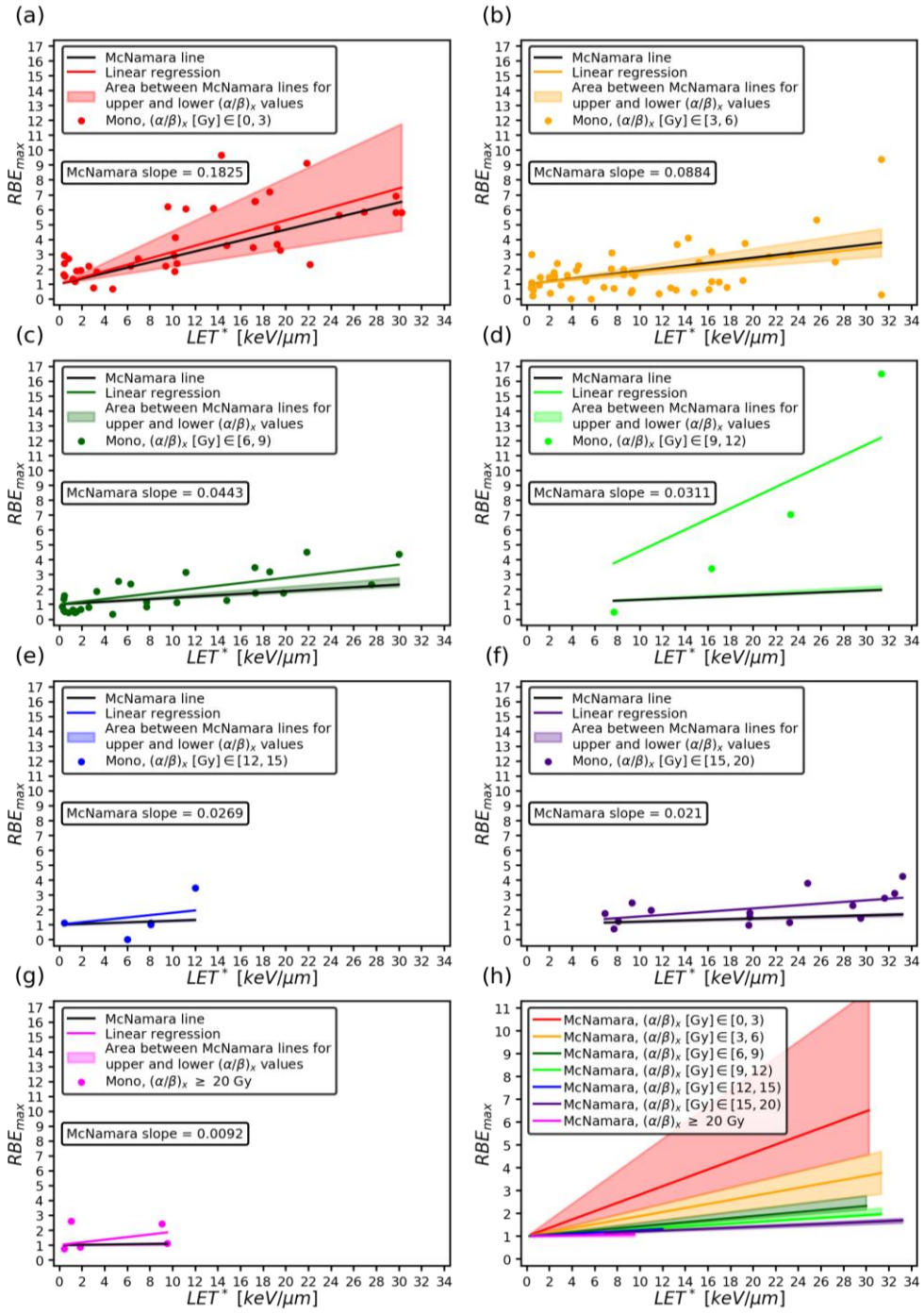

Figure A4: (a)-(g): The linear fits in Figure 21 for monoenergetic proton data divided into seven subsets are shown (colored lines) together with the MCN lines obtained by using the mean $(\alpha/\beta)_x$ value for the data points in each subset as an input in Equation (2.11) (black lines), The slope of the MCN lines are given in each plot. The area between the MCN lines obtained by using the upper and lower limits of each $(\alpha/\beta)_x$ interval as an input in Equation (2.11) are also shown (colored areas). A lower limit of $(\alpha/\beta)_x = 1\ Gy$ is used in (a), assuming that clinical use will not involve $(\alpha/\beta)_x$ values below this. The MCN lines and limit areas for each $(\alpha/\beta)_x$ interval are potted together in (h).

