A prediction model of heterogeneous treatment effects using randomized and observational data

Konstantina Chalkou, Salanti Georgia – University of Bern, Fabio Pellegrini – Biogen International GmbH, Suvitha Subramaniam, Benkert Pascal - University of Basel

 This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement № 825162.

Background

Motivation - Effectiveness of drugs in Relapsing-Remitting Multiple Sclerosis (MS)

- Several drugs, compared in Network Meta-Analyses (NMA)
  #not personalized predictions
  - Tramacere I. et al., 2015
- We focus on ***Dimethyl Fumarate***, ***Glatiramer Acetate***, and ***Natalizumab***
- Outcome: ***Relapse MS in 2 years*** (Yes/No)
- We want to find the drug that minimizes the risk of relapse, subject to patient characteristics: **H**eterogeneous **T**reatment **E**ffects

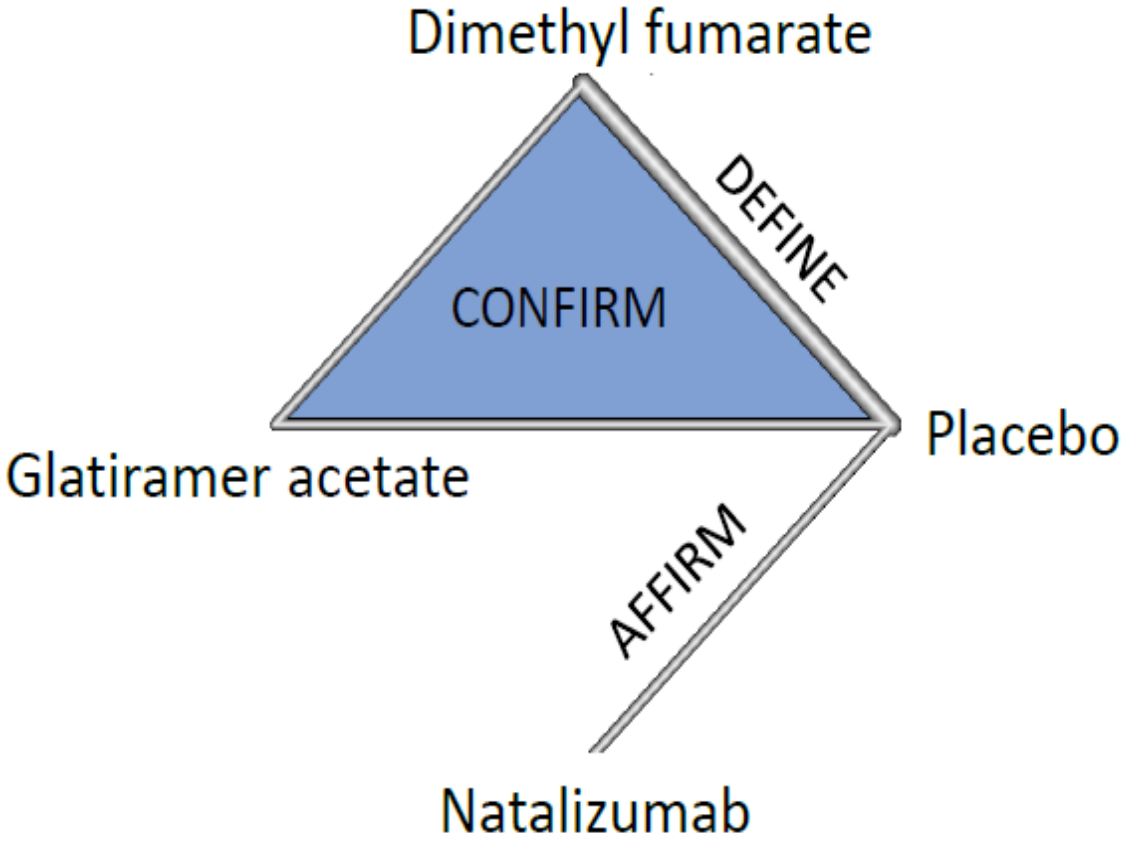
Aim

To develop a ***three-stage*** evidence synthesis ***prediction model*** to predict the most likely outcome under several possible treatment options while accounting for patients' characteristics using ***randomized clinical trials*** and ***observational data***

