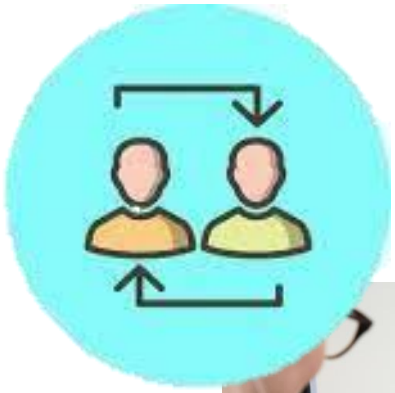
Maastro
in the best of hands

# Using prediction models in clinical practice
## *Introduction of an NTCP-model based selection approach for proton therapy in esophageal cancer*

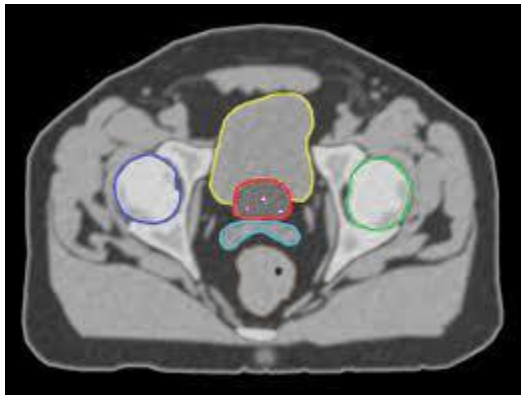
Maaike Berbée, Maastro

# Limited use of models in daily clinical practice

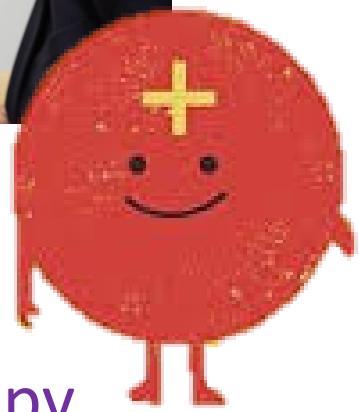

# Models in my daily practice



Shared decision making



Auto-contouring



Patient selection for proton therapy

# *Introduction of an NTCP-model based selection approach for proton therapy in esophageal cancer*

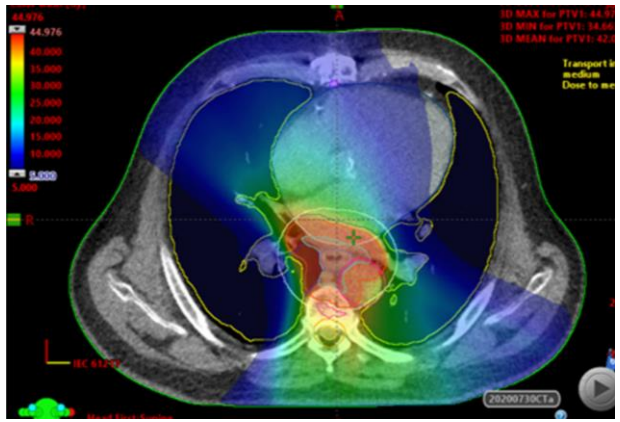
*Introduction of an NTCP-model based selection approach for proton therapy in esophageal cancer*

How to facilitate/improve the use of prediction models in clinical practice?

# Esophageal cancer

- Neo-adjuvant chemoradiotherapy -> overall survival ↑

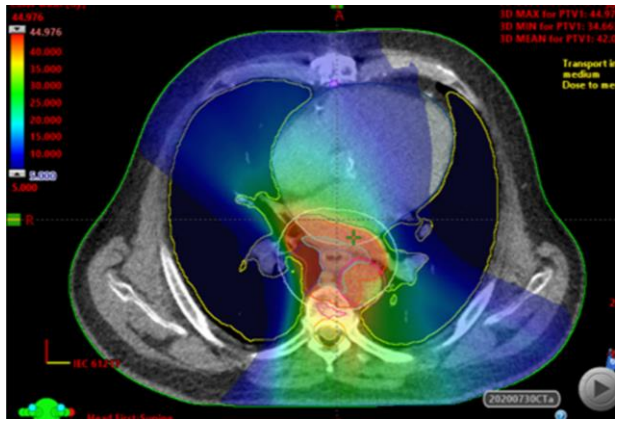
- Radiation exposure OARs -> risk of toxicity ↑
  risk of non-cancer related death ↑

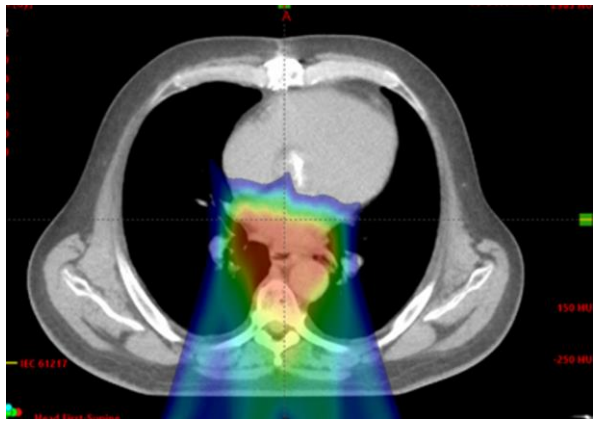


# Esophageal cancer

- Neo-adjuvant chemoradiotherapy -> overall survival ↑

- Radiation exposure OARs -> risk of toxicity ↑
  risk of non-cancer related death ↑





# The Netherlands: model based patient selection approach

Model-based approach: ↓ side effects

**Statement:**

Proton therapy is
cost-effective if good patient selection is performed
based on clinically relevant gain
(evidence based)

