

Modelling heterogeneous treatment effects using baseline risk score

Georgia Salanti, University of Bern, Switzerland

Work done by Dr. Konstantina Chalkou, Dr. Tasnim Hamza, Frederik Philipona

with M. Pletzer, N. Meyer, J. Lorscheider, E. Steyerberg, M. Egger, A. Manca, J. Kuhle, and F. Pellegrini

Relapsing–remitting multiple sclerosis

A chronic inflammatory disease of the central nervous system

Characterized by episodes of neurological worsening (relapses) followed by periods of partial or complete recovery.

OUTCOME

Relapse

Episodes of neurological worsening

EDSS (EXPANDED DISABILITY STATUS SCALE)

Used to measure the level of neurological disability.



TREATMENTS

Several disease modifying drugs



**GLATIRAMER
ACETATE**

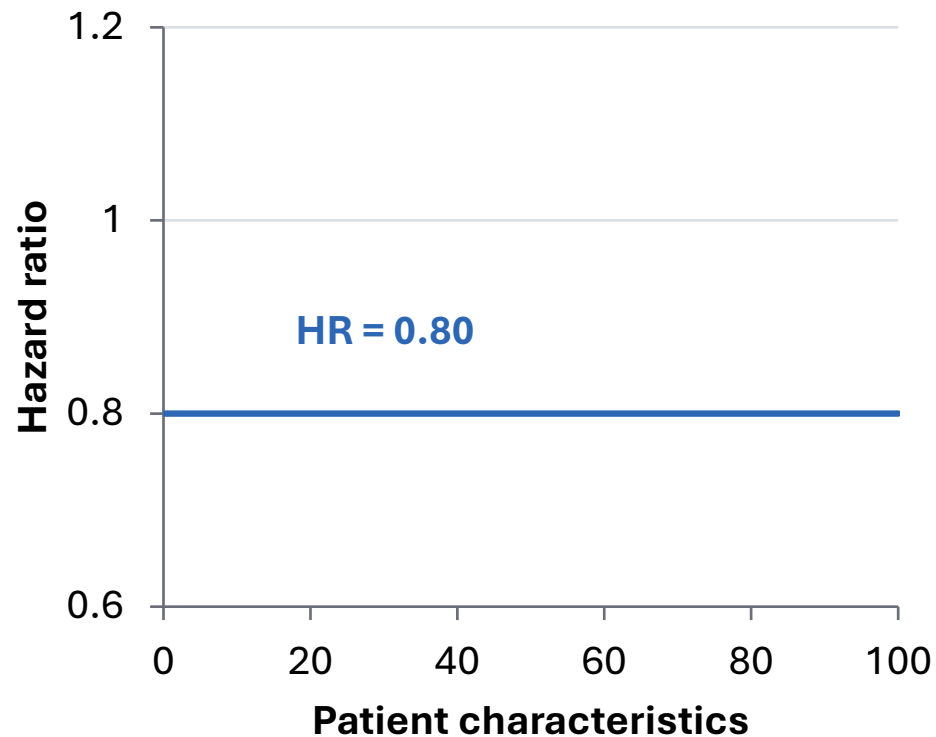


**DIMETHYL
FUMARATE**

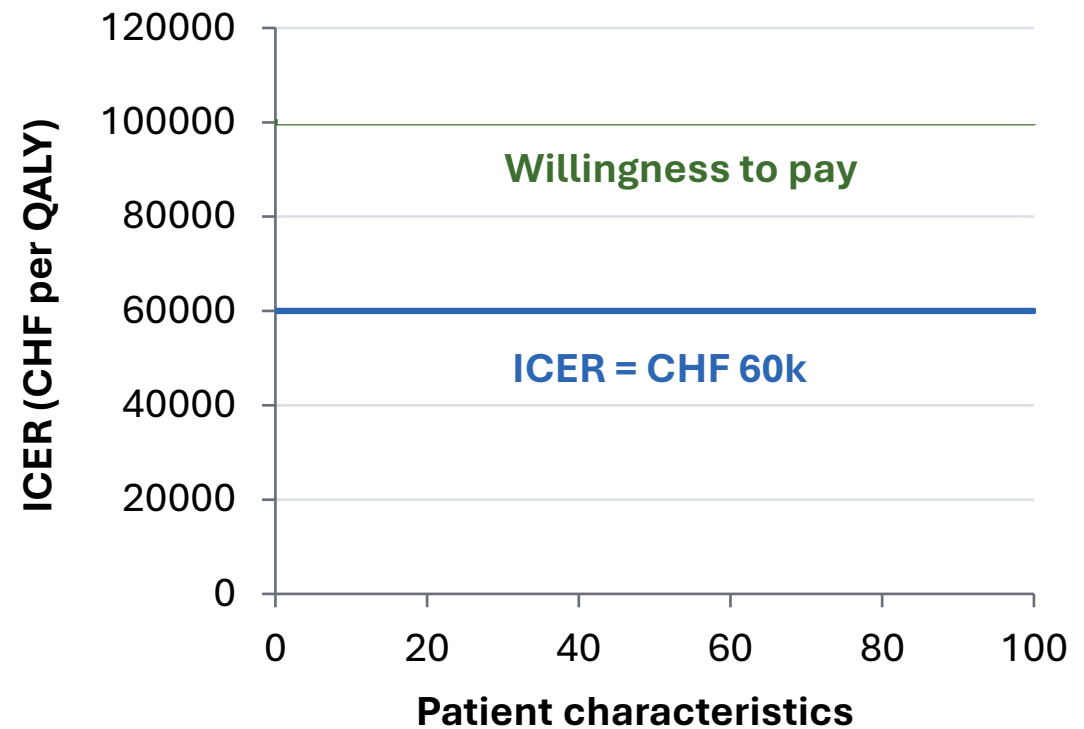
Which treatment is best on average?

dimethyl fumarate vs glatiramer acetate

Effectiveness



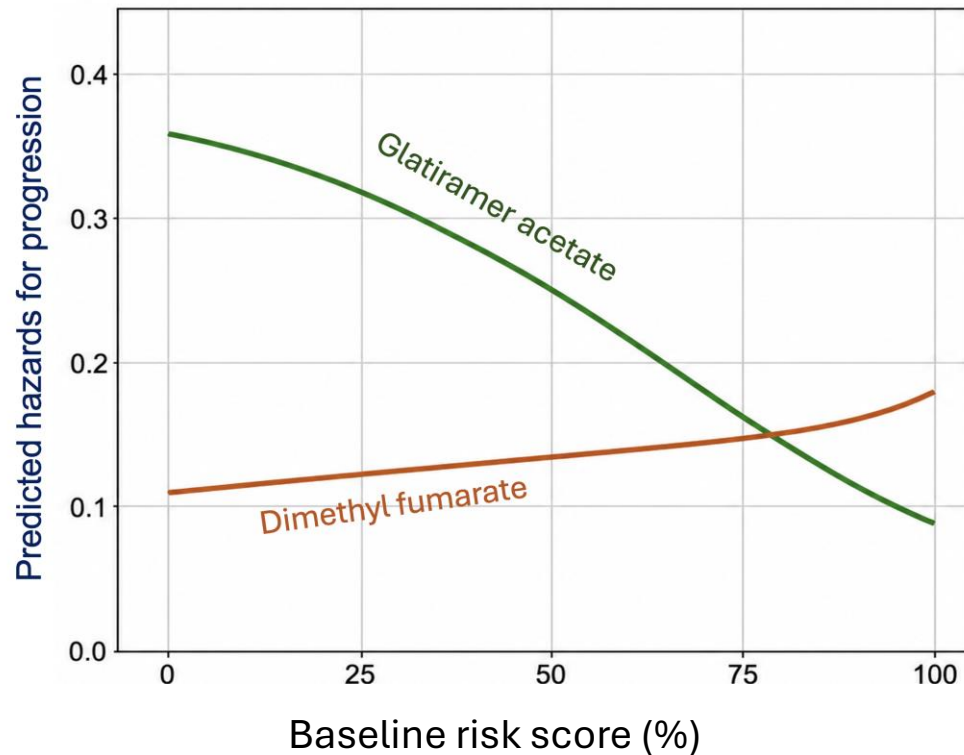
Cost-effectiveness



Incremental Cost-Effectiveness Ratio
 $ICER = \Delta COST / \Delta QALY$

Which treatment is best for whom?

