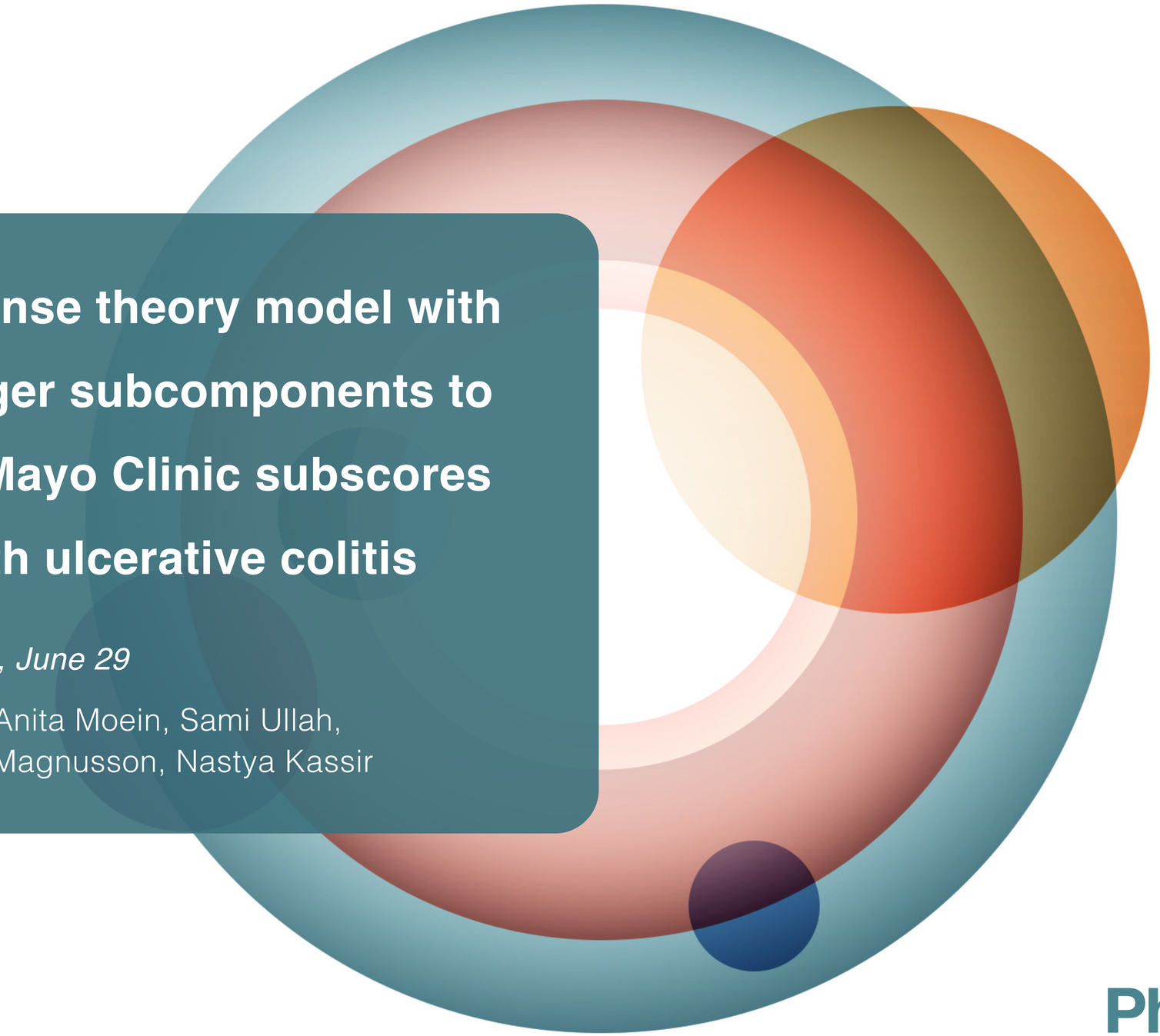
An item response theory model with bounded integer subcomponents to describe the Mayo Clinic subscores in patients with ulcerative colitis

PAGE 2022, Ljubljana, June 29

Jurgen Langenhorst, Anita Moein, Sami Ullah,
Matts Kågedal, Mats Magnusson, Nastya Kassir