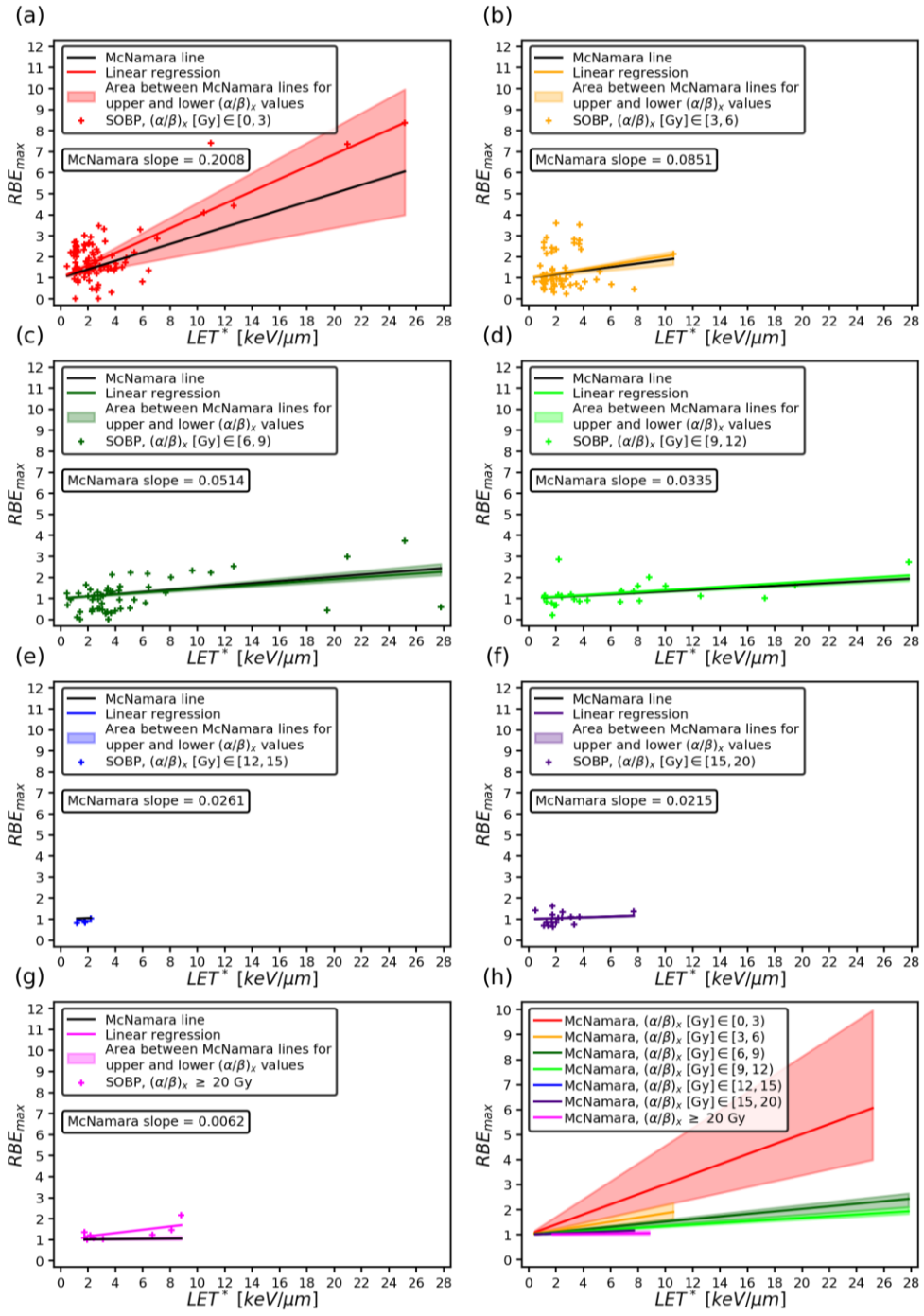

Figure A5: (a)-(g): The linear fits in Figure 22 for SOBP proton data divided into seven subsets are shown (colored lines) together with the MCN lines obtained by using the mean $(\alpha/\beta)_x$ value for the data points in each subset as an input in Equation (2.11) (black lines), The slope of the MCN lines are given in each plot. The area between the MCN lines obtained by using the upper and lower limits of each $(\alpha/\beta)_x$ interval as an input in Equation (2.11) are also shown (colored areas). A lower limit of $(\alpha/\beta)_x = 1\ Gy$ is used in (a), assuming that clinical use will not involve $(\alpha/\beta)_x$ values below this. The MCN lines and limit areas for each $(\alpha/\beta)_x$ interval are potted together in (h).

