SUPPLEMENTARY MATERIAL 1

*Monte Carlo Sampling validation using Semi Analytical Method*

For the two NTCP models based on multivariable logistic regression (swallowing problems and tube feeding dependence) it is possible to combine all sample distributions into one single distribution, which is a validation of the Monte Carlo sampling method described in the materials and methods.

The linear part of the logistic regression formula is given by:

|  |  |  |
| --- | --- | --- |
|  | $$s=b\_{0}+b\_{1}⋅x\_{1}+…+b\_{n}x\_{n}$$ | (1) |

The mean values for both model and dose uncertainty are given by:

|  |  |  |
| --- | --- | --- |
|  | $$μ\_{s\_{model}}=μ\_{b\_{0}}+μ\_{b\_{1}}⋅x\_{1}+…+μ\_{b\_{n}}⋅x\_{n}$$ | (2) |
|  | $$μ\_{s\_{dose}}=b\_{0}+b\_{1}⋅μ\_{x\_{1}}+…+b\_{n}⋅μ\_{x\_{n}}$$ | (3) |

Where $b\_{i}$ represents the $i$th model-coefficient and $x\_{i}$ the patient dependent variable belonging to $b\_{i}$. The standard deviations are quadratically summed while a correlation terms is added in case of dose uncertainty:

|  |  |  |
| --- | --- | --- |
|  | $$SD\_{s\_{model}}=SD\_{b\_{0}}^{2}+\left(SD\_{b\_{1}}⋅x\_{1}\right)^{2}+…+\left(SD\_{b\_{n}}⋅x\_{n}\right)^{2}$$ | (4) |
|  | $$SD\_{s\_{dose}}=\sum\_{i}^{}\left(b\_{i}⋅SD\_{x\_{i}}\right)^{2}+2⋅\sum\_{i,j}^{}ρ\_{i,j}⋅\left(b\_{i}⋅SD\_{x\_{i}}\right)⋅\left(b\_{j}⋅SD\_{x\_{j}}\right)$$ | (5) |

Where $i$ and $j$ are identifiers of the dose variables. The contribution of the clinical variables is $0$ since the uncertainty for those is assumed to be negligible. Since $s$ is assumed to follow a normal distribution it could be proved that $e^{-s}\~ logN\left(-μ\_{s},SD\_{s}\right).$ The values of $e^{-s}$ are sampled from the derived lognormal distribution followed by the calculation of NTCP values for both the photon and proton plan. The results of this semi-analytical method are shown in figures A1 and A2 as a validation of the multiple-sample method. The accuracy levels for different ΔNTCP threshold show the same trend for both approaches.

Nevertheless, this semi-analytical method has not been used since it can not be applied for the model of Dijkema et al. which is not based on logistic regression, it is an LKB model. The coefficients are not linearly dependent. For simplicity we decided to use the Monte Carlo sampling method for all NTCP models.

**Figure A1:** *Selection accuracy due to model uncertainty validation using Semi-Analytical method.*



**Figure A2:** *Selection accuracy due to dose uncertainty validation using Semi-Analytical method. (SDdose = 3 Gy)*