⇒ **Model-based indications**

→Validated models required, for clinically relevant endpoints

→Estimation of NTCP based on individual treatment plan → to allow plan comparison

*Ramaekers et al, 2013*

# Conditions model-based indications

Conditions for approval of a National indication protocol proton therapy (NIPP):

1. *Selection is based on a **clinically relevant outcome measure;***

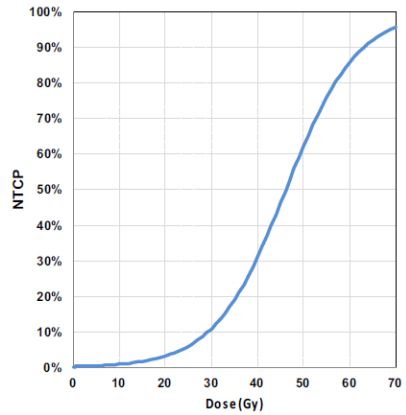
2. *Prediction models of **sufficient quality** are available for the determination of dose-volume-effect relationships;*

3. *It is possible to determine and provide insight into an estimate of the expected clinically relevant benefit (added value), based on a **planning comparison in each individual patient.***

# Model-based approach

**Step1**

- Selection of NTCP model



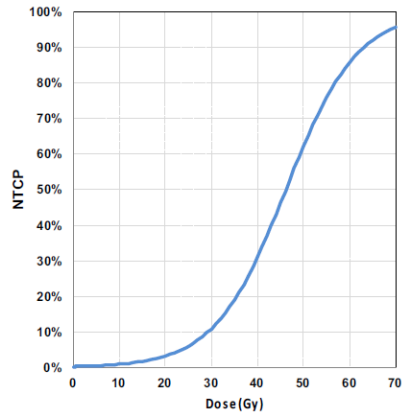
# Model-based approach

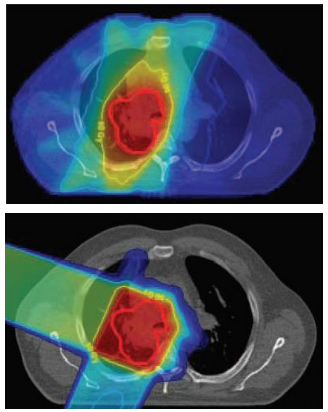
**Step1**

- Selection of NTCP model

**Step 2**

- Make photon & proton plan
- Calculate Δ dose



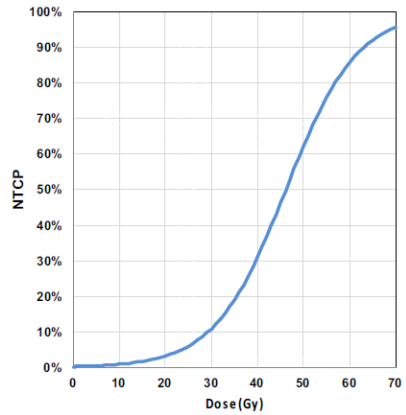



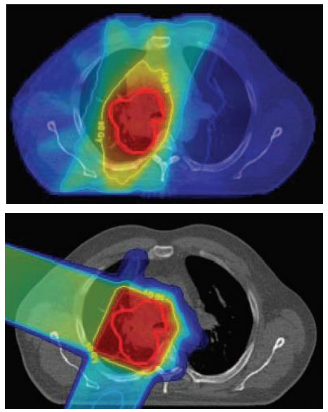
# Model-based approach

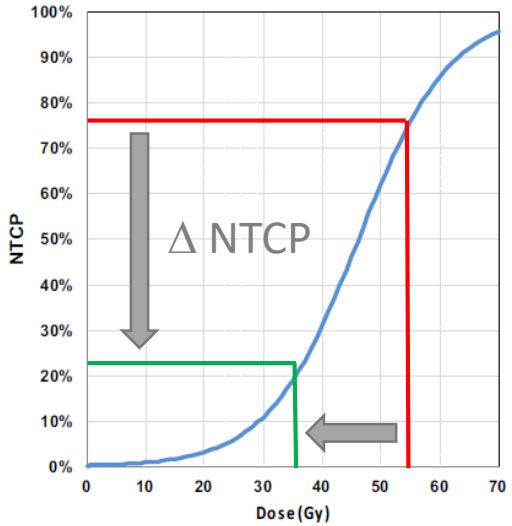
**Step1**
- Selection of NTCP model

**Step 2**
- Make photon & proton plan
- Calculate Δ dose