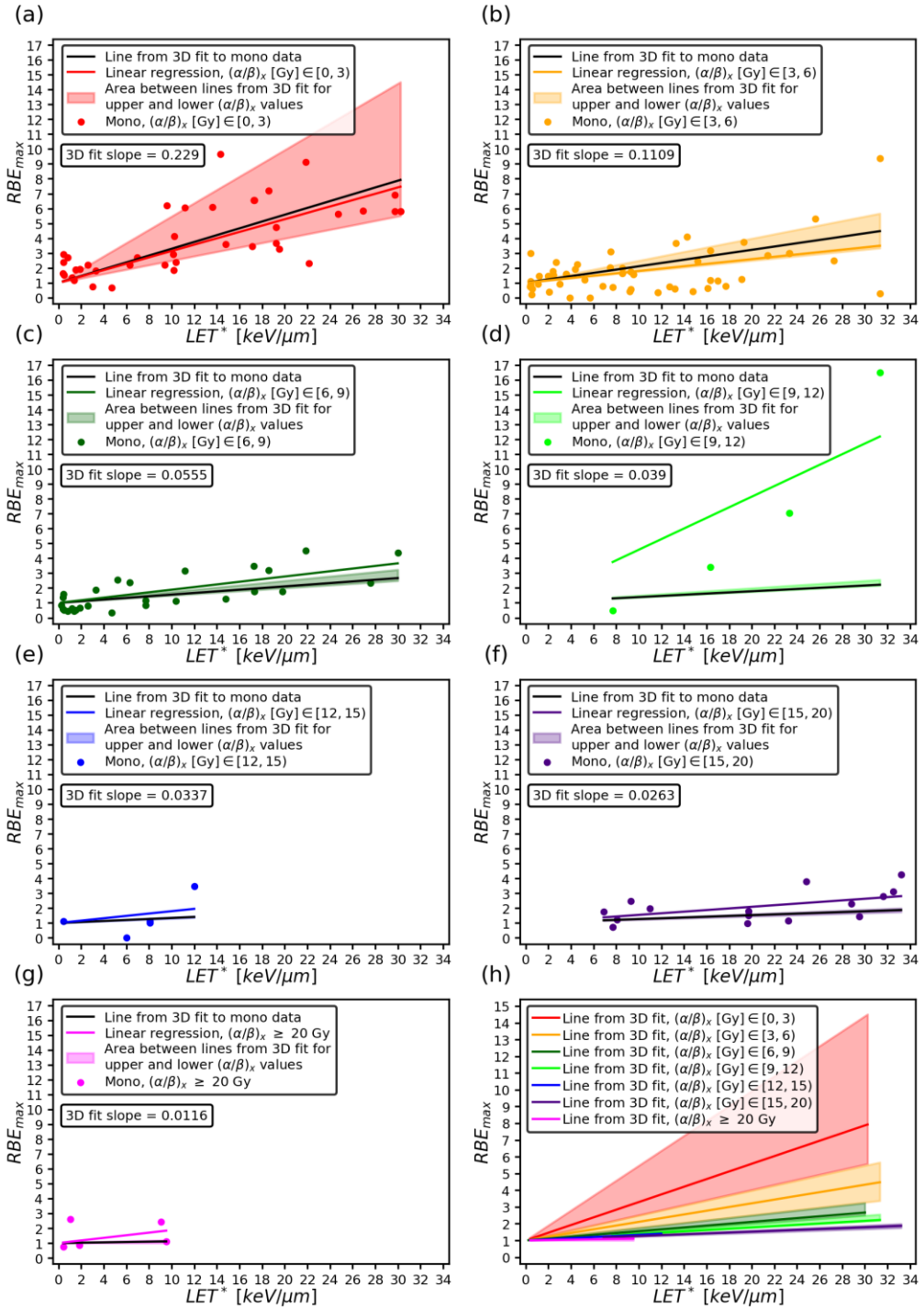

Figure A6: The linear regression fits to monoenergetic proton data points from Figure 21 are shown in (a-g) together with lines obtained from the fitted plane to monoenergetic data from Figure 18(b). These lines are plotted by inserting the obtained fit parameter $k = 0.45$ into Equation (3.2) and using the mean $(\alpha/\beta)_x$ value of the data points in each of the subsets as an input for $(\alpha/\beta)_x$ in the equation. The slopes of the lines from the 3D fit are given in each plot. The colored area in each plot is obtained by using the same equation, and the upper and lower boundaries in each interval of $(\alpha/\beta)_x$ values. In (h), the lines from the 3D fit and the limit areas for each $(\alpha/\beta)_x$ interval are plotted together.