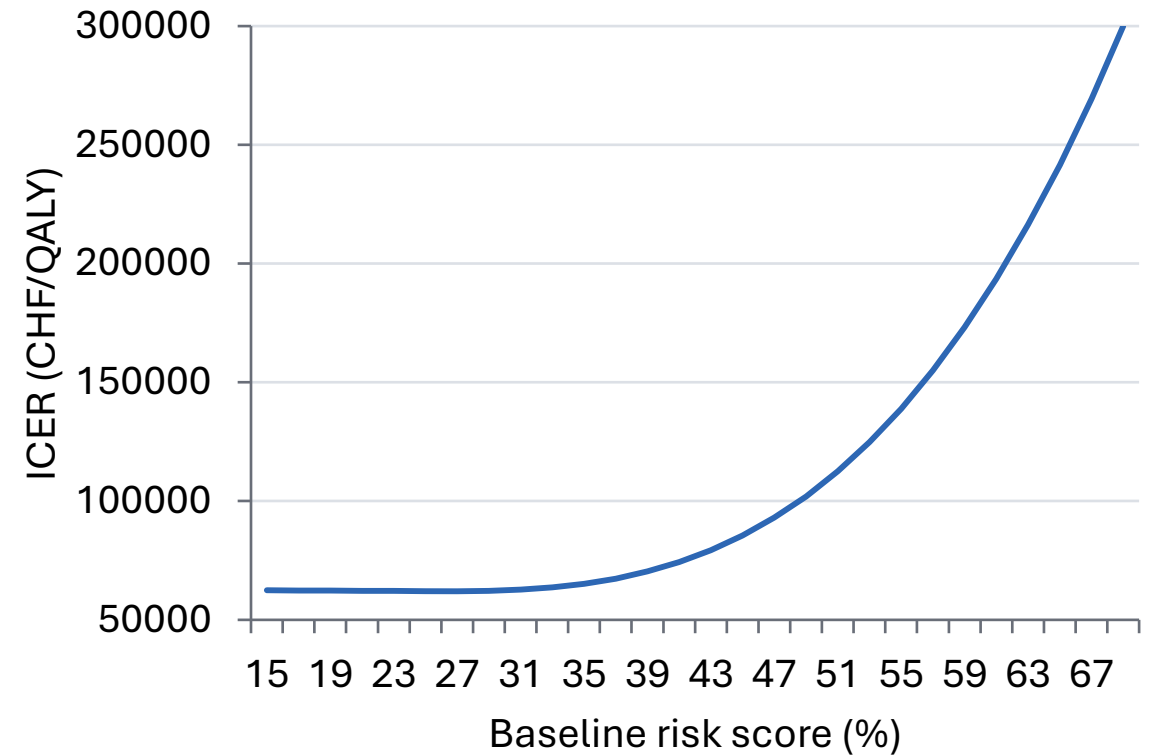
Effectiveness



Network meta-analysis

Treatment outcome as a function of risk

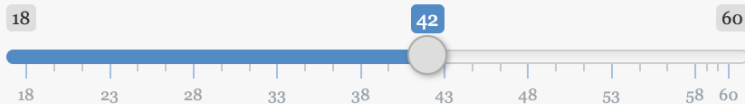
Cost-effectiveness



Cost-effectiveness model

ICER as a function of risk

Age (years)



☐ Male

☐ White

☐ Prior treatment

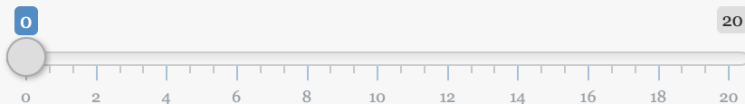
Baseline EDSS



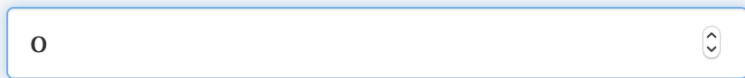
Years since onset of symptoms



Number of relapses the last 1 year



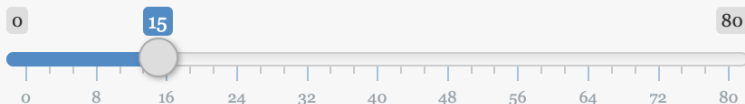
Months since last relapse



Baseline Timed 25-Foot Walk

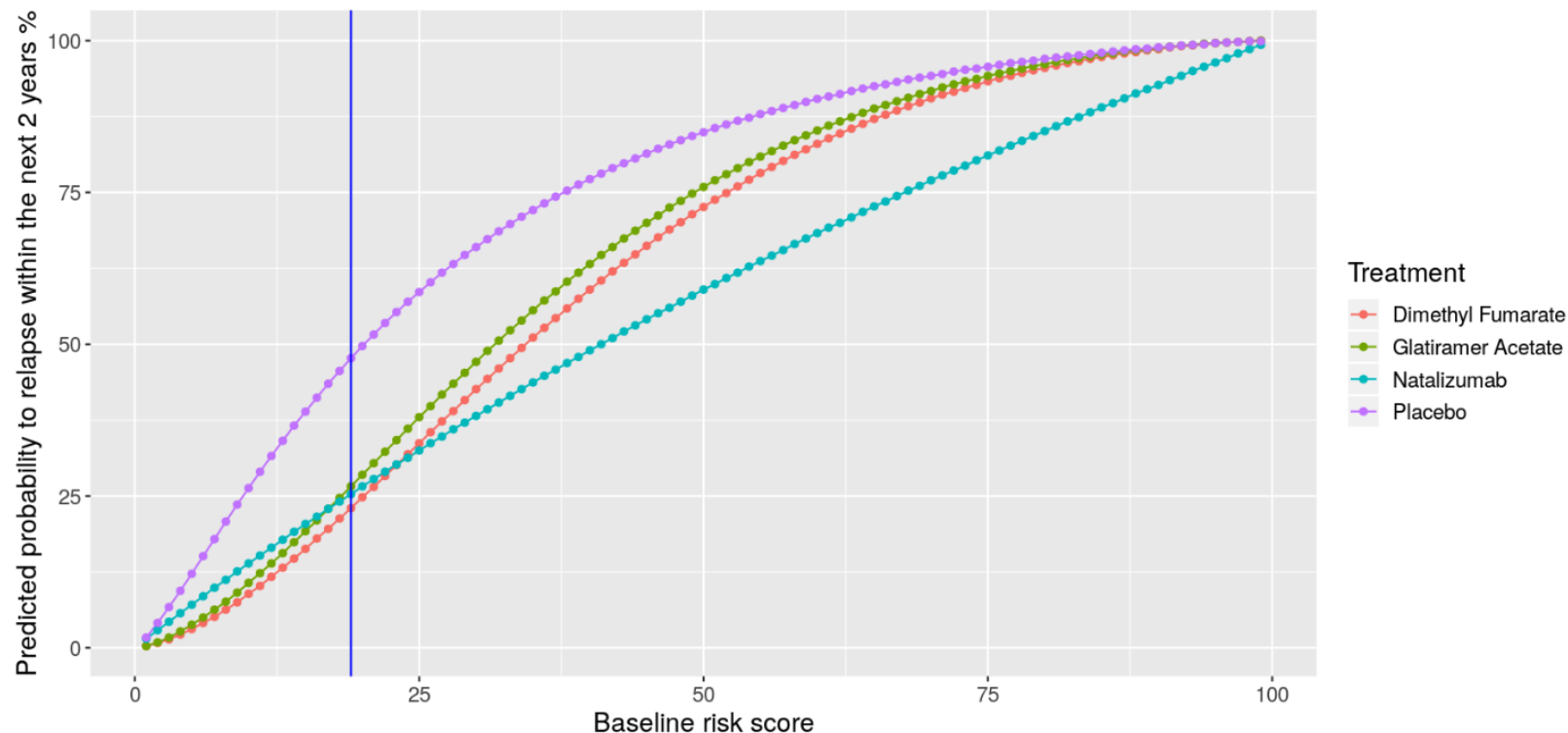


Baseline PASAT 3



Personalised effects for many treatments

Plot of predicted probabilities



A two-stage prediction model for heterogeneous effects of treatments. Stat Med. 2021

Predicted probabilities

Dimethyl Fumarate - 23 % / Glatiramer Acetate - 27 % / Natalizumab - 25 % / Placebo - 48 %

<https://cinema.ispm.unibe.ch/shinies/koms/>



Outline of the talk



Risk modelling approach to predict treatment effects.

What it is, assumptions, and advantages.



Risk modelling approach to predict heterogeneous treatment effects using IPD NMA.



Evaluate the clinical usefulness of the resulting model

Extending the decision curve analysis.



Produce risk-dependent cost-effectiveness analysis.

Chalkou K, Decision Curve Analysis for Personalized Treatment Choice between Multiple Options. Medical Decision Making. 2023;43(3):337-349.

Parsimony versus flexibility

PATH statement by Kent et al AIM 2020

Risk Modelling (RM)



Parsimony. Reduces dimensionality by summarizing patient characteristics into the baseline risk and then using it as the sole covariate

Effect Modification (EM)



Flexibility. Map all complex interactions directly to capture pure effect modifiers.