**Step 3**
- Translate Δ dose into Δ NTCP





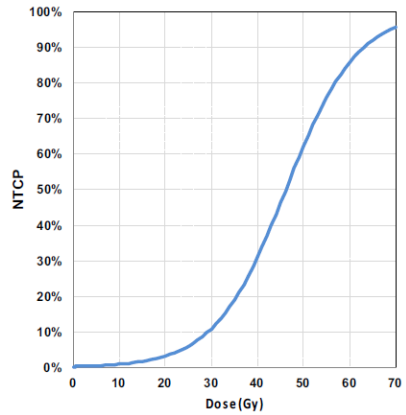
# Model-based approach

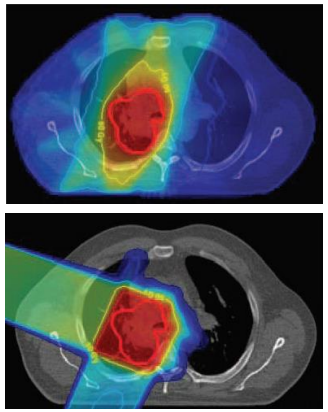
**Step1**
- Selection of NTCP model

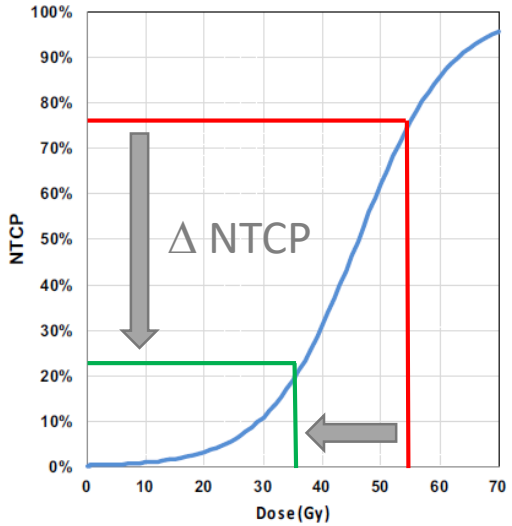
**Step 2**
- Make photon & proton plan
- Calculate Δ dose

**Step 3**
- Translate Δ dose into Δ NTCP

**Step4**
- Δ NTCP > x% → Treatment with protons
- Δ NTCP < x% → Treatment with photons



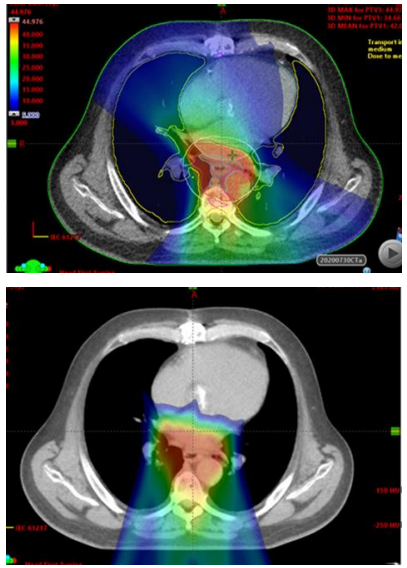


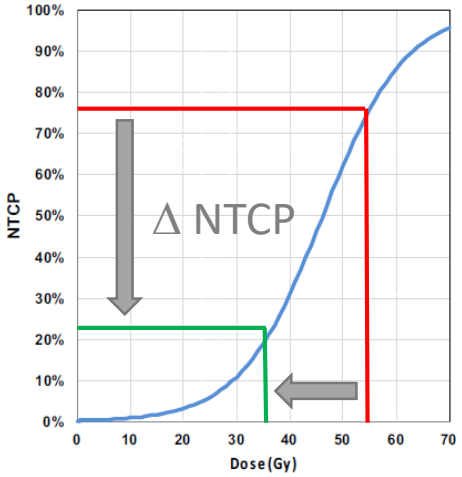
# Nationally approved thresholds for delta NTCP

| Severity of toxicity (Grade) | Delta NTCP- threshold |
|---|---|
| Mild toxicity (Grade 1) | - |
| Moderate toxicity (Grade 2) | ≥10% |
| Severe toxicity (Grade 3) | ≥ 5% |
| Life-threatening tox- death (Grade 4-5) | ≥2% |

**Obligation to prospectively record side-effects, to validate the models !**

# Proton therapy for esophageal cancer






Validated NTCP models are required!

# No suitable validated NTCP model for esophageal cancer

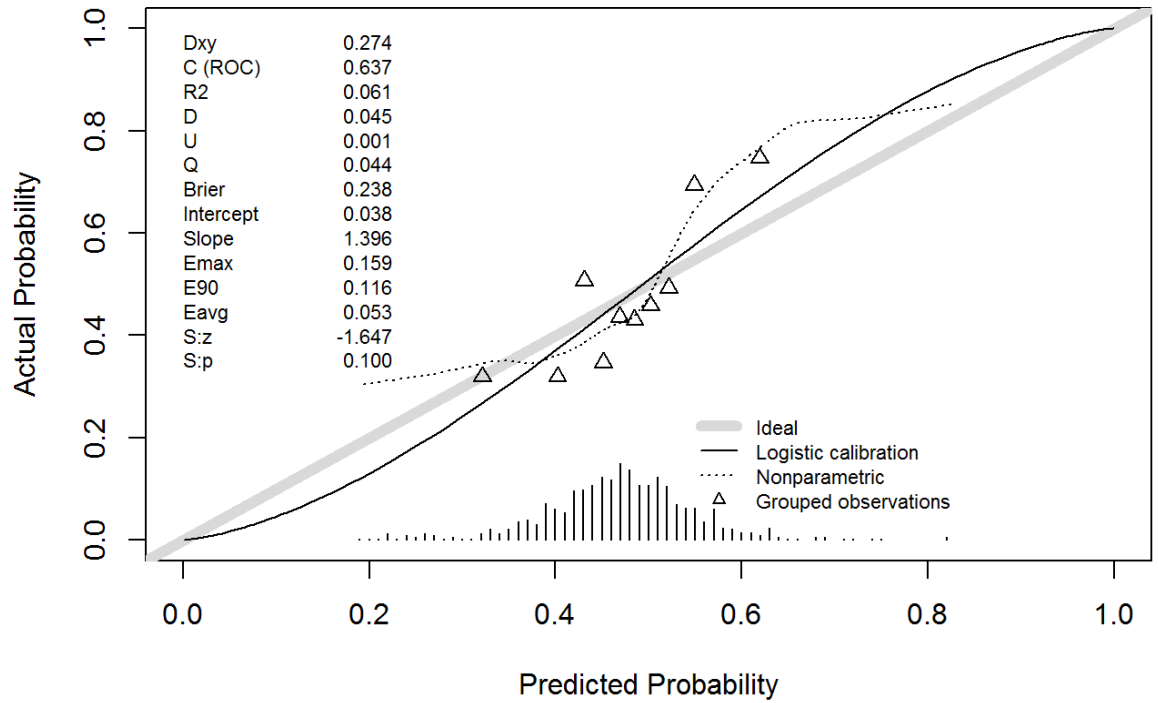

# Alternative: model for two-year mortality in lung cancer

| Defraene-Leuven | Dutch model |
| --- | --- |
| *Tumor volume* | *Tumor volume* |
| *Mean heart dose* | *Mean heart dose* |
| *Smoking (current)* | |