Pharmetheus

Genentech
A Member of the Roche Group

Genentech involvement

Anita Moein

Matts Kågedal

Nastya Kassir

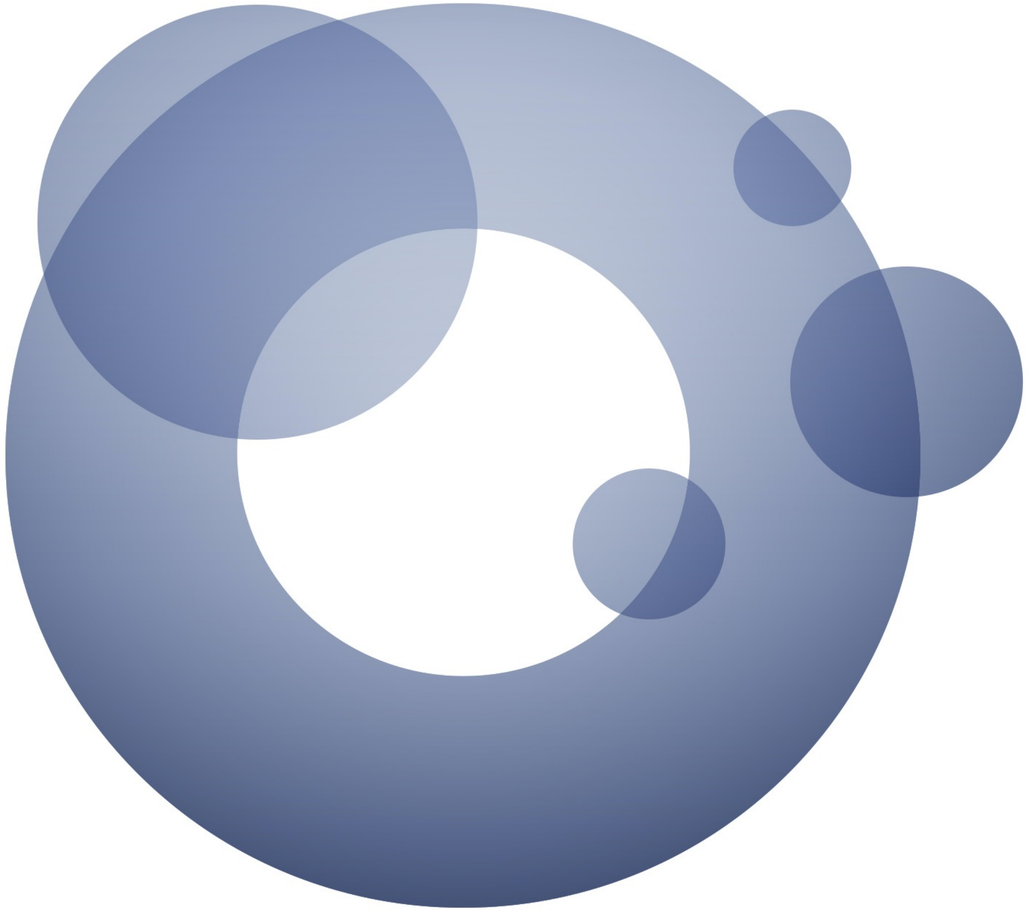
Author disclosures

- Nastya Kassir, Anita Moein, and Matts Kâgedal are employees of Genentech and own Roche stocks

- Jurgen Langenhorst, Sami Ullah, and Mats Magnusson (MM) are employees of Pharmetheus and paid consultants for Genentech, Inc and MM owns stocks in Pharmetheus

Key messages

- Remission in Ulcerative Colitis (UC) is a key clinical parameter, but it binarizes a large quantity of longitudinal integer score data
- An item response theory (IRT) model with bounded integer item models is proposed as a more powerful alternative to the binarized approach
- The IRT model described the analysis data and an external source of data adequately
- The IRT model can be used to improve model informed drug development (MIDD) in UC



MIDD case study

UC is a severe disease of the gut, with a need of improved therapies

- UC is an inflammatory bowel disease affecting the colon and rectum
- Patients suffer from a range of symptoms from persistent diarrhea to rectal bleeding
- Long-term uncontrolled UC associates with severe consequences including increased rates of depression and risk of colon cancer
- Anti tumor necrosis factor (TNF) therapy has lead to stable induction of remission, but
  - A substantial number of patients have refractory disease
  - Anti-TNF therapy associates with significant side-effects

Remission in UC is a key clinical parameter, but effectively binarizes a large quantity of longitudinal integer score data

Rectal bleeding (RB)

0: No blood seen
1: Streaks of blood
2: Obvious blood
3: Blood alone passed

Remission in UC is a key clinical parameter, but effectively binarizes a large quantity of longitudinal integer score data