Types of patient characteristics x

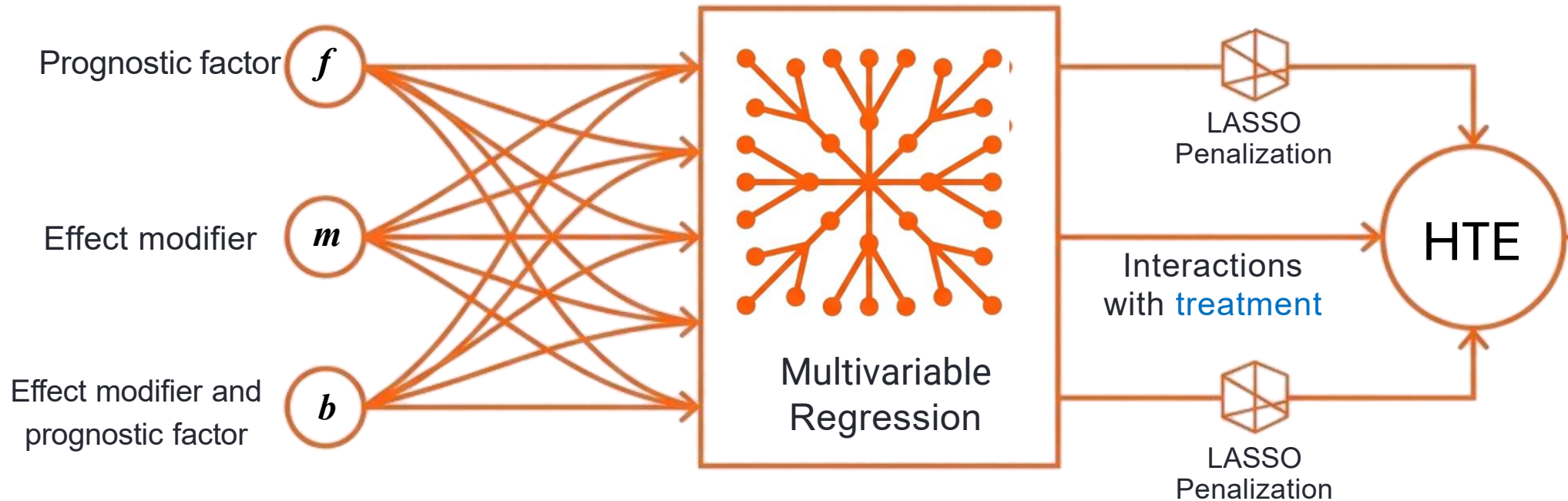
Prognostic **f**actors (f)

Effect **m**odifiers (m)

Characteristics that are **b**oth prognostic factors and effect modifiers (b)

Effect modification approach

$$\text{logit}(E[Y_i | \mathbf{x}_i, t_i]) = \alpha + \beta t_i + \gamma_b b_i + \gamma_f f_i + \delta_b b_i t_i + \delta_m m_i t_i.$$



EM is a one-stage multivariable regression explicitly mapping treatment interactions.

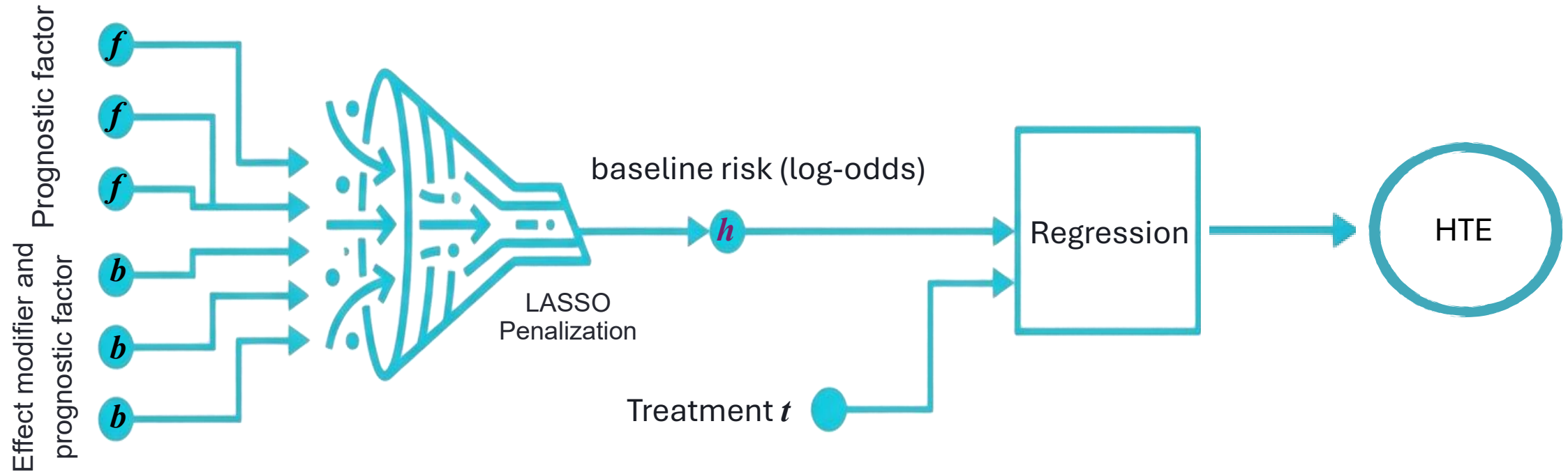
Maximum flexibility.
EM clearly differentiates between “pure” effect modifiers and prognostic factors.

Relies heavily on LASSO penalization to shrink noise and prevent overfitting.

Risk modelling approach

Stage 1 $\text{logit}(E[Y_i | \mathbf{x}_i, t_i = 0]) = \alpha_1 + \gamma_\beta \mathbf{b}_i + \gamma_f \mathbf{f}_i = \mathbf{h}_i$

Stage 2 $\text{logit}(E[Y_i | \mathbf{x}_i, t_i]) = \alpha_2 + \beta t_i + \gamma \mathbf{h}_i + \delta \mathbf{h}_i t_i$



RM is a two-stage process forcing all prognostic factors through a single baseline log-odds h

Core Assumption: A characteristic's significance as an effect modifier is strictly tied to its strength as a prognostic factor

Pure effect modifiers are inherently ignored

Theoretical assumptions and structural trade-offs

	Risk Modelling (RM)	Effect Modelling (EM)
Dimensionality	Low	High: Severe overfitting in small trials
Assumes baseline risk modifies effect?	Yes, explicitly via Stage 2	Yes, implicitly via interactions
Captures pure effect modifiers (m)?	No, structurally blind: only characteristics with prognostic value are considered in the first stage	Yes, directly modelled
Flexibility in modelling interactions	All baseline log-odds (summands of h_i) equally modify the treatment effect (via δ in stage 2): b variables all have similar impact. Pure prognostic factors are put as effect modifiers	Maximum flexibility, each characteristic is modelled directly

Comparison in simulation studies

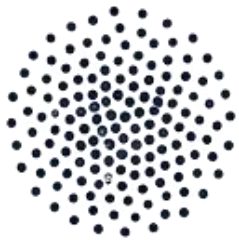


Sample Size of RCT

Main data-generating mechanisms



N=500

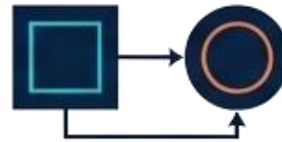


N=2000



N=5000

BOTH



prognostic + modifiers

PF



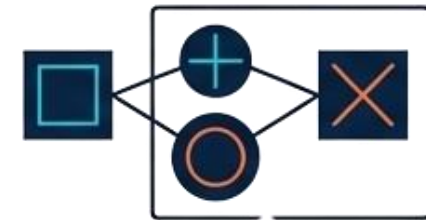
Pure prognostic factors only

BOTH+PF



Mixed prognostic and
dual-role variables

BOTH+PF+M



Includes pure effect modifiers
(The ultimate stress test)

Bias

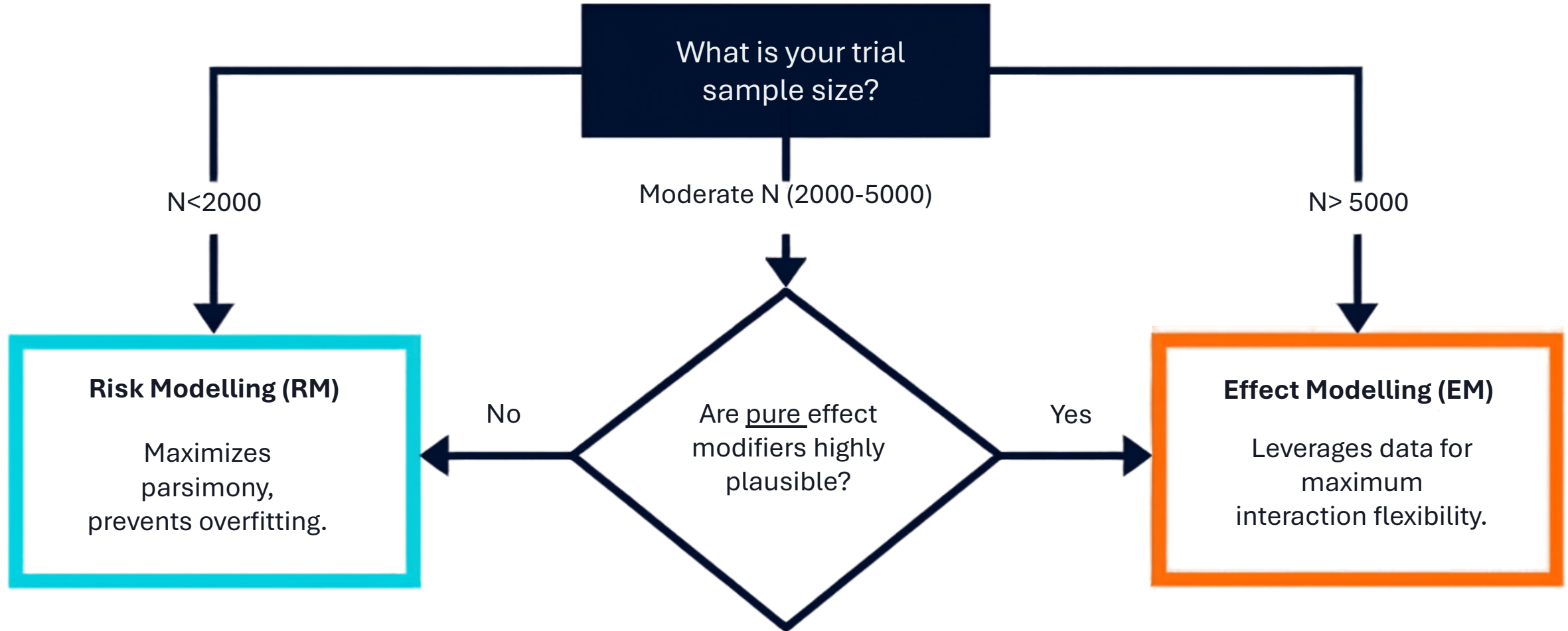


Sqrt(mean squared error)