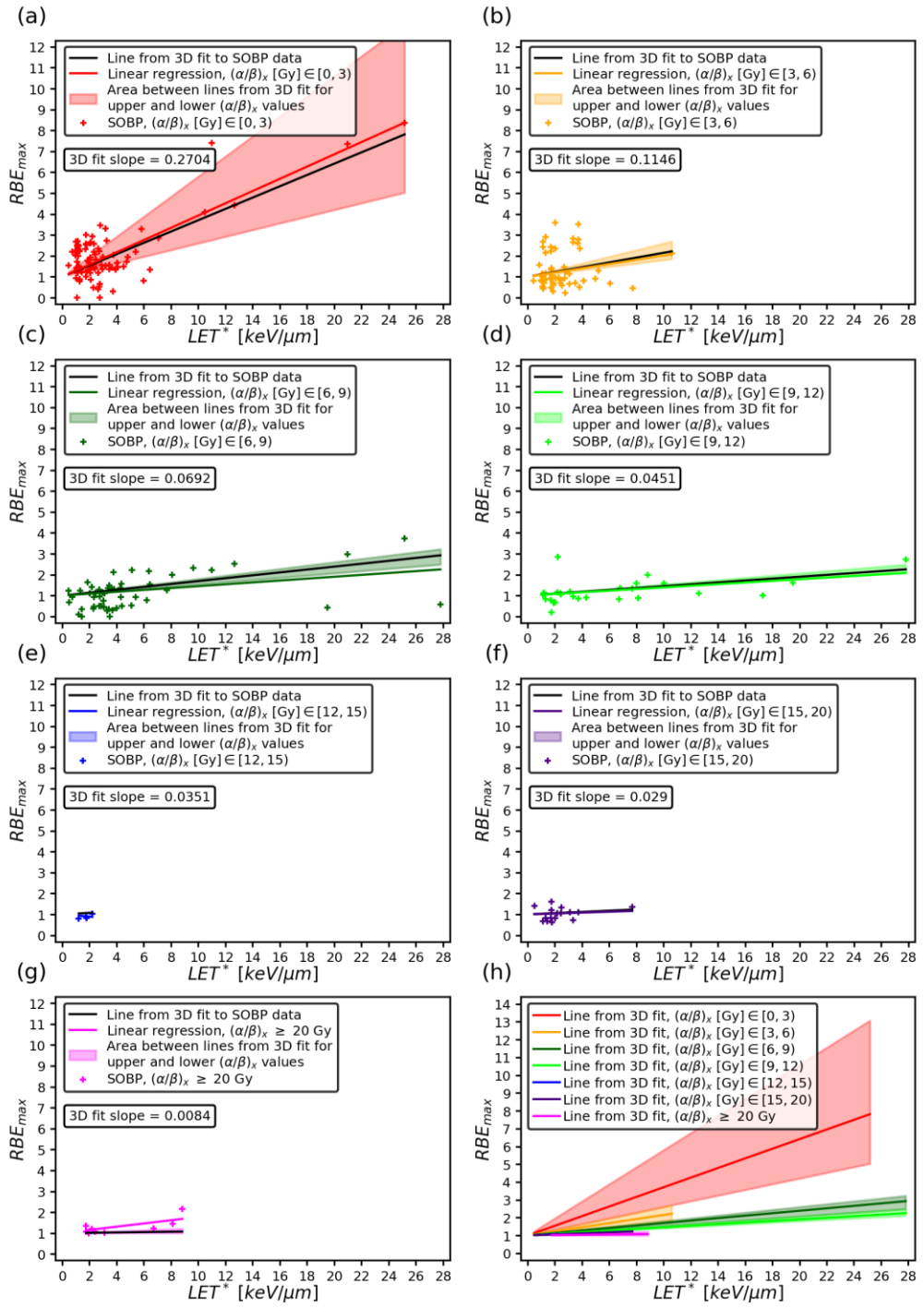

Figure A7: The linear regression fits to SOBP proton data points from Figure 22 are shown in (a-g) together with lines obtained from the fitted plane to SOBP data from Figure 18(c). These lines are plotted by inserting the obtained fit parameter $k = 0.48$ into Equation (3.2) and using the mean $(\alpha/\beta)_x$ value of the data points in each of the subsets as an input for $(\alpha/\beta)_x$ in the equation. The slopes of the lines from the 3D fit are given in each plot. The colored area in each plot is obtained by using the same equation, and the upper and lower boundaries in each interval of $(\alpha/\beta)_x$ values. In (h), the lines from the 3D fit and the limit areas for each $(\alpha/\beta)_x$ interval are plotted together.

Bibliography

1. Observatory, T.G.C. *Cancer today*. 2020; Available from: https://gco.iarc.fr/today/home.

2. Barton, M.B., et al., *Estimating the demand for radiotherapy from the evidence: A review of changes from 2003 to 2012.* Radiotherapy and Oncology, 2014. **112**(1): p. 140-144.

3. Mohan, R. and D. Grosshans, *Proton therapy - Present and future.* Adv Drug Deliv Rev, 2017. **109**: p. 26-44.

4. Paganetti, H., *Relative biological effectiveness (RBE) values for proton beam therapy. Variations as a function of biological endpoint, dose, and linear energy transfer.* Physics in Medicine and Biology, 2014. **59**(22): p. R419-R472.