Data

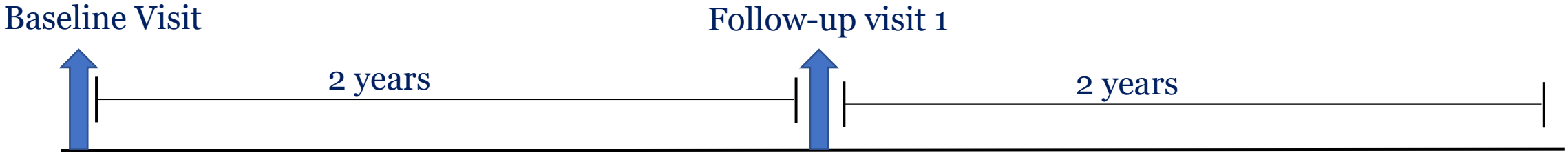
RCTs

- 3 randomized clinical trials (phase III), 2990 observations in total
- **Disease:** Relapsing-remitting Multiple Sclerosis (MS)
- **Outcome:** Relapse MS in 2 years

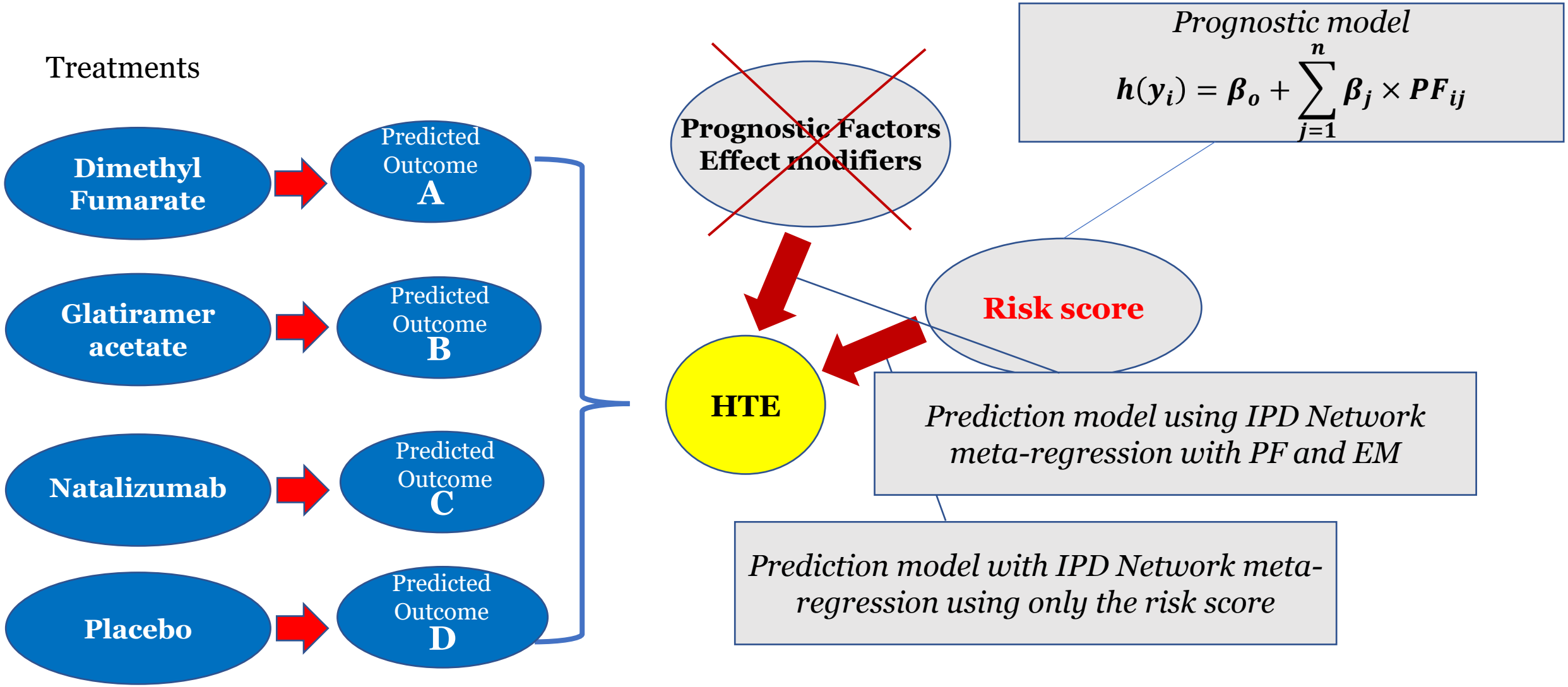


Observational data – Swiss MS Cohort (SMSC)

- **Inclusion criteria:** Patients with confirmed RRMS and at least two-year follow-up period from the baseline visit date
- **Patients:** 935 patients, each one with 1, 2, or 3 treatment cycles (i.e. repeated measures)