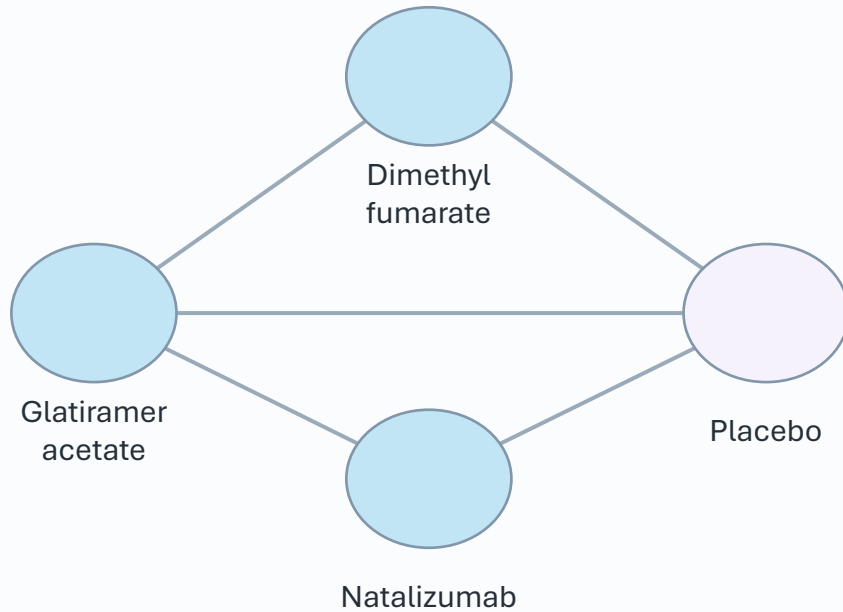




Which model should you use?



Risk modelling in network meta-regression



Evidence base

3 randomized RCTs
with individual participant data

2,990 participants

Outcome: **Relapse** in the next two years

NMA odds ratios

	Dimethyl fumarate	Glatiramer acetate	Natalizumab	Placebo
Dimethyl fumarate	—	1.24 [0.84, 1.83]	—	0.71 [0.54, 0.94]
Glatiramer acetate	1.17 [0.84, 1.66]	—	—	0.63 [0.45, 0.89]
Natalizumab	2.3 [1.4, 3.7]	1.95 [1.16, 3.28]	—	0.31 [0.20, 0.46]
Placebo	0.71 [0.54, 0.93]	0.60 [0.44, 0.83]	0.31 [0.20, 0.47]	—

Odds ratios with 95% confidence intervals • Reference direction: column treatment vs row treatment

Development of prognostic risk scores (stage 1)

Internal stage 1: use the RCT baseline covariates blinded to allocation to predict the outcome. Define this as $\text{expit}(h)$

LASSO

Prognostic factors: 9/28 selected via LASSO **Shrinkage of coefficients.**

Pre-specified model

Prognostic factors: 14 prognostic factors identified by Pellegrini et al.

Shrinkage of coefficients: penalized maximum estimation likelihood (ridge regression)

Chalkou K et al. Combining randomized and non-randomized data to predict heterogeneous effects of competing treatments. Res Synth Methods. 2024 Jul;15(4):641-656.

External model: use the Swiss Cohort Study and develop a prognostic model. Estimate h for the RCT data

Chalkou K et al. Development, validation and clinical usefulness of a prognostic model for relapse in relapsing-remitting multiple sclerosis. Diagn Progn Res. 2021 Oct 27;5(1):17.

IPD network meta-regression for relapse (stage 2)

$$\text{logit}(p_{ijk}) = \begin{cases} u_j + \gamma \times h_{ij} & \text{if } k = b_j \\ u_j + \beta_{b_jk} + \gamma \times h_{ij} + \delta_{b_jk} \times h_{ij}, & \text{if } k \neq b_j \end{cases}$$

$$\begin{aligned} \beta_{b_jk} &= \beta_{b_j} - \beta_k \\ \beta_k &\sim N(B_k, \tau^2) \end{aligned}$$

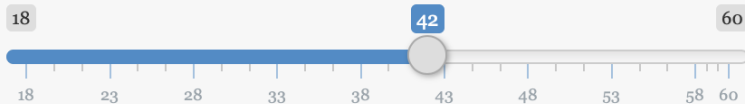
Results and details when h is from internal stage 1

	OR drug vs placebo at the study mean logit-risk ($\exp(B_k)$)	OR drug vs placebo for one unit increase in the logit-risk
Natalizumab	0.18	0.67
Glatiramer Acetate	0.41	0.87
Dimethyl Fumarate	0.43	1.06

Results and details when h is from external stage 1

Chalkou et al. *Combining randomized and non-randomized data to predict heterogeneous effects of competing treatments. Res Syn Meth.* 2024; 15(4): 641-656.
doi:[10.1002/jrsm.1717](https://doi.org/10.1002/jrsm.1717)

Age (years)