# Datasets: esophageal cancer patients

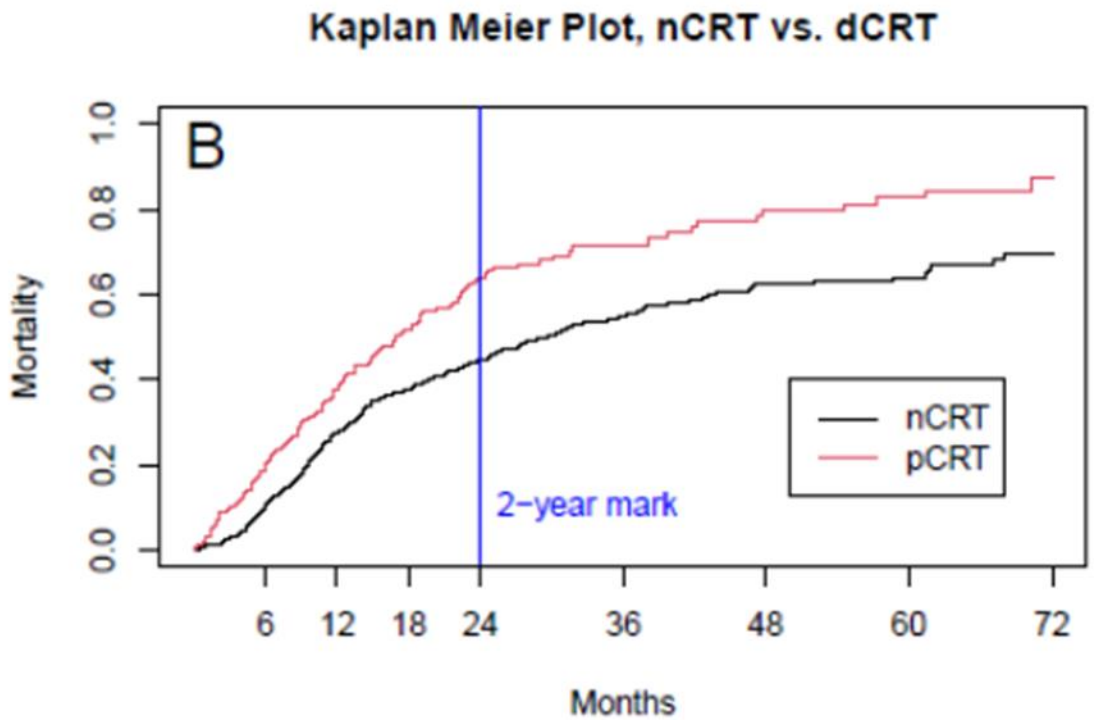
| | nCRT | dCRT | Total |
|---|---|---|---|
| UMCG | 224 | 70 | |
| Maastro | 248 | 86 | |
| AvL | 65 | | |
| *Total* | *537* | *156* | 693 |

# Validity in combined group….clinically relevant?



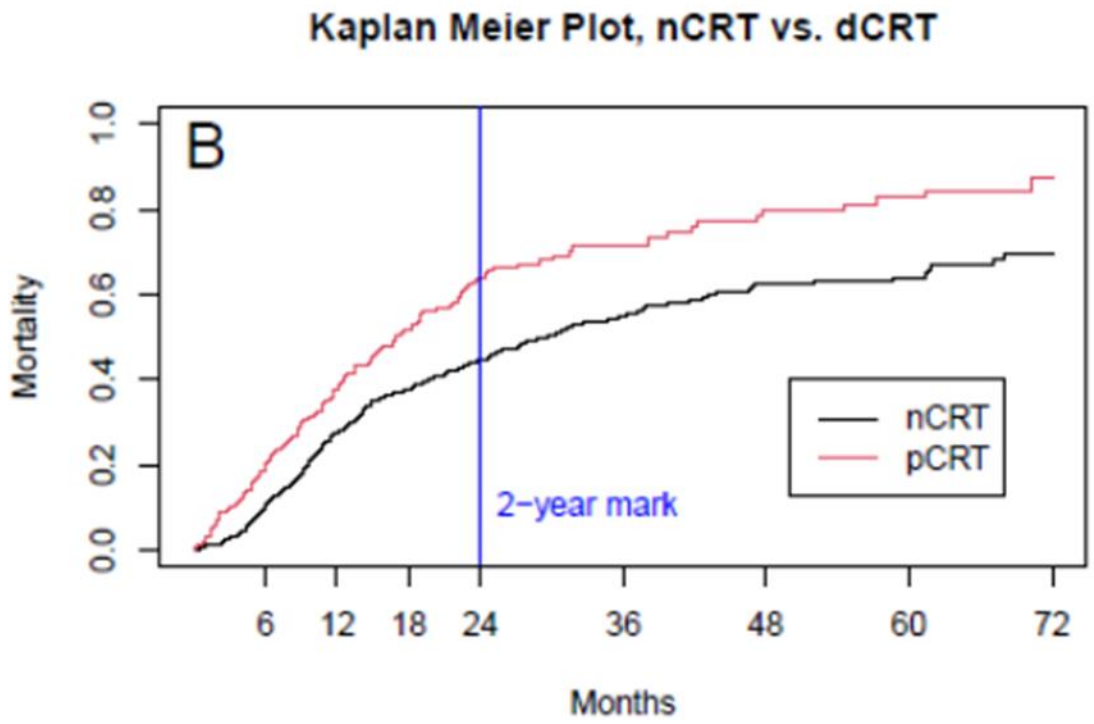

- According to the closed testing procedure, the adjustment of the intercept only is sufficient