- **Observations:** 1752 follow-up cycles


Treatments
Dimethyl Fumarate
Glatiramer acetate
Natalizumab
Placebo
Predicted Outcome A
Predicted Outcome B
Predicted Outcome C
Predicted Outcome D
Prognostic Factors
Effect modifiers
HTE
Prognostic model
$h(y_i) = \beta_o + \sum_{j=1}^{n} \beta_j \times PF_{ij}$
Risk score
Prediction model using IPD Network meta-regression with PF and EM
Prediction model with IPD Network meta-regression using only the risk score

IPD from Observational Data SMSC

Treatments

Dimethyl Fumarate → Predicted Outcome **A**

Glatiramer acetate → Predicted Outcome **B**

Natalizumab → Predicted Outcome **C**

Placebo → Predicted Outcome **D**

HTE

IPD from RCTs

Prognostic model

$$h(y_i) = \beta_o + \sum_{j=1}^{n} \beta_j \times PF_{ij}$$

#STAGE1

Risk score

#STAGE2

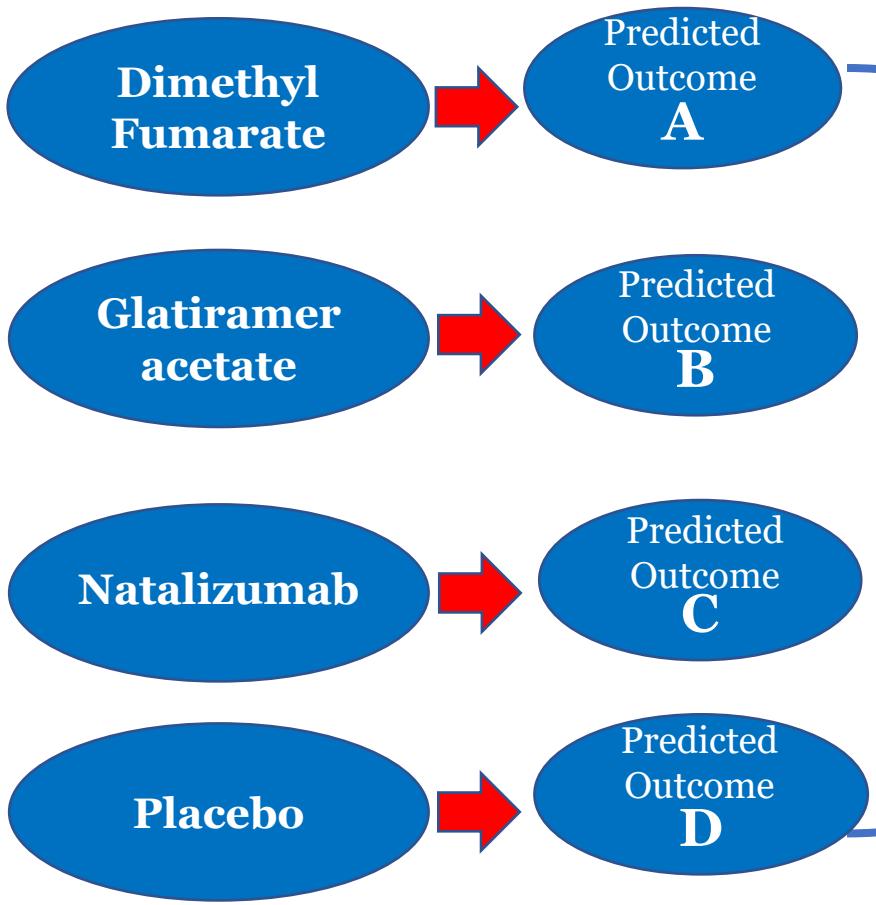
Re-calibration and re-estimation of the developed risk model and calculation of the new risk score in the RCTs dataset

Prediction model with IPD Network meta-regression using only the risk score

#STAGE3

IPD from Observational Data SMSC

Treatments



HTE

IPD from RCTs

Prognostic model

$$h(y_i) = \beta_o + \sum_{j=1}^{n} \beta_j \times PF_{ij}$$

#STAGE1

Risk score

#STAGE2

Re-calibration and re-estimation of the developed risk model and calculation of the new risk score in the RCTs dataset

