1. Introduction

This repository contains the replication data and code of the study "Evaluation of indirect damage and damage saturation in hypofractionated radiotherapy treatments of NSCLC and brain metastases through analysis of dose-response curves".

The repository contains the following files, organized in three folders:

1.1. Database files: database.xls data.mat

1.2. Code files:

1.2.1 Main functions: repeatability_study.m dose_variation_study.m
probability_variation_study.m dose_and_prob_variation_study.m

1.2.2 Other functions: MaxLikelihood_TCP.m loglikelihood_logistic.m loglikelihood_poisson.m
EQD2_calculator.m weight_calculator.m

1.3. Results files:

1.3.1 Brain metastases: Brain_central_parameter_values.mat Brain_logit_alphabeta_study.mat
Brain_logit_gamma_study.mat Brain_logit_tk_study.mat
Brain_poisson_g_study.mat Brain_poisson_alphabeta_study.mat
Brain_poisson_tk_study.mat Brain_logit_dose_and_TCP_perturbations.mat

1.3.2 Lung: Lung_central_parameter_values.mat Lung_logit_alphabeta_study.mat
Lung_logit_gamma_study.mat Lung_logit_tk_study.mat
Lung_poisson_g_study.mat Lung_poisson_alphabeta_study.mat
Lung_poisson_tk_study.mat Lung_logit_dose_and_TCP_perturbations.mat

2. Database of treatments

The full dataset used for this study is presented in the **database.xls** file, in excel format. Two clinical databases are used with information corresponding to early NSCLC and brain metastases.

The datasets are also presented in MATLAB (The Mathworks, Natwick, MA) format in the file **data.mat**. This Matlab file contains two arrays with the clinical databases, data_TCPexp_lung and data_TCPexp_brain. Data in these arrays is organized in columns with the following information:

1) Number of patients/metastases treated, in NSCLC/brain databases, respectively.
2) Kaplan-Meier TCP value 3 years and 1 year after treatment for NSCLC and brain, respectively.
3) Agresti-Coull estimate of the tumor control probability.
4) Lower bound of the confidence interval associated to the tumor control probability, calculated with the Clopper-Pearson estimation.
5) Upper bound of the confidence interval associated to the tumor control probability, calculated with the Clopper-Pearson estimation.
6) Treatment duration in days.
7) Number or treatment schedules used (1 for NSCL, 2 for brain).
8) Number of treatment fractions used in the first treatment schedule.
9) Average dose per fraction associated to the first treatment schedule.
10) Number of treatment fractions used in the second treatment schedule.
11) Average dose per fractions associated to the second treatment schedule.

Additionally, the data.mat file also contains two vectors, dep_a and dep_b, which specify the SF functional dependencies analyzed in this work corresponding to: i) the LQ model, using SF1 (Eq 1) with a=0 and b=0; ii) 24 LQ model modifications using SF1 (Eq 1) with dep_a and dep_b equal to all the combinations with 1/2, 2/3, 1 or 2; and iii) the LQL model, using SF2 (Eq 2).

log (SF1)= n.*d.*alpha.*(1+a.*d.^dep_a+d.*(1+b.*d.^dep_b)/alphabeta)-lambda*max(0,t_trat-tk) (1)

log(SF2)= n.*alpha.*(d+2.*(a.*d+exp(-a.*d)-1)./(alphabeta.*a^2))-lambda*max(0,t_trat-tk) (2)

Detailed information about the other parameters of the model is included in the manuscript.

3. Code

The code consists of a set of Matlab functions to fit RT dose response clinical data to several tumor control probability (TCP) models. Regarding the TCP models, the user can choose between two TCP formulations, Logistic and Poisson. Both TCP formulations use an expression for the tumor cells surviving fraction (SF model). Fits are performed using 26 SF models with different functional dependencies in order to find the SF model that provides the best data fitting.

Four main functions are required to perform four fitting studies. All these functions use a simulated annealing algorithm (SA algorithm) to find the best-fitting parameters that maximize the likelihood of the fit

1) repeatability_study.m: This function performs n optimizations to fit the clinical dose response data to 26 Poisson or logistic TCP models. The function optimizes the value of 4 parameters to maximize the loglikelihood associated to the fit.

2) dose_variation_study.m: This function performs 200 optimizations to fit the clinical dose response data to 26 logistic TCP models using 20 arrays with perturbed values of the dose per fraction associated to the treatment schedules included the clinical database. 10 TCP optimizations are run using each dose perturbation, and the best one is selected (in order to minimize the possibility of partial convergence of the SA algorithm). This is done to analyse the sensitivity of the fits to uncertainty in the database, in this case dose uncertainty.