# Differences in 2 year mortality between nCRT and dCRT group




| nCRT | dCRT | |
|---|---|---|
| 45% | 64% | 2-year mortality |

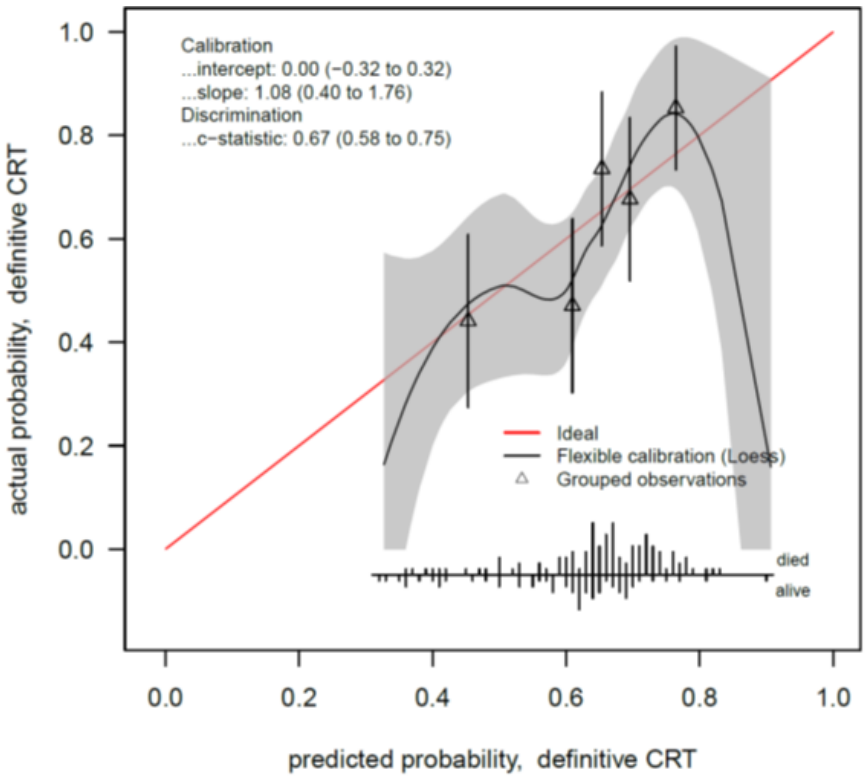
# Differences in 2 year mortality between nCRT and dCRT group




| nCRT | dCRT | |
|---|---|---|
| 45% | 64% | 2-year mortality |

- *Separate validation in dCRT and nCRT cohort*
- *Intercept adjustment before external validation*