Rectal bleeding (RB)

0: No blood seen
1: Streaks of blood
2: Obvious blood
3: Blood alone passed

Stool frequency (SF)

0: Normal number of stools
1: 1-2 stools more than normal
2: 3-4 more stools than normal
3: ≥5 more stools than normal

Remission in UC is a key clinical parameter, but effectively binarizes a large quantity of longitudinal integer score data

Rectal bleeding (RB)

0: No blood seen
1: Streaks of blood
2: Obvious blood
3: Blood alone passed

Stool frequency (SF)

0: Normal number of stools
1: 1-2 stools more than normal
2: 3-4 more stools than normal
3: ≥5 more stools than normal

Physician's global assesment (PGA)

0: Normal
1: Mild disease
2: Moderate disease
3: Severe disease

Remission in UC is a key clinical parameter, but effectively binarizes a large quantity of longitudinal integer score data

Rectal bleeding (RB)

0: No blood seen
1: Streaks of blood
2: Obvious blood
3: Blood alone passed

Endoscopy (END)

0: Normal or inactive disease
1: Mild disease
2: Moderate disease
3: Severe disease

Stool frequency (SF)

0: Normal number of stools
1: 1-2 stools more than normal
2: 3-4 more stools than normal
3: ≥5 more stools than normal

Physician's global assesment (PGA)

0: Normal
1: Mild disease
2: Moderate disease
3: Severe disease

Remission in UC is a key clinical parameter, but effectively binarizes a large quantity of longitudinal integer score data

Rectal bleeding (RB)

0: No blood seen
1: Streaks of blood
2: Obvious blood
3: Blood alone passed

Endoscopy (END)

0: Normal or inactive disease
1: Mild disease
2: Moderate disease
3: Severe disease

Mayo clinic score (MCS, 0-12):
sum of all subscores

Stool frequency (SF)

0: Normal number of stools
1: 1-2 stools more than normal
2: 3-4 more stools than normal
3: ≥5 more stools than normal

Physician's global assesment (PGA)

0: Normal
1: Mild disease
2: Moderate disease
3: Severe disease

Remission in UC is a key clinical parameter, but effectively binarizes a large quantity of longitudinal integer score data

Rectal bleeding (RB)

0: No blood seen
1: Streaks of blood
2: Obvious blood
3: Blood alone passed

Endoscopy (END)

0: Normal or inactive disease
1: Mild disease
2: Moderate disease
3: Severe disease

Mayo clinic score (MCS, 0-12):
sum of all subscores

Remission (yes/no, 1/0):
MCS < 3
All subscores < 2
RB of 0

Stool frequency (SF)

0: Normal number of stools
1: 1-2 stools more than normal
2: 3-4 more stools than normal
3: ≥5 more stools than normal

Physician's global assesment (PGA)