☐ Male

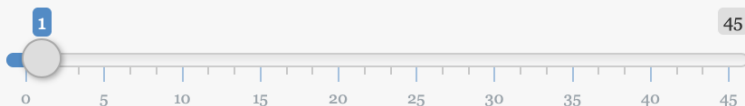
☐ White

☐ Prior treatment

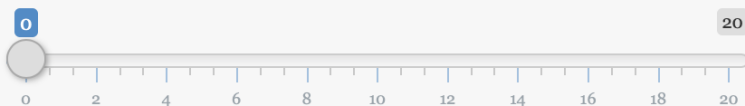
Baseline EDSS



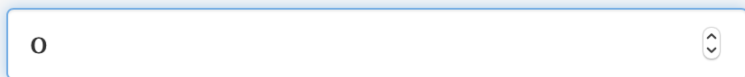
Years since onset of symptoms



Number of relapses the last 1 year



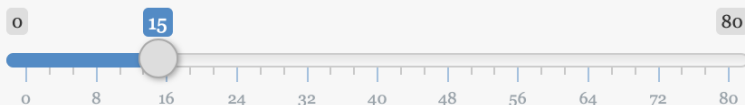
Months since last relapse



Baseline Timed 25-Foot Walk

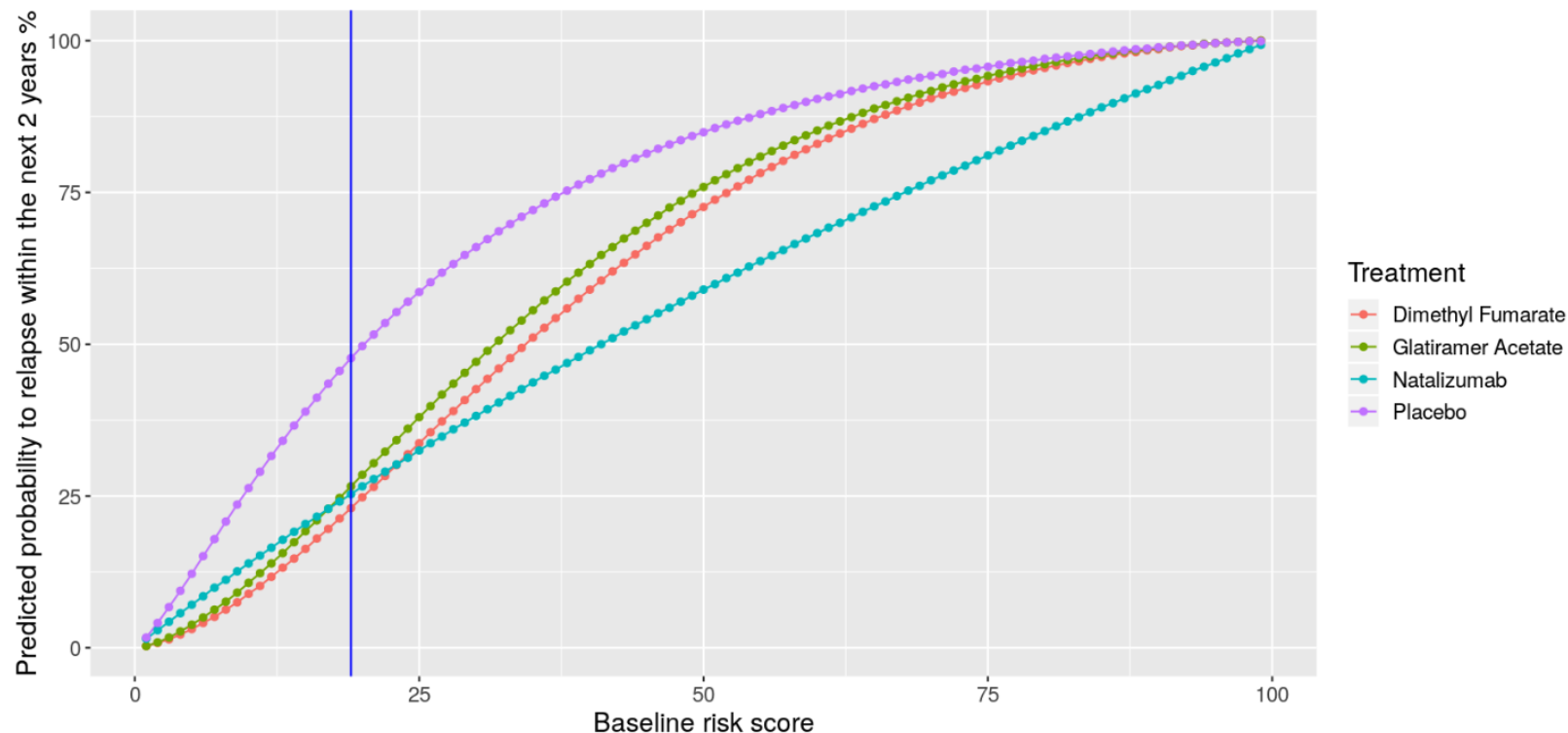


Baseline PASAT 3



Personalised effects for many treatments

Plot of predicted probabilities



A two-stage prediction model for heterogeneous effects of treatments. Stat Med. 2021

Predicted probabilities

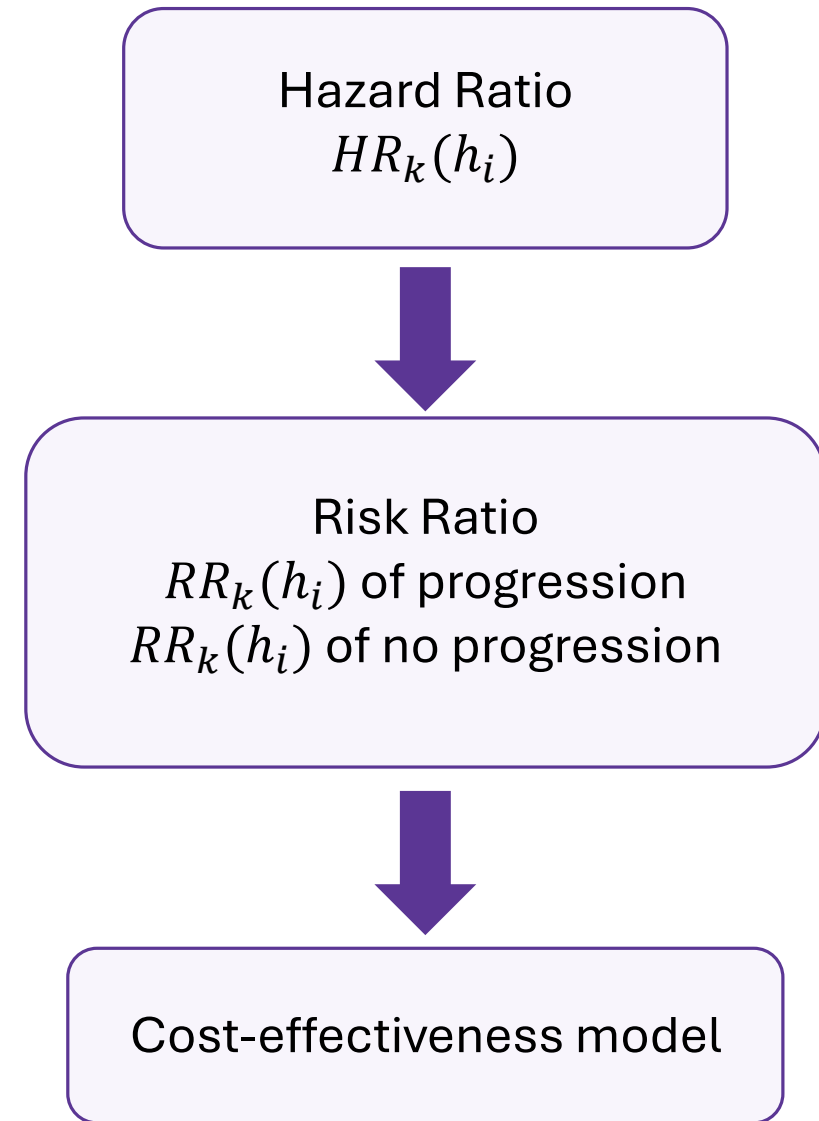
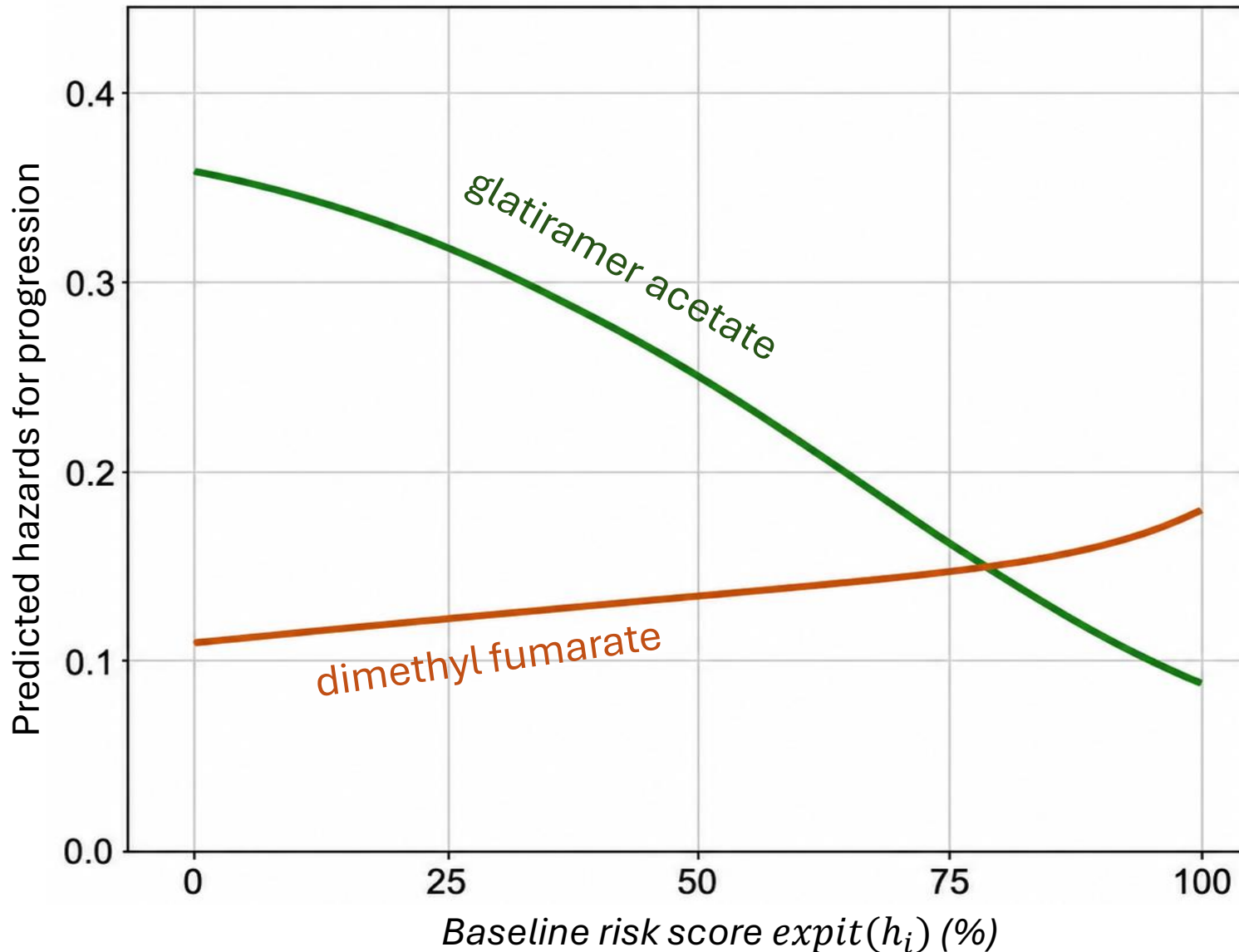
Dimethyl Fumarate - 23 % / Glatiramer Acetate - 27 % / Natalizumab - 25 % / Placebo - 48 %

<https://cinema.ispm.unibe.ch/shinies/koms/>



IPD network meta-regression for time to progression (stage 2)