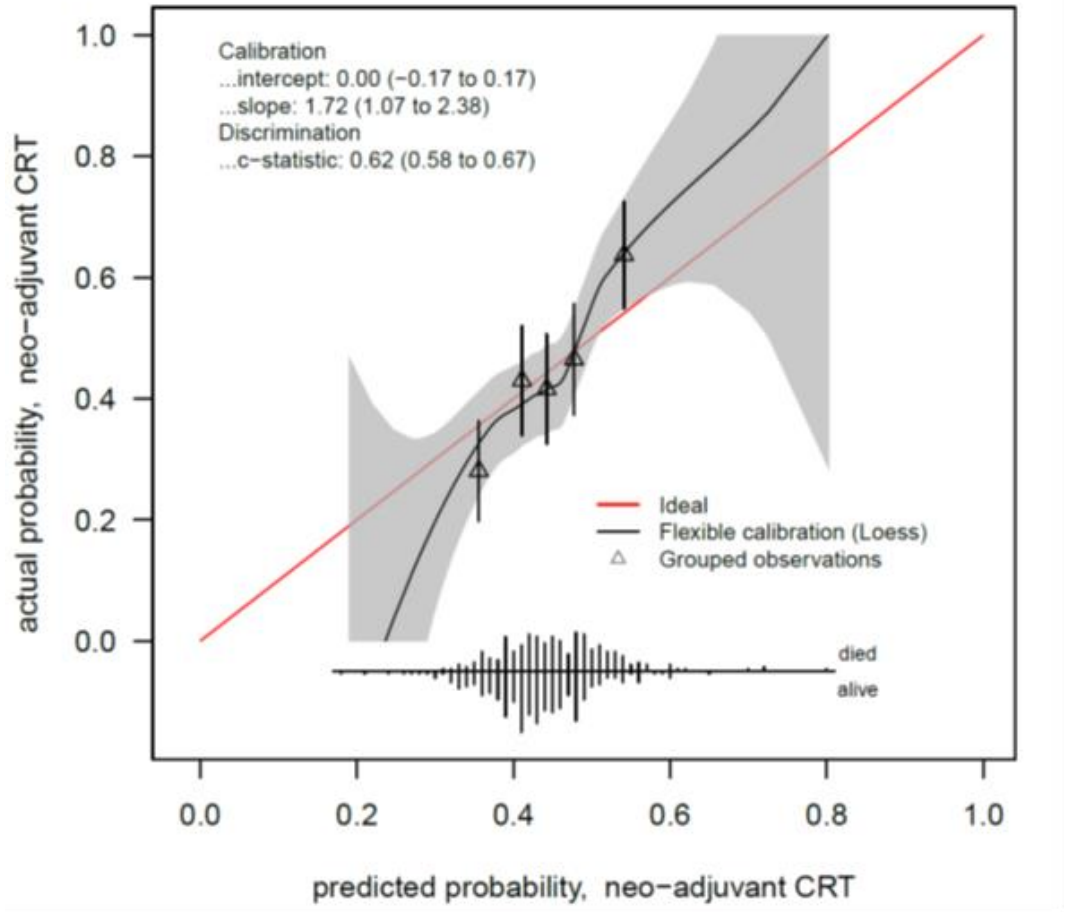
# Definitive chemo-radiotherapy



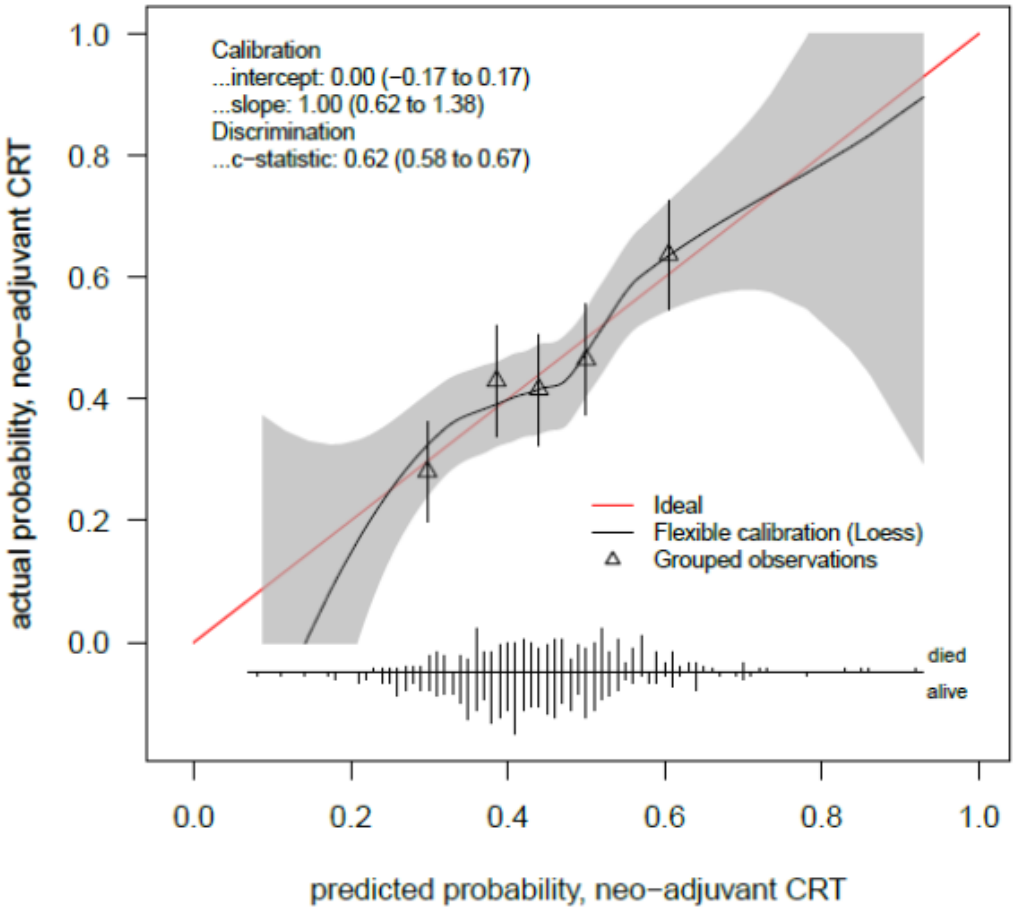

- Adjustment of intercept

| | Original | dCRT |
|---|---|---|
| Intercept | -1.3409 | -1.0421 |
| √ gross tumor volume ($cm^3$) | 0.059 | 0.059 |
| √ mean heart dose (Gy) | 0.263 | 0.263 |

# Neo-adjuvant chemoradiotherapy



# Neo-adjuvant chemoradiotherapy



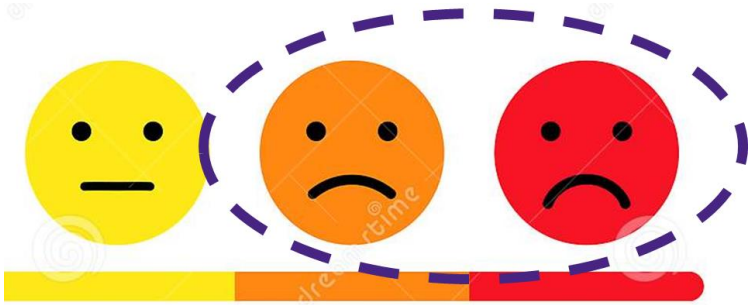

## Recalibration

| | Original | dCRT | nCRT |
|---|---|---|---|
| Intercept | -1.3409 | -1.0421 | -3.0818 |
| √ gross tumor volume ($cm^3$) | 0.059 | 0.059 | 0.1016 |
| √ mean heart dose (Gy) | 0.263 | 0.263 | 0.4529 |

# 2 NTCP models: everybody happy?

# 2 NTCP models: everybody happy?

# 2 NTCP models: everybody happy?



- Independent association MHD?