0: Normal
1: Mild disease
2: Moderate disease
3: Severe disease

Proposed MIDD solution

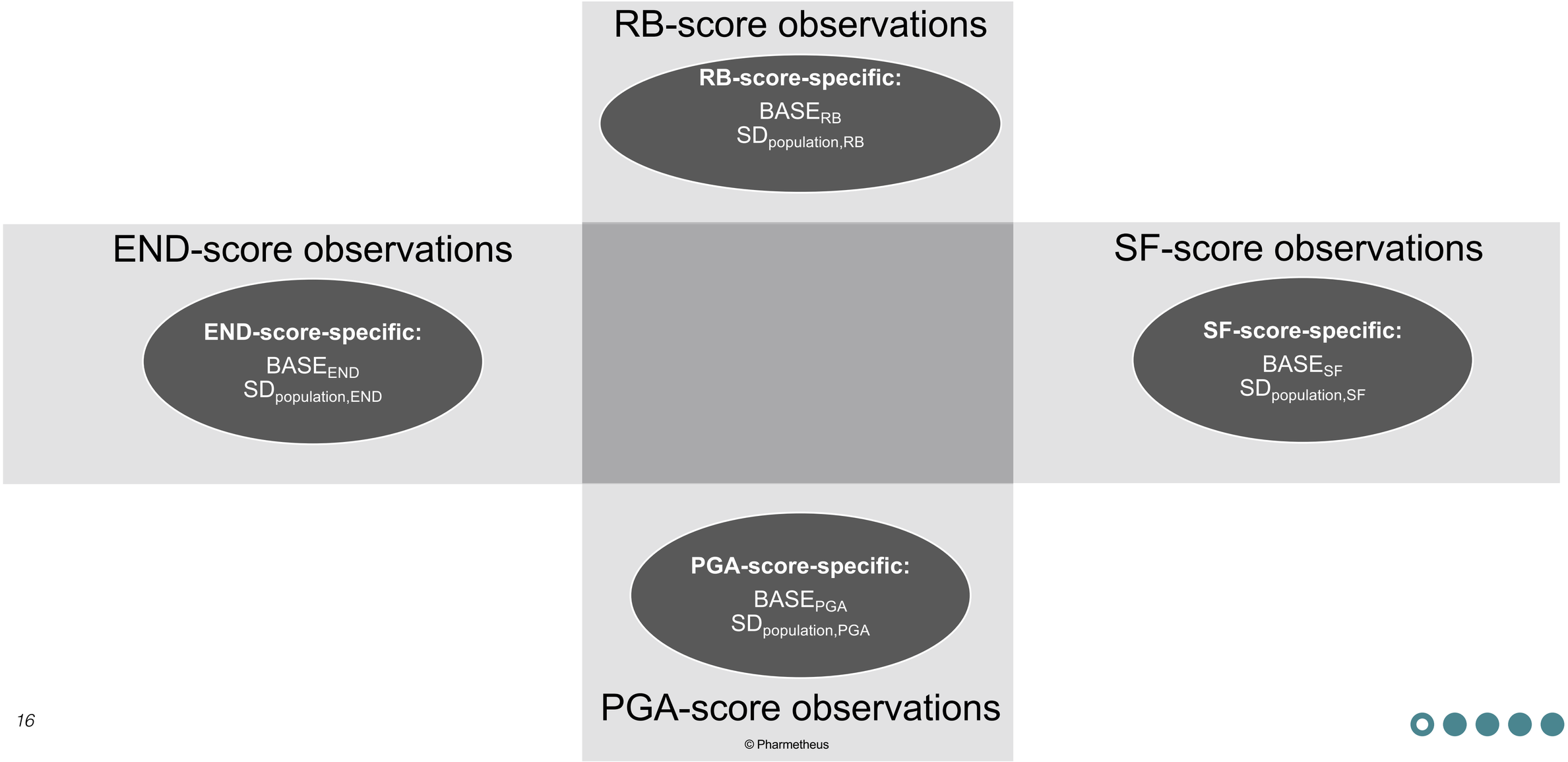
An IRT model with bounded integer item models is proposed to model four UC efficacy subscores simultaneously

A typical IRT model for such data	
Proportional odds models with discrimination parameter describing the item (i.e. subscore) characteristics	
Each item model needs $N_{categories}$ parameters (4)	
A latent variable that links the item models	
Increasing latent variable ➡️ increasing probability of higher scores	
Each subscore/item and derived parameter can be predicted at any timepoint	

An IRT model with bounded integer item models is proposed to model four UC efficacy subscores simultaneously

A typical IRT model for such data	The proposed model
Proportional odds models with discrimination parameter describing the item (i.e. subscore) characteristics	Bounded integer models describing the subscore/item characteristics
Each item model needs $N_{categories}$ parameters (4)	Each item model needs 2 parameters, regardless of $N_{categories}$
A latent variable that links the item models	
Increasing latent variable ➡️ increasing probability of higher scores	
Each subscore/item and derived parameter can be predicted at any timepoint	

The IRT model is partitioned in item-specific parameters and a latent disease model



The IRT model is partitioned in item-specific parameters and a latent disease model