Only dimethyl fumarate and glatiramer acetate



Treatment-specific transition probability matrix

10 EDSS states & death

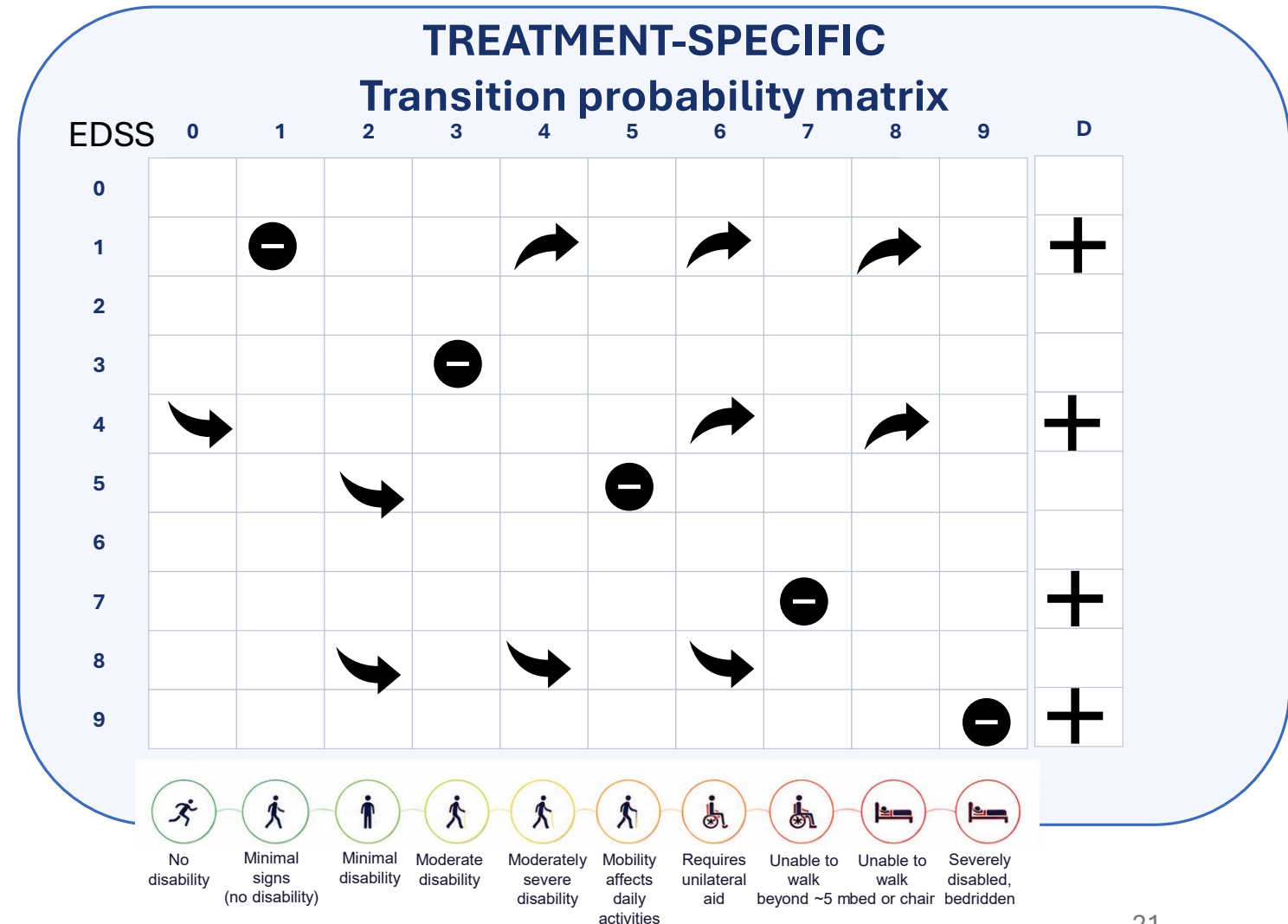
Rows=current state

Columns=next state

Patients can:

progress (top triangle),
remain stable (diagonal),
improve (lower triangle)
or die (last column)

These transition probabilities are
modified according to the treatment
received **and** the baseline risk



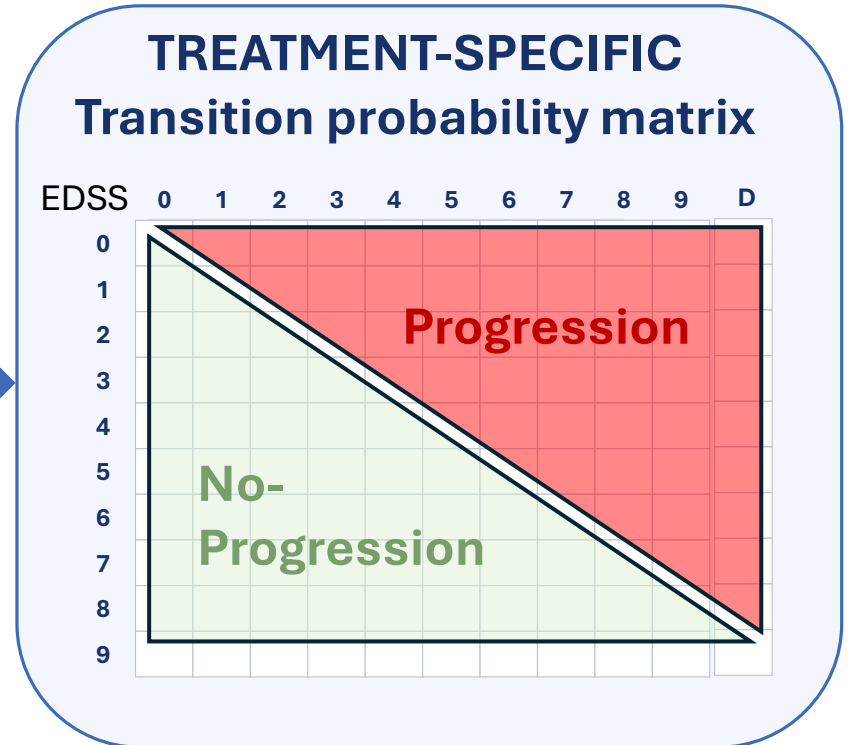
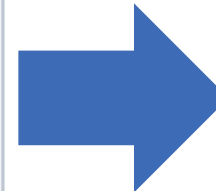
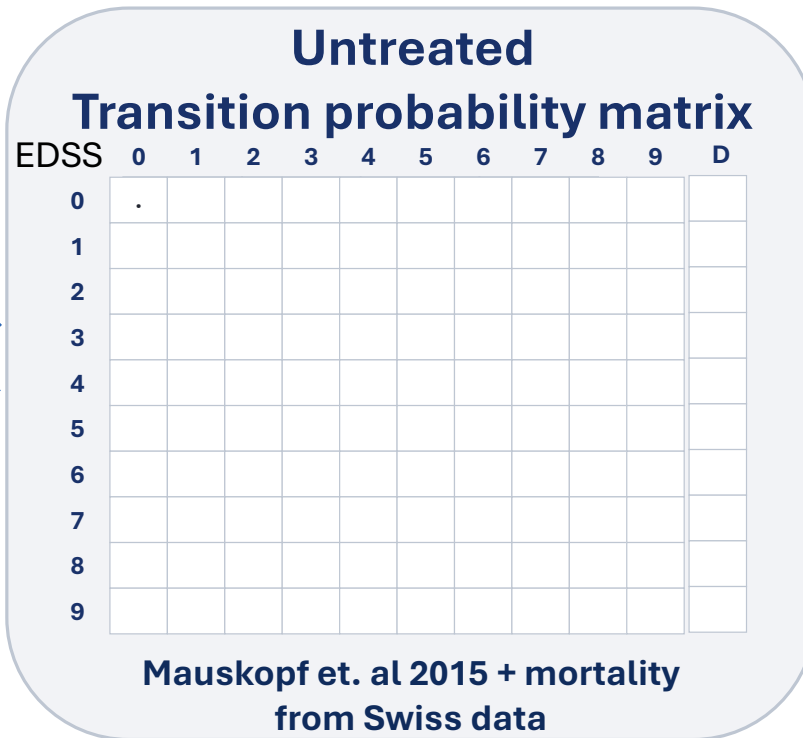
Treatment-specific transition probability matrix

We apply the risk ratios for progression and no-progression to the natural history model