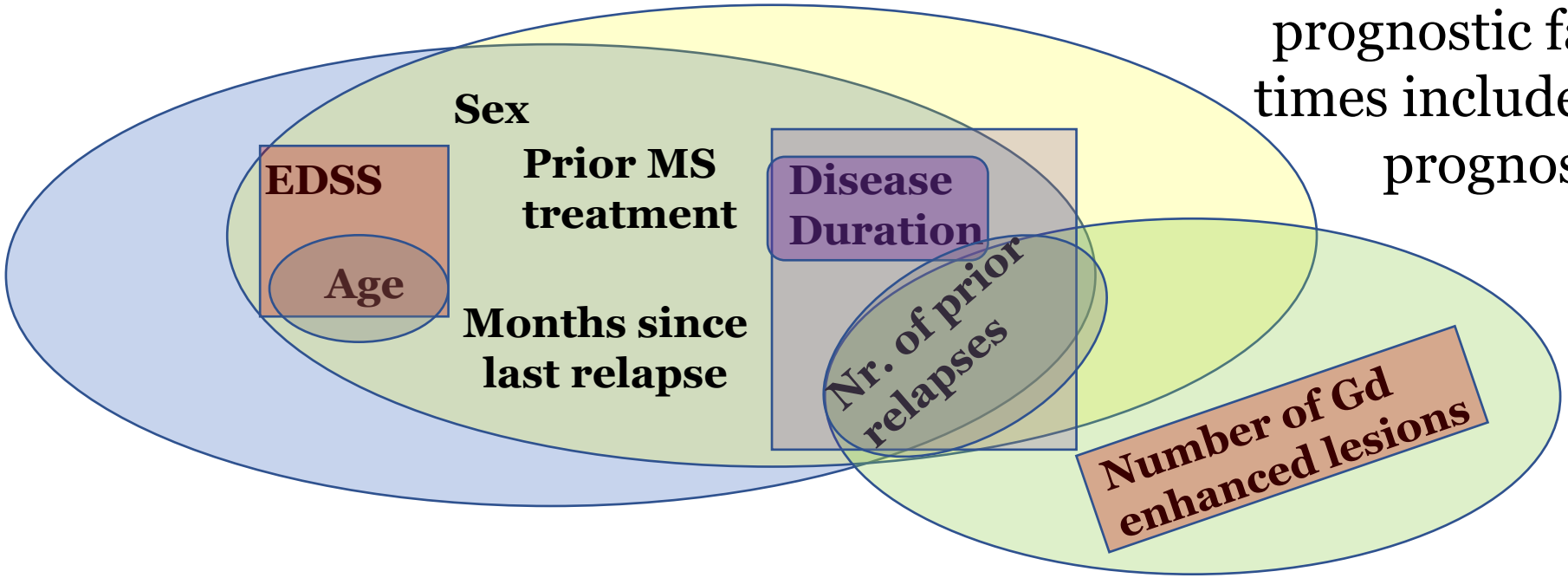
Prediction model with IPD Network meta-regression using only the risk score

#STAGE3

Stage 1: Development of the prognostic model

Selection of prognostic factors

8 previously identified prognostic factors (at least 2 times included in pre-existing prognostic models)



Pellegrini | Stühler | Cree | Held | Sormani | Signori | Kalincik

European Union's Horizon 2020 research and innovation programme under grant agreement № 825162.

Stage 1: Development of the prognostic model

Generalized linear mixed effects model – Bayesian framework

$$Y_{ij} \sim \text{Bernoulli}(R_{ij})$$

$$\text{logit}(R_{ij}) = \beta_0 + u_{oi} + \sum_{k=1}^{P} (\beta_k + u_{ki}) \times PF_{k,j}$$

Notation

i: individuals, where $i = 1, 2, \ldots, N$

j: time point, where $j = 1, 2, 3$

$PF_{k,j}$: kth prognostic factor at j^{th} time point,

where $k = 1, 2, \ldots, P$

β_0: fixed effect intercept

u_{0i}: random effect intercept

β_k: fixed effect slopes of k^{th} prognostic factor

u_{ki}: the individual-level random slopes of k^{th} prognostic factor

Stage 1: Development of the prognostic model

Shrinkage

Bayesian shrinkage methods use a prior on the regression coefficients

- O'Hara et al., 2009

Laplace prior distributions for regression coefficients

- Genkin et al., 2007

$$\pi(\beta) = \prod_{k=1}^{p} \frac{\lambda}{2} e^{-\lambda|\beta_k|}, \quad p: number\ of\ regression\ coefficients$$