RB-score observations

RB-score-specific:
$BASE_{RB}$
$SD_{population,RB}$

END-score observations

END-score-specific:
$BASE_{END}$
$SD_{population,END}$

SF-score observations

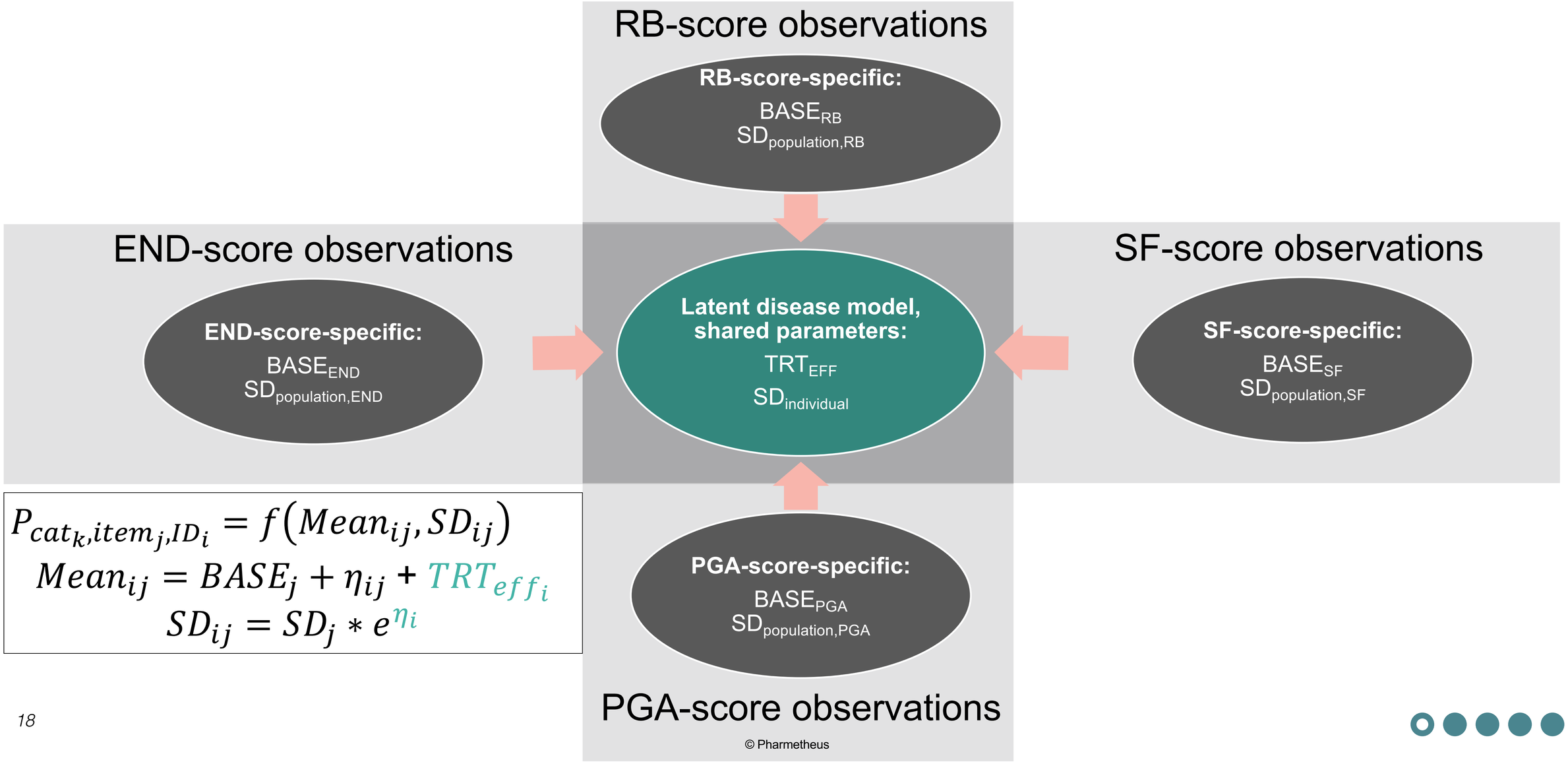
SF-score-specific:
$BASE_{SF}$
$SD_{population,SF}$

$$P_{cat_k,item_j,ID_i} = f(Mean_{ij}, SD_{ij})$$
$$Mean_{ij} = BASE_j + \eta_{ij} + TRT_{eff_i}$$
$$SD_{ij} = SD_j * e^{\eta_i}$$

PGA-score-specific:
$BASE_{PGA}$
$SD_{population,PGA}$

PGA-score observations

The IRT model is partitioned in item-specific parameters and a latent disease model



$$P_{cat_k,item_j,ID_i} = f\left(Mean_{ij}, SD_{ij}\right)$$

$$Mean_{ij} = BASE_j + \eta_{ij} + TRT_{eff_i}$$

$$SD_{ij} = SD_j * e^{\eta_i}$$

Data available

Analysis data came from three phase III trials performed by Genentech/Roche to evaluate the efficacy of etrolizumab

UC TNF-NAIVE

HIBISCUS I & II

NCT02163759 & NCT02171429
(GA28948 & GA28949)
2 induction trials:
Etrolizumab vs adalimumab vs placebo