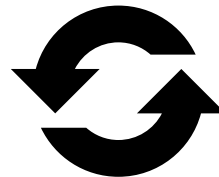
Risk Ratio

$RR_k(h_i)$ of progression

$RR_k(h_i)$ of no progression



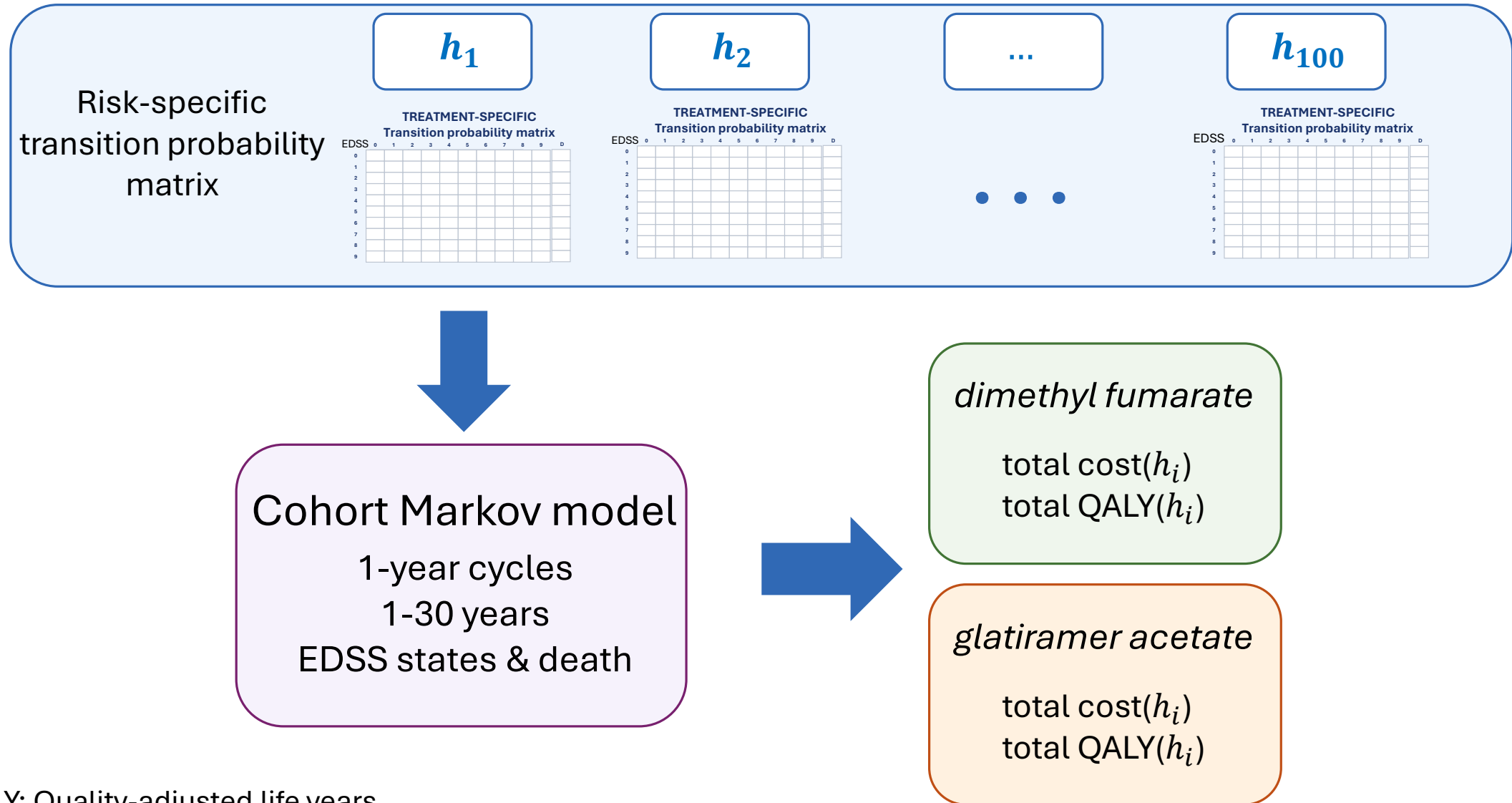
Repeat for every baseline risk score value



Repeat for every treatment in the network

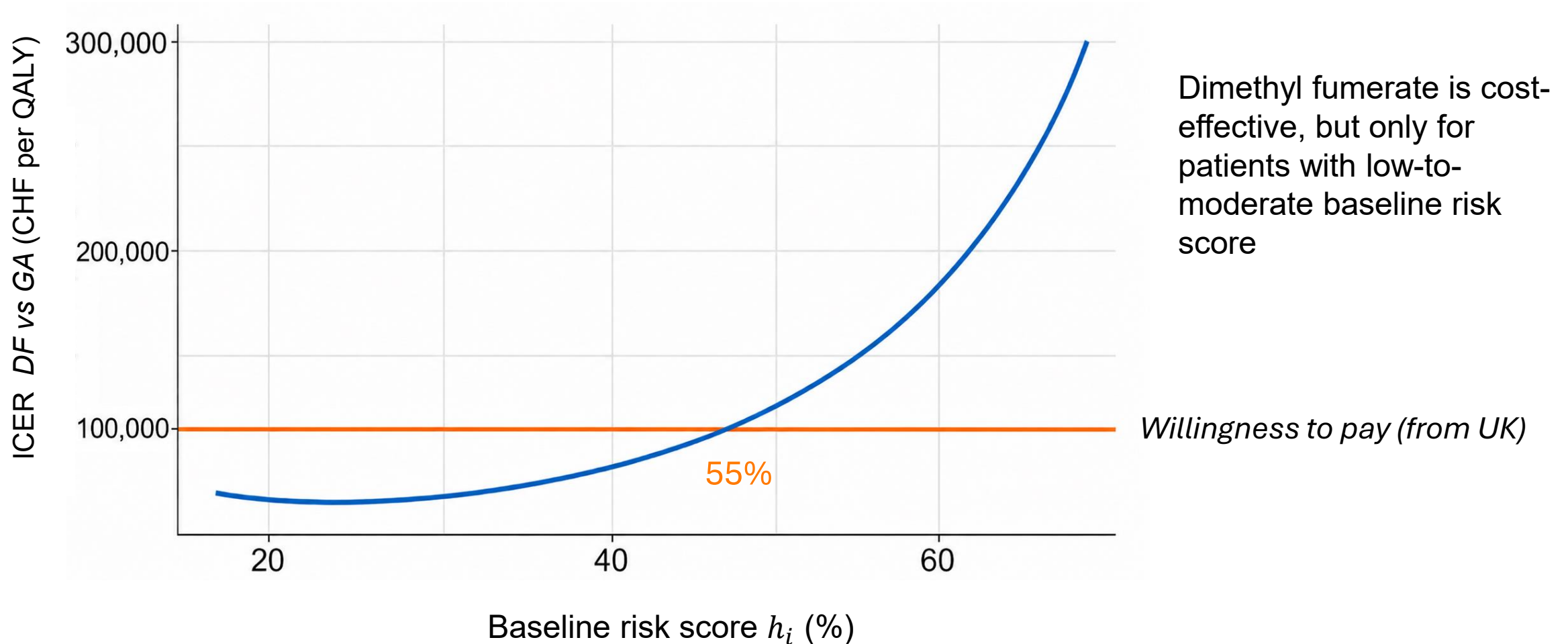
Calculate total costs and total QALYs per baseline risk value

Each risk-specific transition matrix generates a risk-specific disease trajectory



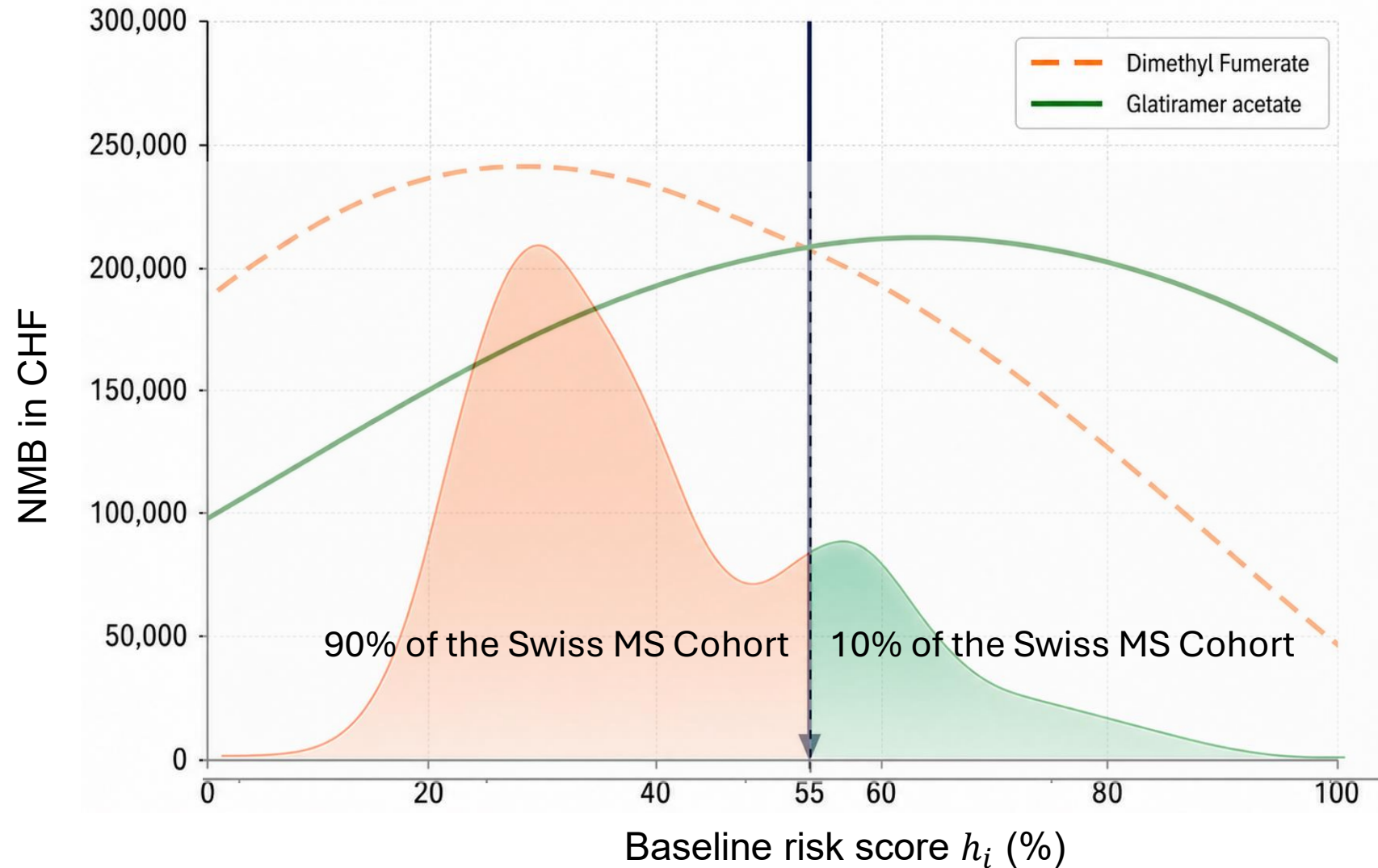
Risk-dependant incremental cost-effectiveness ratio (ICER)