5. Wouters, B.G., et al., *Radiobiological intercomparison of the 160 MeV and 230 MeV proton therapy beams at the Harvard Cyclotron Laboratory and at Massachusetts General Hospital.* Radiat Res, 2015. **183**(2): p. 174-87.

6. Polster, L., et al., *Extension of TOPAS for the simulation of proton radiation effects considering molecular and cellular endpoints.* Phys Med Biol, 2015. **60**(13): p. 5053-70.

7. Rørvik, E. *Analysis and Development of Phenomenological Models for the Relative Biological Effectiveness in Proton Therapy*. 2019; Available from: http://hdl.handle.net/1956/20923.

8. Wedenberg, M., B.K. Lind, and B. Hårdemark, *A model for the relative biological effectiveness of protons: the tissue specific parameter α/β of photons is a predictor for the sensitivity to LET changes.* Acta Oncol, 2013. **52**(3): p. 580-8.

9. Guan, F., et al., *Spatial mapping of the biologic effectiveness of scanned particle beams: towards biologically optimized particle therapy.* Scientific Reports, 2015. **5**(1): p. 9850.

10. Rorvik, E., et al., *A phenomenological biological dose model for proton therapy based on linear energy transfer spectra.* Med Phys, 2017. **44**(6): p. 2586-2594.

11. Newhauser, W.D. and R. Zhang, *The physics of proton therapy.* Physics in Medicine and Biology, 2015. **60**(8): p. R155-R209.

12. Particle Data, G., et al., *Review of Particle Physics*. Physical Review D, 2018. **98**(3): p. 030001.

13. Ytre-Hauge, K.S., *Measurements and Monte Carlo Simulations of Neutron Doses from Radiation Therapy with Photons, Protons and Carbon Ions.* 2013.