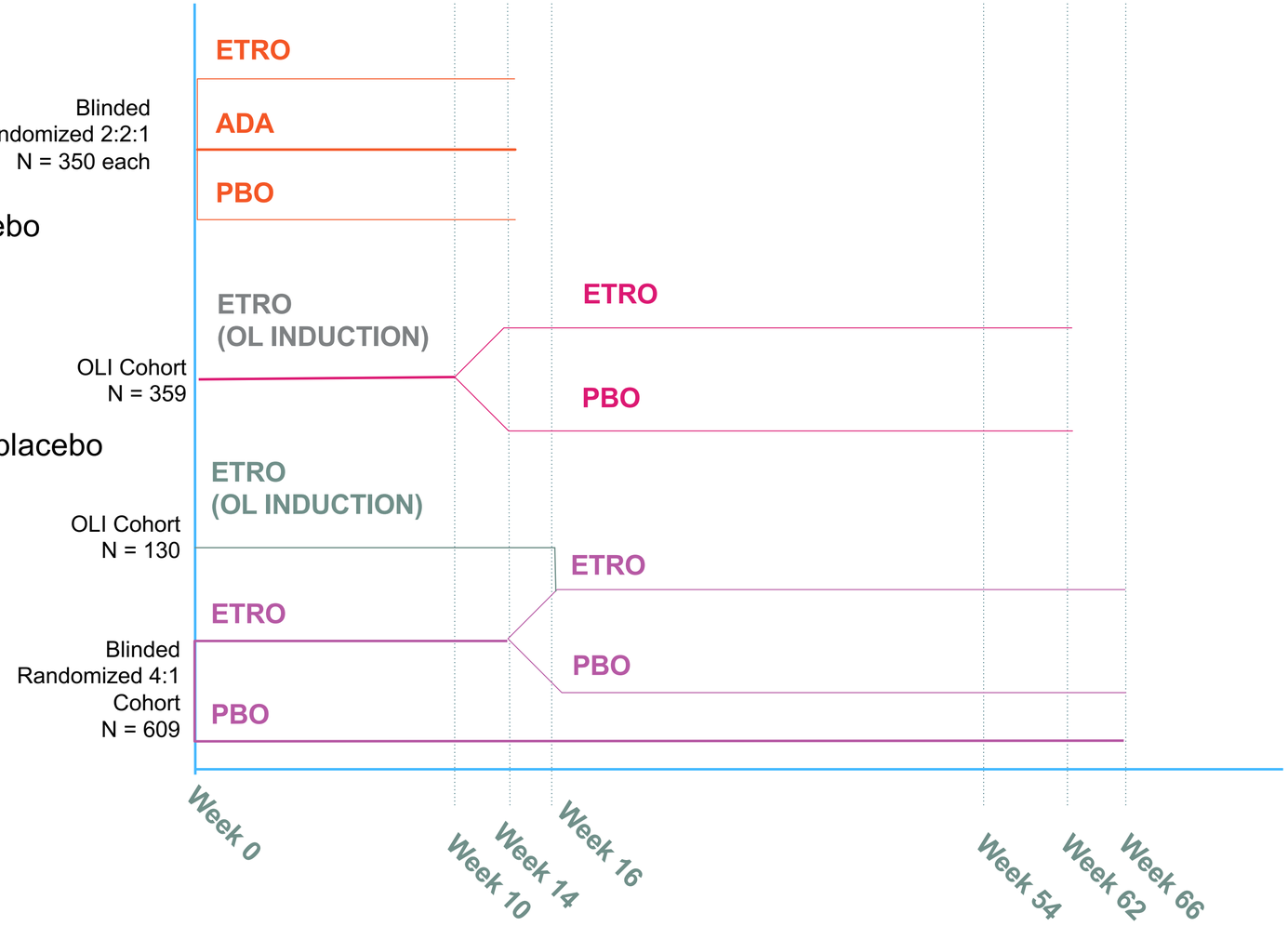
LAUREL

NCT02165215
(GA29102)
1 maintenance trial: Etrolizumab vs placebo

UC TNF-EXP

HICKORY

NCT02100696
(GA28950)
1 induction/maintenance trial:
Etrolizumab vs placebo



Analysis data came from three phase III trials performed by Genentech/Roche to evaluate the efficacy of etrolizumab

UC TNF-NAIVE

HIBISCUS I & II

NCT02163759 & NCT02171429
(GA28948 & GA28949)
2 induction trials:
Etrolizumab vs adalimumab vs placebo