3) probability_variation_study.m: This function performs 200 repeated optimizations to fit the clinical dose response data to 26 logistic TCP models using 20 arrays with perturbed values of the tumor control probability associated to the clinical database. 10 TCP optimizations are run using each probability perturbation, and the best one is selected (in order to minimize the possibility of partial convergence of the SA algorithm). This is done to analyse the sensitivity of the fits to uncertainty in the database, in this case TCP uncertainty.

4) dose_and_prob_variation_study.m: This function performs 500 repeated optimizations to fit the clinical dose response data to 26 logistic TCP models using 50 arrays with perturbed values of the dose per fraction and 50 arrays with perturbed values of the tumor control in the clinical database. 10 TCP optimizations are run using each dose and probability simultaneaous perturbation, and the best one is selected (in order to minimize the possibility of partial convergence of the SA algorithm). This is done to simultaneously analyse the sensitivity of the fits to two sources of uncertainty in the database.

The four main functions used in the simulation studies call several functions to perform the fitting optimizations:

1) Maxlikelihood_TCP.m: Simulated annealing optimization algorithm that finds the TCP model parameter values that maximize the loglikelihood associated to a set of dose response TCP fits using different expressions for the surviving fraction, SF.

2) Loglikelihood_Poisson.m: This function calculates the loglikelihood associated to the fit of a set of dose response data to a Poisson TCP model.

3) Loglikelihood_logit.m: This function calculates the loglikelihood associated to the fit of a set of dose response data to a Logistic TCP model.

4) EQD2_calculator.m: This function calculates the issoeffective total dose corresponding to a conventional

fractionation regimen using 2 Gy per fraction.

5) weight_calculator.m: This function calculates the Akaike weights associated to the models using the different expressions for the SF, with respect to the linear quadratic model.

More information on the purpose of each function, including information about the inputs and outputs, is presented within the body of each function.

4. Results

The results corresponding to the four fitting studies are in matlab format (.mat files) in two folders: **results_brain** and **results_lung**, for the brain and NSCLC databases, respectively. Results from each simulation study are structured as follows:

4.1 One Matlab structure containing n structures named "simn" with information about the bestfitting TCP models calculated in each of the n repeated optimizations performed.

Each sim n structure contains 7 arrays:

i) bp: 26x8 array containing the values of 8 parameters associated to the 26 different TCP models;
ii) aic: 26x1 array containing the AICc values corresponding to the fit to 26 TCP models; iii) loglike: 26x1 array containing the loglikelihood associated to the fit to 26 TCP models; iv) TCPmodel: 26xcohorts array with the TCP values associated to each of the 26 TCP models for each of the cohorts (fractionation regimens) included in the database.
v) EQD2: a 26xcohorts array with the EQD2 values associated to each of the 26 TCP models for each of the cohorts (fractionation regimens) included in the database.
vi) Effect: a 26xcohorts array with the treatment effects associated to each of the 26 TCP models for each of the cohorts included in the database;
vii) prob: a 26xcohorts array with the fitting probability associated to the fit of each of the 26 TCP models for each of the cohorts' data points included in the database. The number of cohorts included in the database is 55 for brain and 61 for lung, respectively.

4.2 One results array: Results are summarized in a (n+8)x26 results array, where n is the number of optimizations run in each study for 26 radiobiological models. Some lung studies, for which the LQL model was no longer tested, have (n+8)x25 results arrays. The information provided in these arrays is:

i) rows 1 to n: AICc values achieved with every model in each of the n repeated optimization processes. In dose and TCP perturbation studies, where 20 (or 50) perturbations are performed and 10 optimizations are run per perturbation, rows 1 to 20 (or 50) show the bestfitting AICc values achieved with every perturbation
ii) row n+1: relative standard deviation of the AICc values achieved for each model in n optimizations.
iii) row n+2: overall bestfiting (minimum) AICc values achieved for each model.
iv) row n+3: Akaike weights corresponding to the bestfitting models.
v) rows n+4 and n+5: Mean and standard deviation of the AICc values calculated in n optimizations.
vi) row n+6: Evidence ratios associated to each of the bestfitting models.
vii) row n+7 and n+8: Mean and standard deviation of the evidence ratios calculated in n optimizations.

4.3 Dose_vector/p_vector: Results from dose and TCP perturbation studies have a nxcohorts vector with the n perturbed values of dose and TCP used to fit the data. This vector is included in the output .mat file.