Small coefficients → **towards zero faster**

Large coefficients → **smaller shrinkage**

Stage 1: Development of the prognostic model

Missing data - Multilevel Joint Modelling Multiple Imputations

4 prognostic factors with missing data

Steps

1. Check for **auxiliary variables** – **1 variable was identified and used as auxiliary**

2. Creation of 10 imputed datasets – Use of the same model (i.e. the substantive one) to impute the datasets – `mitml` R-package

$$Y_{1ij} = \beta_0 + u_{oi} + \sum_{k=1}^{P}(\beta_k + u_{ki}) \times X_{k,j}$$

$$Y_{2ij} = \beta_0 + u_{oi} + \sum_{k=1}^{P}(\beta_k + u_{ki}) \times X_{k,j}$$

Y_{1ij} and Y_{2ij} factors with missing values,
$X_{k,j}$ complete factors used in the substantive model & auxiliary variables,
Use of random intercept (u_{oi}) and random slope (u_{ki}) as in the substantive model

Stage 1: Development of the prognostic model

Missing data - Multilevel Joint Modelling Multiple Imputations

4 prognostic factors with missing data

Steps

3. Application of the Bayesian model to all 10 imputed datasets

4. Pooled estimates via Rubin's rules for m imputed datasets

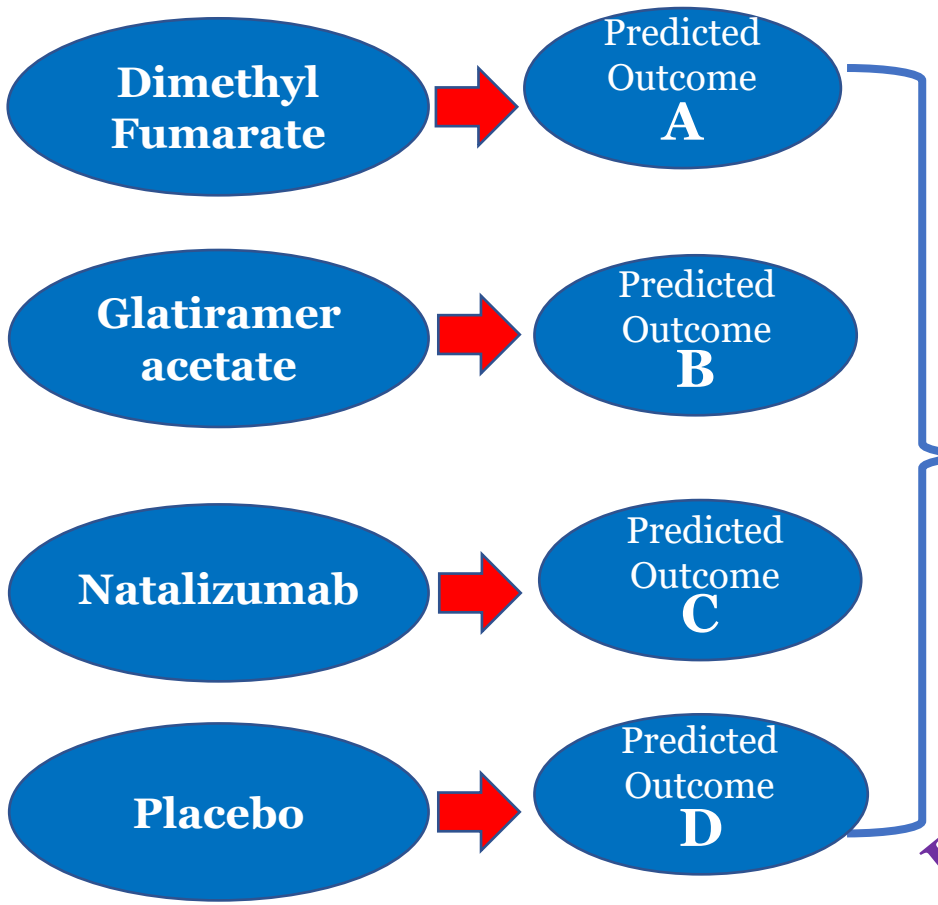
Stage 1: Development of the prognostic model

Estimated coefficients

Prognostic factors	Estimations
Intercept	-2.25
Age	-0.04
Disease Duration	0.36
Edss	0.12
Gd enhanced lesions	0.00
Number of previous Relapses (1 vs 0)	-0.08
Number of previous Relapses (more than 2 vs 0)	0.15
MonthsSinceRelapse	-0.45
Treatment Naive	0.15
Gender	0.28
Sigma	0.04

IPD from Observational Data SMSC

Treatments



Prognostic model

$$h(y_i) = \beta_o + \sum_{j=1}^{n} \beta_j \times PF_{ij}$$

#STAGE1

Risk score

HTE

IPD from RCTs

#STAGE2

Re-calibration and re-estimation of the developed risk model and calculation of the new risk score in the RCTs dataset