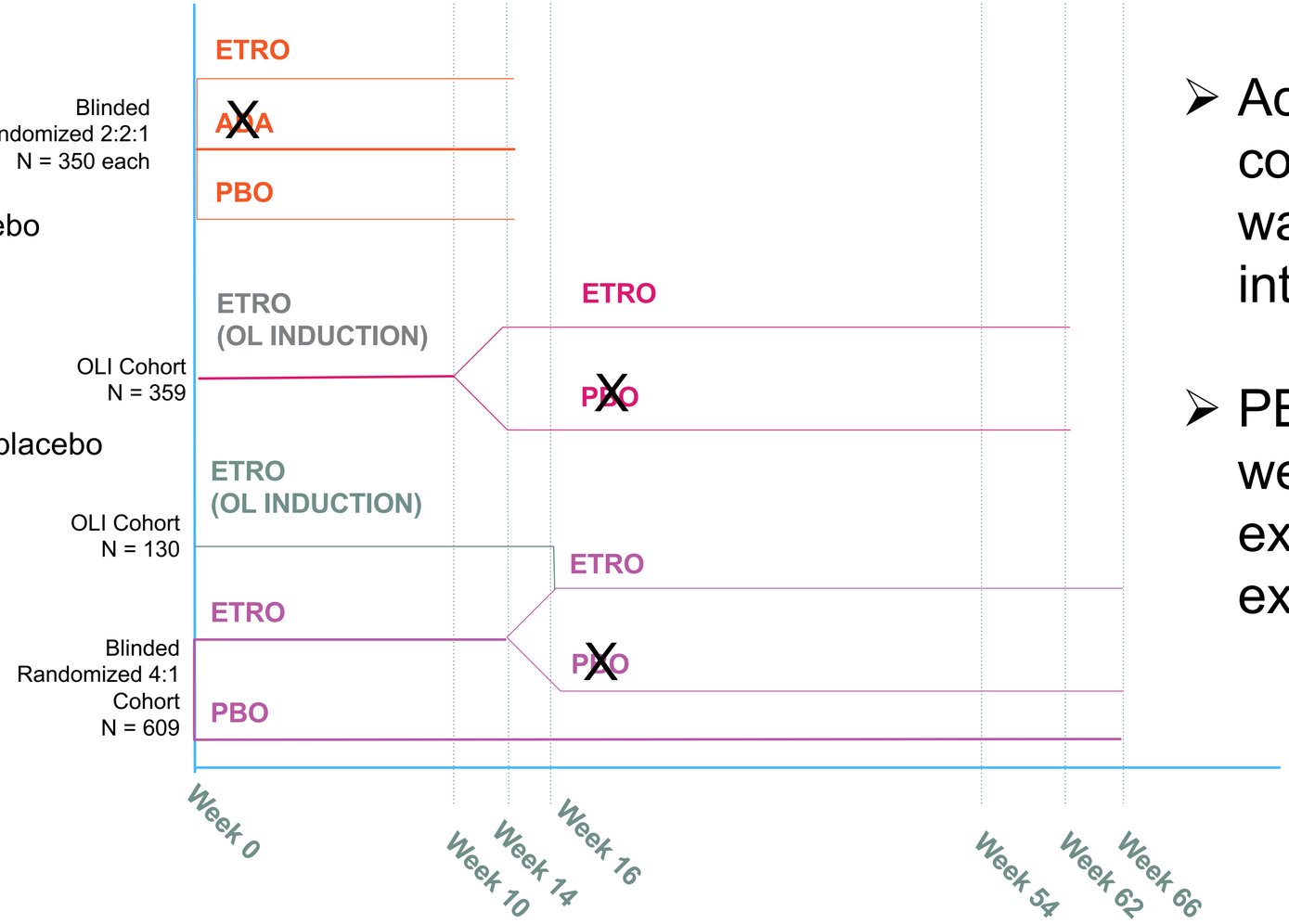
LAUREL

NCT02165215
(GA29102)
1 maintenance trial: Etrolizumab vs placebo

UC TNF-EXP

HICKORY

NCT02100696
(GA28950)
1 induction/maintenance trial:
Etrolizumab vs placebo



- Active comparator was not of interest
- PBO arms that were ETRO-exposed were excluded

Analysis data came from three phase III trials performed by Genentech/Roche to evaluate the efficacy of etrolizumab

UC TNF-NAIVE

HIBISCUS I & II

NCT02163759 & NCT02171429
(GA28948 & GA28949)
2 induction trials:
Etrolizumab vs adalimumab vs placebo