- Surrogate endpoint, not a real toxicity

- Many prognostic factors are not taken into account:
  - Age, TNM status, WHO performance status etc

# Exploratory analyses: Multivariable analysis

| | nCRT | dCRT |
|---|---|---|
| √ gross tumor volume ($cm^3$) | 1.13 (1.06 to 1.20)<br>*p < 0.001* | 1.07 (0.97 to 1.20)<br>*p = 0.19* |
| √ mean heart dose (Gy) | 1.43 (1.06 to 1.95)<br>*p = 0.02* | 1.31 (1.03 to 1.70)<br>*p = 0.03* |

Both the tumor volume and mean heart dose were associated to 2-year mortality

# Exploratory analyses
# Incremental value of the MHD on top of other known predictors

- Is the MHD associated with 2y-mortality after adjustment for other prognostic factors? (multivariable analysis)

- Does the MDH have additional value on top of multivariable models with an increasing number of predictors? (likelihood ratio test, IDI)

# Exploratory analyses
# Incremental value of the MHD on top of other known predictors

- Is the MHD associated with 2y-mortality after adjustment for other prognostic factors? (multivariable analysis)

  Combined *population: +*

  *nCRT/dCRT seperately: +/-*

- Does the MDH have additional value on top of multivariable models with an increasing number of predictors? (likelihood ratio test, IDI)

  Combined polulation *population: +*

  *nCRT/dCRT seperately: +/-*

# Surrogate endpoint, other prognostic factors

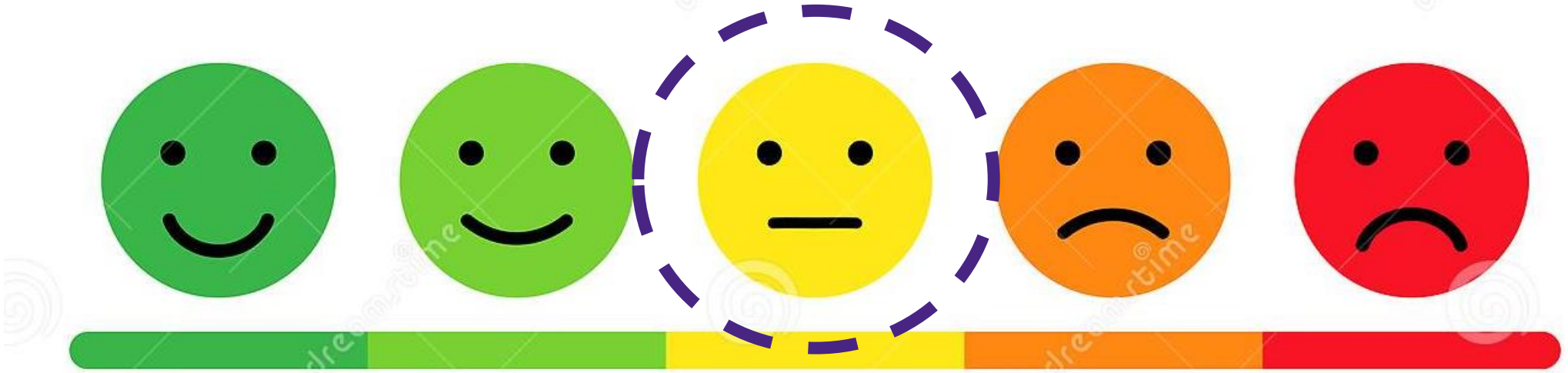
Additional selection criteria

- WHO 0-1
- No T4
- No N3

Delta NTCP > 5%

- Development of models for specific toxicities

# 2 NTCP models: everybody happy?



National implementation of the protocol

# Lessons I learned from this project

- Relevant endpoint?
- Relevant prognostic factors?
- Other factors?
- Development and validation in the right population?
- Model performance

# Lessons I learned from this project

- Relevant endpoint?
- Relevant prognostic factors?
- Other factors?
- Development and validation in the right population?
- Model performance

# Clinical practice can benefit from model implementation

# How to facilitate/improve the use of prediction models in clinical practice?

# How to facilitate/improve the use of prediction models in clinical practice?

- *Data scientists and computer programmers AND clinicians need to collaborate*

# How to facilitate/improve the use of prediction models in clinical practice?

- Data scientists and computer programmers AND clinicians need to collaborate
- *Good models need good predictors*

# How to facilitate/improve the use of prediction models in clinical practice?

- Data scientists and computer programmers AND clinicians need to collaborate
- Good models need good predictors
- *Good predictors still need good modelling practices and a lot of tough external validation*