Prediction model with IPD Network meta-regression using only the risk score

#STAGE3

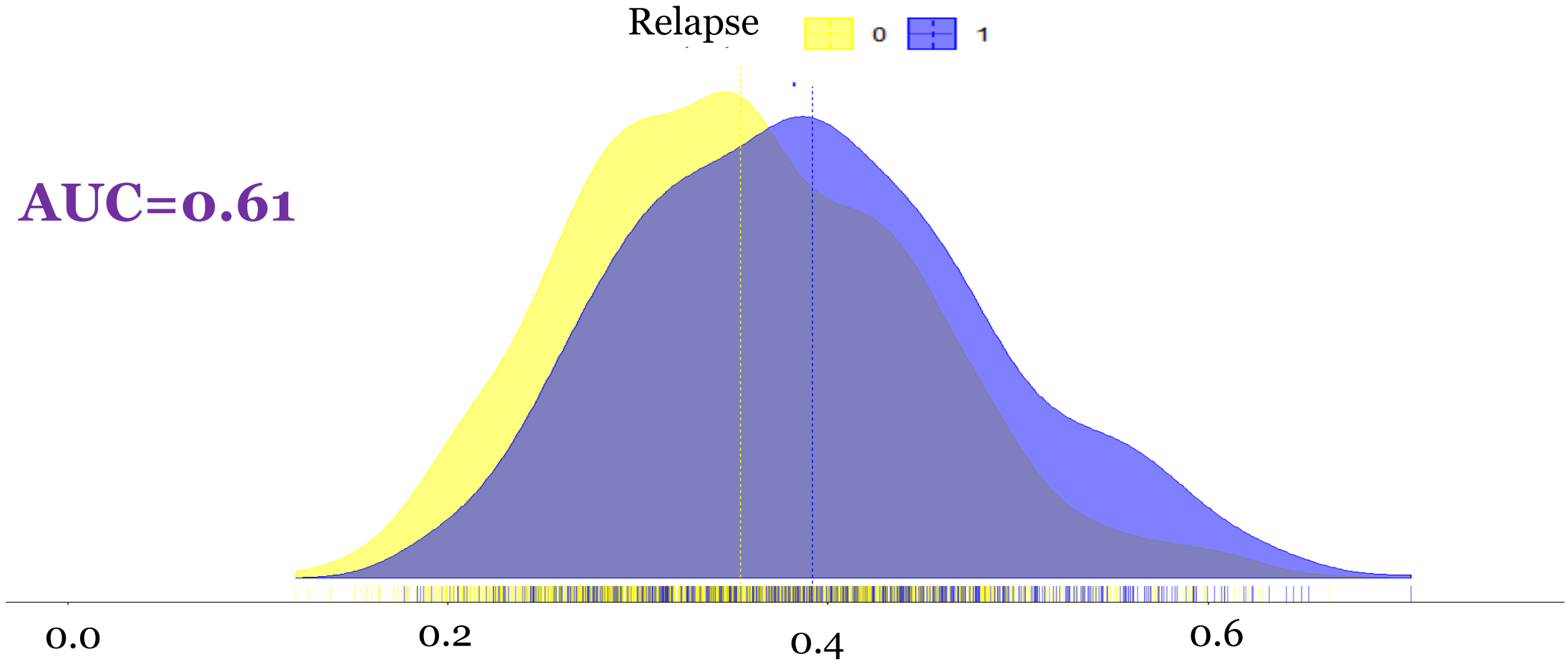
Stage 2: Re-calibration and re-estimation of the risk model to RCTs

Aim

To update the model to improve predictions for new patients from the new setting (i.e. RCTs)