14. Khan, F.M. and J.P. Gibbons, *Khan's the physics of radiation therapy*. 2014.

15. Durante, M., *New challenges in high-energy particle radiobiology.* Br J Radiol, 2014. **87**(1035): p. 20130626.

16. Rørvik, E., *Models for the Relative Biological Effectiveness in Proton Therapy.* 2018.

17. Mohan, R., et al., *Radiobiological issues in proton therapy.* Acta Oncologica, 2017. **56**(11): p. 1367-1373.

18. Carante, M. and F. Ballarini, *Modelling cell death for cancer hadrontherapy.* AIMS Biophysics, 2017. **4**: p. 465-490.

19. McMahon, S.J., *The linear quadratic model: usage, interpretation and challenges.* Phys Med Biol, 2018. **64**(1): p. 01tr01.

20. Rørvik, E., S. Thörnqvist, and K.S. Ytre-Hauge, *The experimental dose ranges influence the LETd dependency of the proton minimum RBE (RBEmin).* Physics in Medicine & Biology, 2019. **64**(19): p. 195001.

21. Belli, M., et al., *RBE-LET relationships for cell inactivation and mutation induced by low energy protons in V79 cells: further results at the LNL facility.* Int J Radiat Biol, 1998. **74**(4): p. 501-9.

22. Rorvik, E., et al., *Exploration and application of phenomenological RBE models for proton therapy.* Phys Med Biol, 2018. **63**(18): p. 185013.

23. Bettega, P.C.P.C.A.C.J.H.N.I.R.M.P.M.G.L.P.L.T.D., *Radiobiological studies on the 65MeV therapeutic proton beam at Nice using human tumour cells.* International Journal of Radiation Biology, 2000. **76**(10): p. 1297-1303.

24. Baggio, L., et al., *Relative Biological Effectiveness of Light Ions in Human Tumoural Cell Lines: Role of Protein p53.* Radiation Protection Dosimetry, 2002. **99**(1-4): p. 211-214.

25. Joiner, M.C. and A. van der Kogel, *Basic Clinical Radiobiology Fourth Edition*. 2009: Taylor & Francis.

26. Carabe, A., et al., *Range uncertainty in proton therapy due to variable biological effectiveness.* Phys Med Biol, 2012. **57**(5): p. 1159-72.

27. McNamara, A.L., J. Schuemann, and H. Paganetti, *A phenomenological relative biological effectiveness (RBE) model for proton therapy based on all published in vitro cell survival data.* Phys Med Biol, 2015. **60**(21): p. 8399-416.

28. Paganetti, H., et al., *Calculation of relative biological effectiveness for proton beams using biological weighting functions.* Int J Radiat Oncol Biol Phys, 1997. **37**(3): p. 719-29.

29. Belli, M., A. Campa, and I. Ermolli, *A Semi-Empirical Approach to the Evaluation of the Relative Biological Effectiveness of Therapeutic Proton Beams: The Methodological Framework.* Radiation Research, 1997. **148**(6): p. 592-598.

30. Wilkens, J.J. and U. Oelfke, *A phenomenological model for the relative biological effectiveness in therapeutic proton beams.* Physics in Medicine and Biology, 2004. **49**(13): p. 2811-2825.

31. Tilly, N., et al., *The influence of RBE variations in a clinical proton treatment plan for a hypopharynx cancer.* Physics in Medicine and Biology, 2005. **50**(12): p. 2765-2777.

32. Chen, Y. and S. Ahmad, *Empirical model estimation of relative biological effectiveness for proton beam therapy†.* Radiation Protection Dosimetry, 2011. **149**(2): p. 116-123.

33. Jones, B., *Towards Achieving the Full Clinical Potential of Proton Therapy by Inclusion of LET and RBE Models.* Cancers, 2015. **7**(1): p. 460-480.

34. Mairani, A., et al., *A phenomenological relative biological effectiveness approach for proton therapy based on an improved description of the mixed radiation field.* Physics in Medicine and Biology, 2017. **62**(4): p. 1378-1395.

35. Peeler, C.R., et al., *Clinical evidence of variable proton biological effectiveness in pediatric patients treated for ependymoma.* Radiotherapy and Oncology, 2016. **121**(3): p. 395-401.

36. Frese, M.C., et al., *Application of Constant vs. Variable Relative Biological Effectiveness in Treatment Planning of Intensity-Modulated Proton Therapy.* International Journal of Radiation Oncology*Biology*Physics, 2011. **79**(1): p. 80-88.