# Steyerberg et al recommends against SPLIT validation

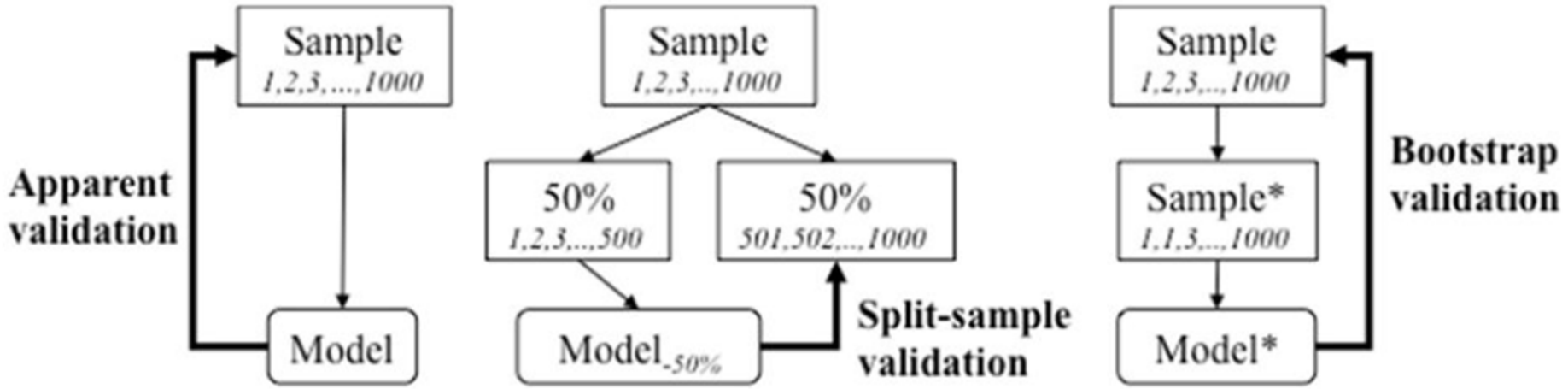
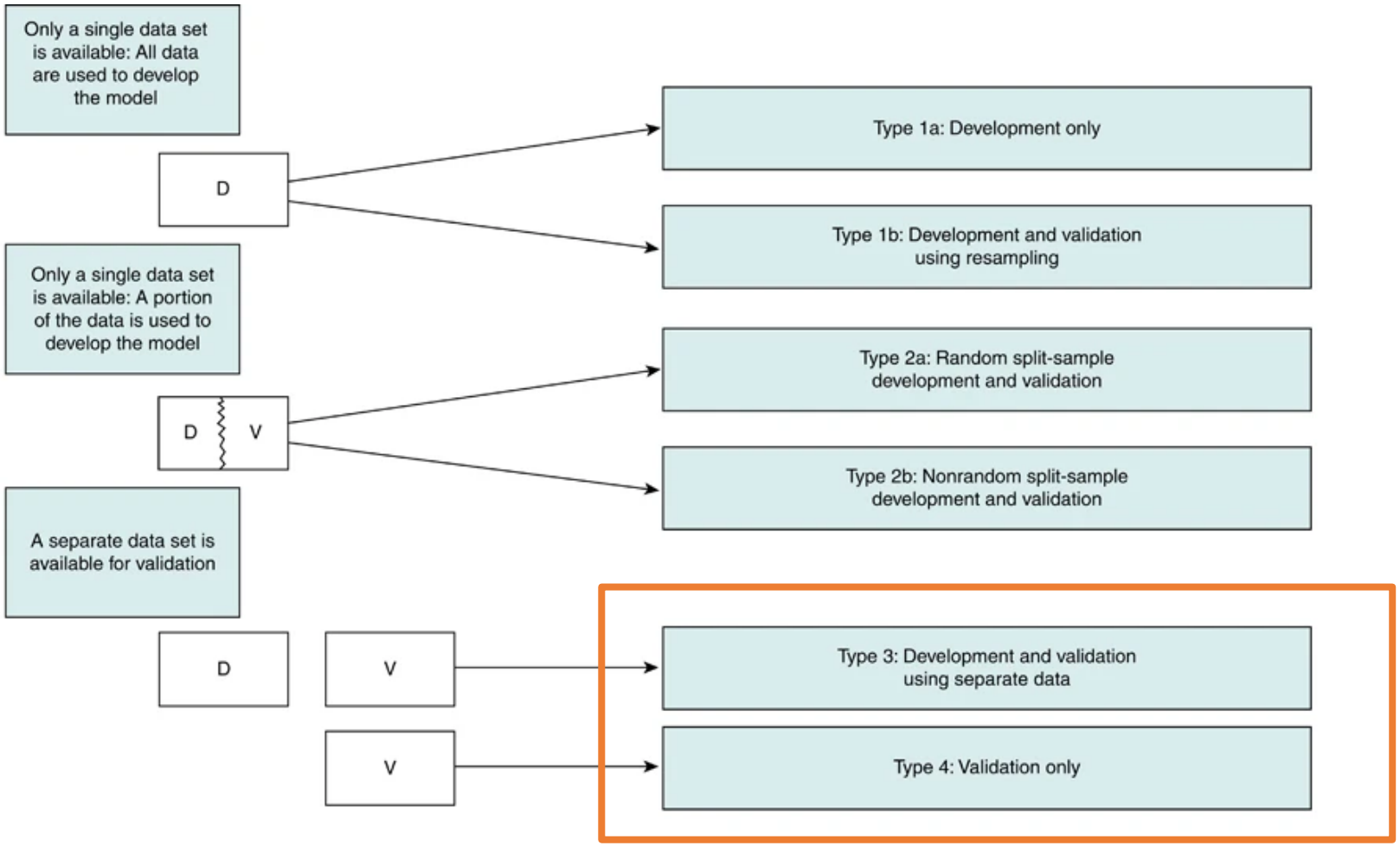



**Figure 1.**
Schematic representation of apparent, split-sample, and bootstrap validation. Suppose we have a development sample of 1,000 subjects (numbered 1,2,3,..1000). Apparent validation assesses performance of a model estimated in these 1000 subjects on the sample. Split-sample validation may consider 50% for model development, and 50% for validation. Bootstrapping involves sampling with replacement (e.g., subject number 1 is drawn twice, number 2 is out, etcetera), with validation of the model developed in the bootstrap sample (Sample*) in the original sample.

# TRIPOD guidelines advise strong validation



# How to facilitate/improve the use of prediction models in clinical practice?

- Data scientists and computer programmers AND clinicians need to collaborate
- Good models need good predictors
- Good predictors still need good modelling practices and a lot of tough external validation
- *In the end…..it might need an RCT to prove that using a predictive model really does give the desired benefit and cost-effectiveness in clinical use*

# Take home messages

- Limited use of predictive models in clinical practice
- Clinical practice can benefit from model implementation
- Data scientist and clinicians need to work at the same table

# Questions?

UMCG
- C.T. Muijs
- E. Oldehinkel
- A. van der Schaaf

UMCU/ Julius Center
- E. Schuit
- J.B. Reitsma

NKI/AvL
- F.E.M. Voncken

Maastro
- L. Wee