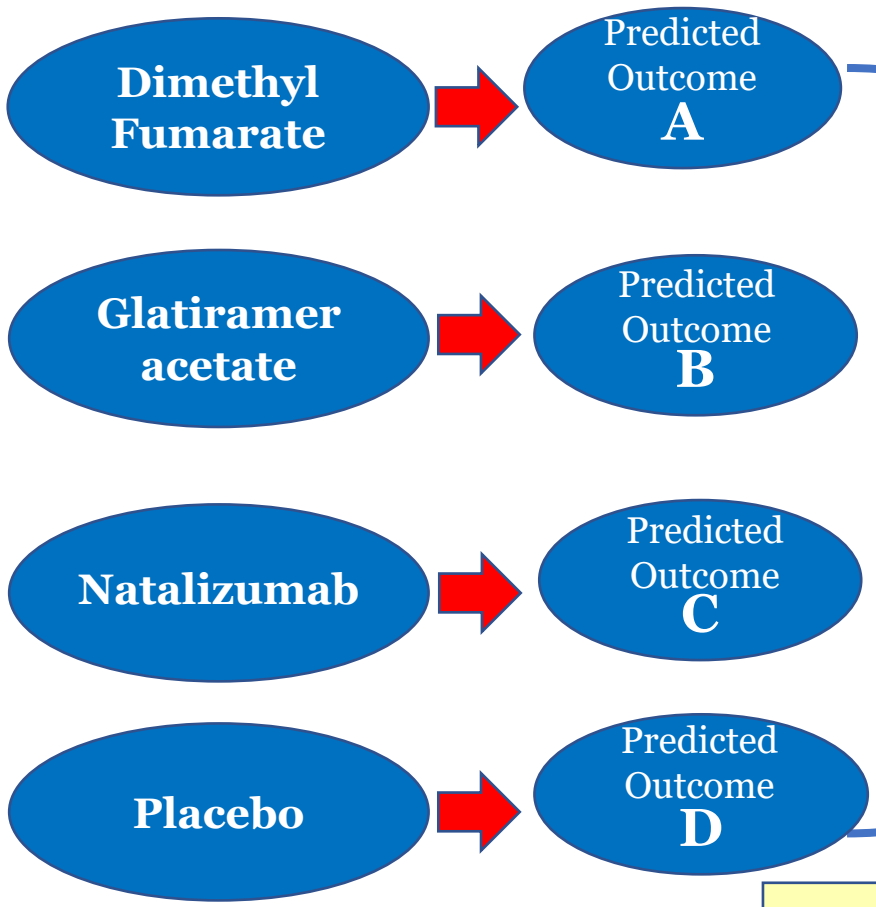
Methods	AUC
No Update	0.57
Update only the intercept (Re-calibration)	0.50
Update intercept and coefficients (Re-calibration)	0.57
Model revision (Re-calibration & selective re-estimation)	0.61

Risk of relapse in two years in RCTs



IPD from Observational Data SMSC

Treatments



Prognostic model

$$h(y_i) = \beta_o + \sum_{j=1}^{n} \beta_j \times PF_{ij}$$

#STAGE1

Risk score

HTE

IPD from RCTs

#STAGE2

Re-calibration and re-estimation of the developed risk model and calculation of the new risk score in the RCTs dataset

Prediction model with IPD Network meta-regression using only the risk score

#STAGE3

Stage 3: IPD Network Meta-regression

$$Y_{ijk} \sim Bernoulli(p_{ijk})$$

$$logit(p_{ijk}) = \begin{cases} u_j + B \times (logitR_{ij} - \overline{logitR_j}) if \ \ k = b_j \\ u_j + D_{b_j k} + B \times (logitR_{ij} - \overline{logitR_j}) + G_{b_j k} \times (logitR_{ij} - \overline{logitR_j}), \quad if\ k \neq b_j \end{cases}$$

Notation

i: Individuals

j: study

k: treatment

b_j: baseline treatment in study j

B: Individual level covariate regression term for Risk / the impact of Risk as prognostic factor

$D_{b_j k}$: the treatment effect of treatment k versus placebo / ***fixed effect***

$G_{b_j k}$: The interaction of treatment and risk. Different for each treatment vs study's control / the impact of Risk as effect modifier