37. Unkelbach, J., et al., *Reoptimization of Intensity Modulated Proton Therapy Plans Based on Linear Energy Transfer.* International Journal of Radiation Oncology*Biology*Physics, 2016. **96**(5): p. 1097-1106.

38. Mairani, A., et al., *Data-driven RBE parameterization for helium ion beams.* Phys Med Biol, 2016. **61**(2): p. 888-905.

39. Abolfath, R., et al., *A model for relative biological effectiveness of therapeutic proton beams based on a global fit of cell survival data.* Sci Rep, 2017. **7**(1): p. 8340.

40. Hawkins, R.B., *A microdosimetric-kinetic theory of the dependence of the RBE for cell death on LET.* Med Phys, 1998. **25**(7 Pt 1): p. 1157-70.

41. Hawkins, R.B., *A microdosimetric-kinetic model for the effect of non-Poisson distribution of lethal lesions on the variation of RBE with LET.* Radiat Res, 2003. **160**(1): p. 61-9.

42. GSI, H.f.S., *Particle Irradiation Data Ensemble, version 3.1 (PIDE3.1); Available at (project page): http://www.gsi.de/bio-pide.*

43. Alvestad, O.N., *Models for the relative biological effectiveness in proton therapy: on the impact of cell-irradiation design.* 2020.

44. Mara, E., et al., *Investigating the impact of alpha/beta and LETd on relative biological effectiveness in scanned proton beams: An in vitro study based on human cell lines.* Med Phys, 2020. **47**(8): p. 3691-3702.

45. Belli, M., et al., *Effectiveness of Monoenergetic and Spread-Out Bragg Peak Carbon-Ions for Inactivation of Various Normal and Tumour Human Cell Lines.* Journal of radiation research, 2008. **49**: p. 597-607.

46. Wulf, H., et al., *Heavy-Ion Effects on Mammalian Cells: Inactivation Measurements with Different Cell Lines.* Radiation Research, 1985. **104**(2): p. S122-S134.

47. Okayasu, R., et al., *Repair of DNA Damage Induced by Accelerated Heavy Ions in Mammalian Cells Proficient and Deficient in the Non-homologous End-Joining Pathway.* Radiation research, 2006. **165**: p. 59-67.

48. Cox, R., et al., *Mutation and inactivation of mammalian cells by various ionising radiations.* Nature, 1977. **267**(5610): p. 425-427.

49. Kraft, G.H., E.A. Blakely, and W. Kraft-Weyrather, *Super heavy ions: uranium radiobiology and physics*. 1985: United States. p. 128-137.

50. Belli, M., et al., *RBE-LET Relationship for the Survival of V79 Cells Irradiated with Low Energy Protons.* International Journal of Radiation Biology, 1989. **55**(1): p. 93-104.

51. Schuff, J.A., et al., *Relative biological effectiveness measurements of low energy proton and lithium beams on tumor cells.* Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms, 2002. **187**(3): p. 345-353.

52. Howard, M.E., *Characterization of Relative Biological Effectiveness for Proton Therapy in Human Cancer Cell Lines*. 2017, College of Medicine - Mayo Clinic: Ann Arbor. p. 114.

53. Matsuura, T., et al., *Apparent absence of a proton beam dose rate effect and possible differences in RBE between Bragg peak and plateau.* Med Phys, 2010. **37**(10): p. 5376-81.

54. Williams, J.R., et al., *Relative Survival of Hybrid X-Ray-Resistant, and Normally Sensitive Mammalian Cells Exposed to X Rays and Protons under Aerobic and Hypoxic Conditions.* Radiation Research, 1978. **73**(3): p. 585-590.

55. Prise, K.M., et al., *The Irradiation of V79 Mammalian Cells by Protons with Energies below 2 MeV. Part II. Measurement of Oxygen Enhancement Ratios and DNA Damage.* International Journal of Radiation Biology, 1990. **58**(2): p. 261-277.

56. Saager, M., et al., *Determination of the proton RBE in the rat spinal cord: Is there an increase towards the end of the spread-out Bragg peak?* Radiotherapy and Oncology, 2018. **128**(1): p. 115-120.