$$ICER_i = \frac{\Delta Cost_i}{\Delta Effect_i}$$



Risk-dependant net monetary benefit (NMB)

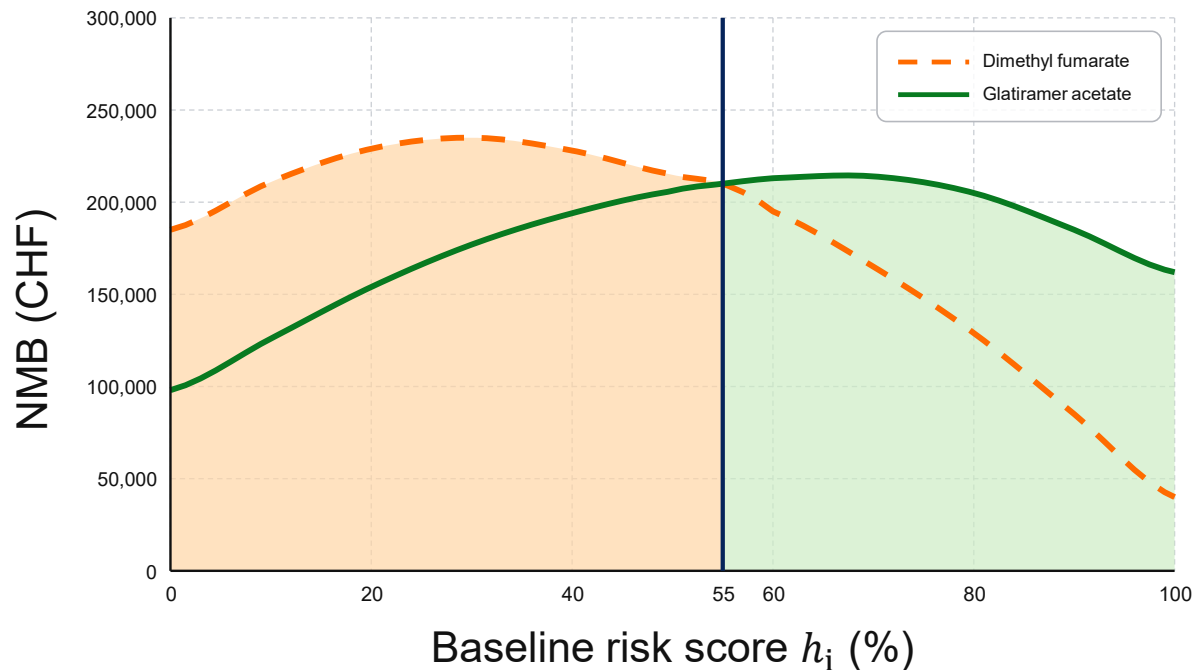
$$NMB_k(h_i) = (\lambda \times QALY_k(h_i)) - Cost_k(h_i)$$



What is the value of individualising the treatment?

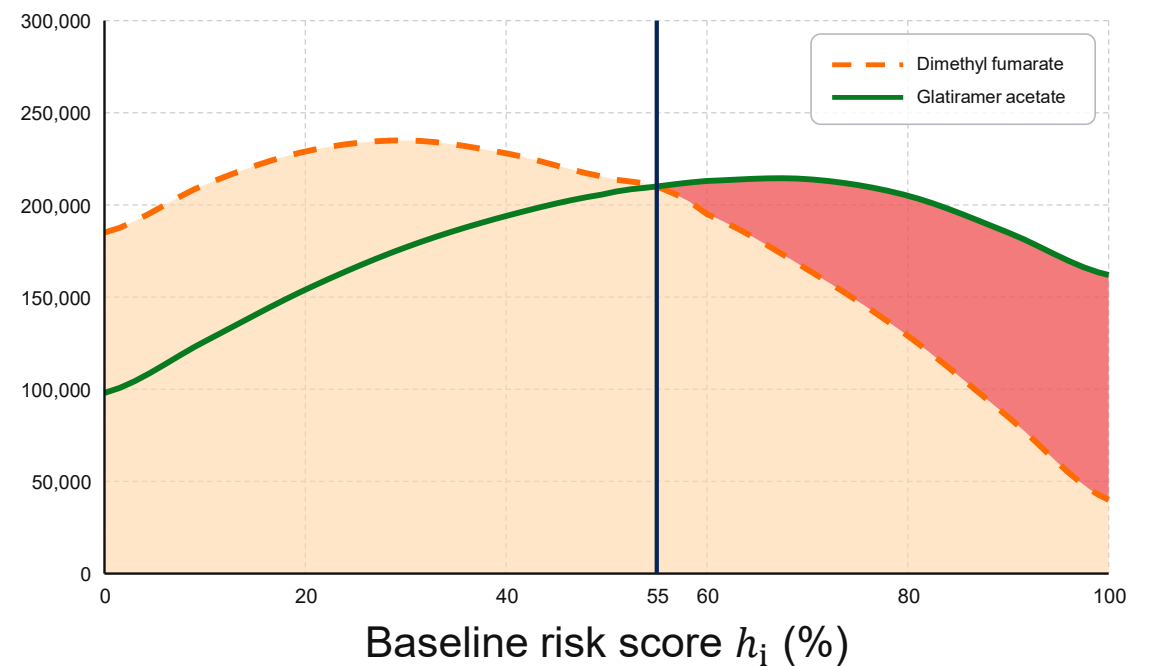
INDIVIDUALIZED STRATEGY

Dimethyl fumarate for low/-moderate risk
Glatiramer acetate for high risk



POPULATION-AVERAGE STRATEGY

Dimethyl fumarate for everyone



Expected value of individualized care:
Using individualized treatment strategy saves Switzerland **CHF 3.2M**

Use more risk modelling

We need to estimate heterogeneous treatment effects in many settings



Risk modelling approach reduces dimensionality by summarizing patient characteristics into the baseline risk and then using it as the only covariate

- Very handy when you have many different IPD in different servers!
- Reduces the means square error and it is better for studies that have small sample size (500)
- A nice way to summarize the population characteristics and see how treatment effects and cost-effectiveness changes in population subgroups

But be careful with risk modelling. Actually, with any modelling.

Data requirements



- IPD are required – access to IPD can be difficult.
- Relevant treatments may be unavailable in the evidence network; the example included only 3 active drugs.

Modelling assumptions

- Baseline risk must capture both prognosis and effect modification.
- Pure effect modifiers that do not contribute to prognosis may be missed.
- NMA assumptions still apply: transitivity, consistency and appropriate modelling of between-study variation.
- Study-level effect modifiers (e.g. risk of bias, year of randomization) may also influence treatment effects.
- Recalibration across populations and combining RCT and observational data introduce additional modelling assumptions.

References

- Philipona et al. Estimating heterogeneous treatment effects from randomised trials: a comparison of the risk modelling and treatment effect modelling approaches. arXiv:2609.21526 arXiv:2609.21526
- Hamza T et al. Next Generation Health Technology Assessment (HTx) Work Package 2 group. The Use of Baseline Score in Cost-Effectiveness Modeling of Competing Interventions. Value Health. 2026 Aug;29(8):1434-1442.
- Chalkou K et al. Combining randomized and non-randomized data to predict heterogeneous effects of competing treatments. Res Synth Methods. 2024 Jul;15(4):641-656.
- Chalkou K et al. Decision Curve Analysis for Personalized Treatment Choice between Multiple Options. Med Decis Making. 2023 Apr;43(3):337-349.
- Chalkou K et al. Development, validation and clinical usefulness of a prognostic model for relapse in relapsing-remitting multiple sclerosis. Diagn Progn Res. 2021 Oct 27;5(1):17.
- Chalkou K et al. A two- stage prediction model for heterogeneous effects of treatments. Stat Med. 2021 Sep 10;40(20):4362-4375.