Saramago et al., 2012

Stage 3: IPD Network Meta-regression

Results: Estimation of model parameters

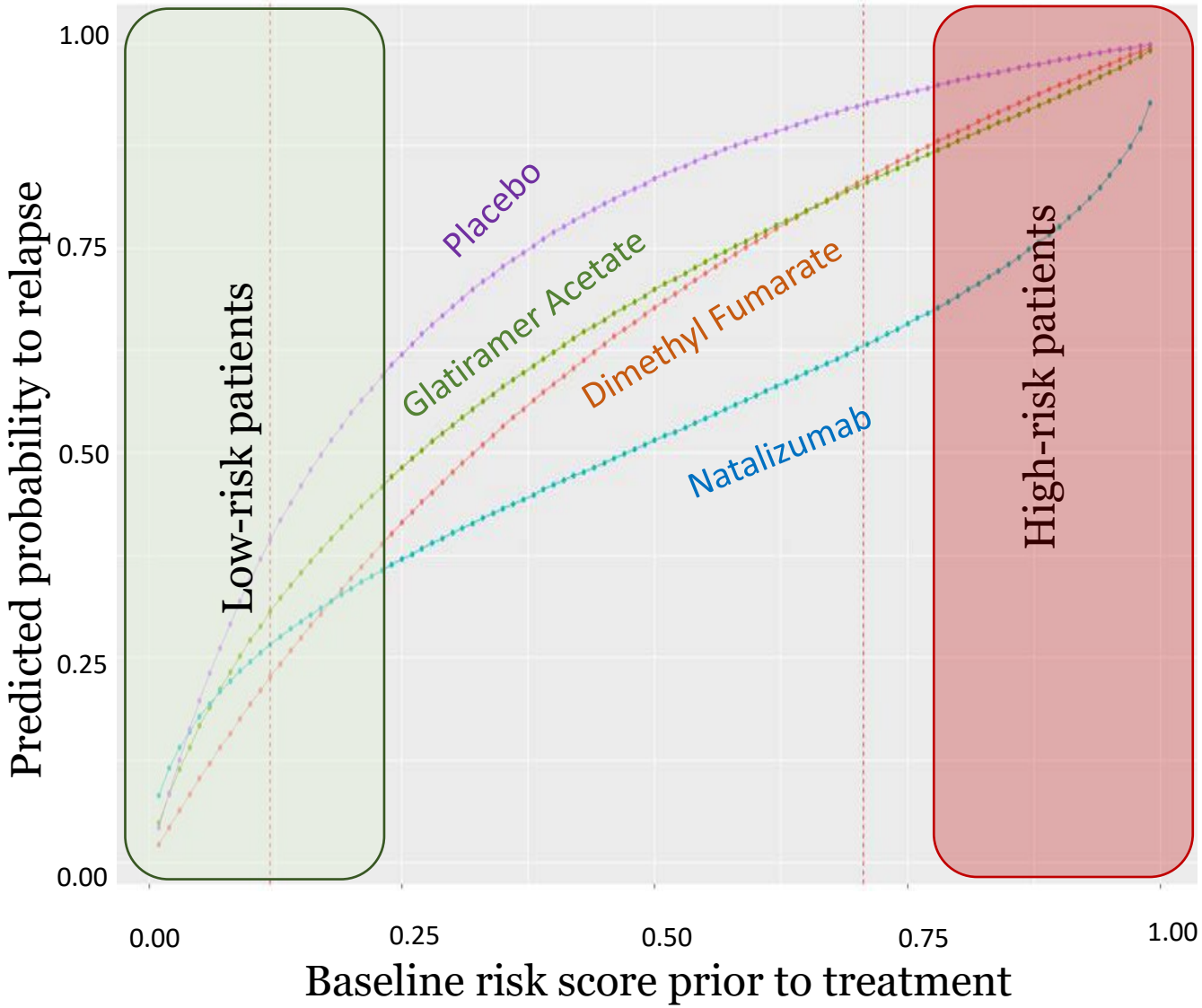
OR for relapse for one unit increase in logit-risk in untreated patients (placebo) - **(exp(B))** = 2.8 (2.1, 3.9)

	OR for relapse versus placebo at the study mean risk **(exp(D))** & 95% Cr. Intervals	OR versus placebo for one unit of increase in the logit risk **(exp(G))** & 95% Cr. Intervals
Natalizumab	0.28 (0.21, 0.37)	0.62 (0.31, 1.15)
Glatiramer Acetate	0.52 (0.34, 0.78)	0.83 (0.32, 2.10)
Dimethyl Fumarate	0.43 (0.3, 0.57)	0.96 (0.50, 1.87)

$$logit(p_{ijk}) = \begin{cases} u_j + B \times (logitR_{ij} - \overline{logitR_j}) if \ \ k = b_j \\ u_j + D_{b_jk} + B \times (logitR_{ij} - \overline{logitR_j}) + G_{b_jk} \times (logitR_{ij} - \overline{logitR_j}), \qquad if \ k \neq b_j \end{cases}$$

Stage 3: IPD Network Meta-regression

Results: Estimation of model parameters



Treatment	Mean	Less than 25% Risk	More than 75%
Natalizumab	51%	26%	76%
Glatiramer Acetate	65%	30%	92%
Dimethyl Fumarate	62%	23%	93%

Best treatment **Dimethyl fumarate - 3% Absolute benefit** compared to Natalizumab

Best treatment **Natalizumab- 17% Absolute benefit** compared to Dimethyl Fumarate

Conclusions & further research

Conclusions

The risk score blinded to treatment modifies the absolute benefit of treatments

Further research

We plan to use measures relevant to clinical usefulness to validate the model

Thank you for your attention!

Questions?