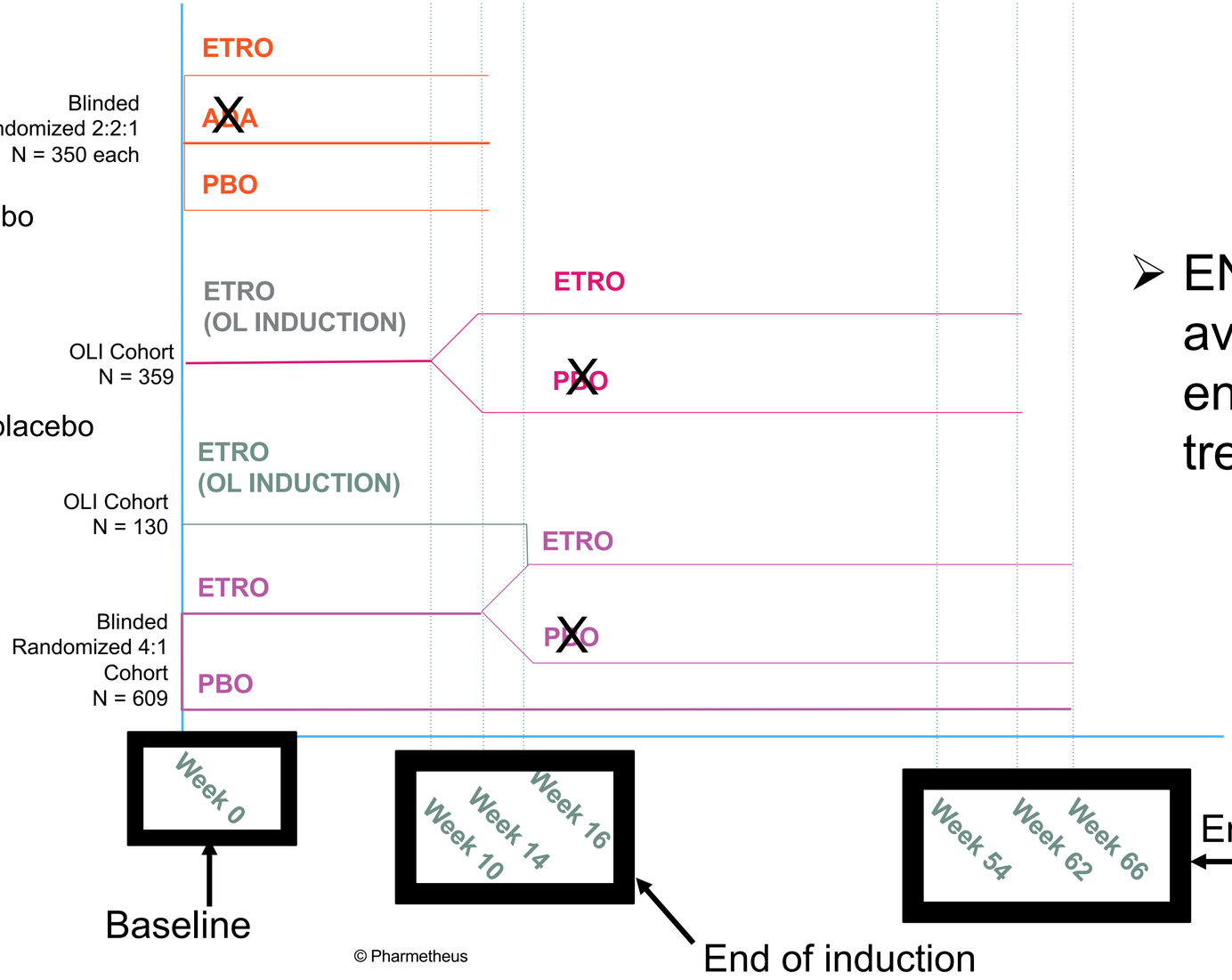
LAUREL

NCT02165215
(GA29102)
1 maintenance trial: Etrolizumab vs placebo

UC TNF-EXP

HICKORY

NCT02100696
(GA28950)
1 induction/maintenance trial:
Etrolizumab vs placebo



- END was only available at end/start of treatment phase

Analysis data came from three phase III trials performed by Genentech/Roche to evaluate the efficacy of etrolizumab

UC TNF-NAIVE

HIBISCUS I & II

NCT02163759 & NCT02171429
(GA28948 & GA28949)
2 induction trials:
Etrolizumab vs adalimumab vs placebo

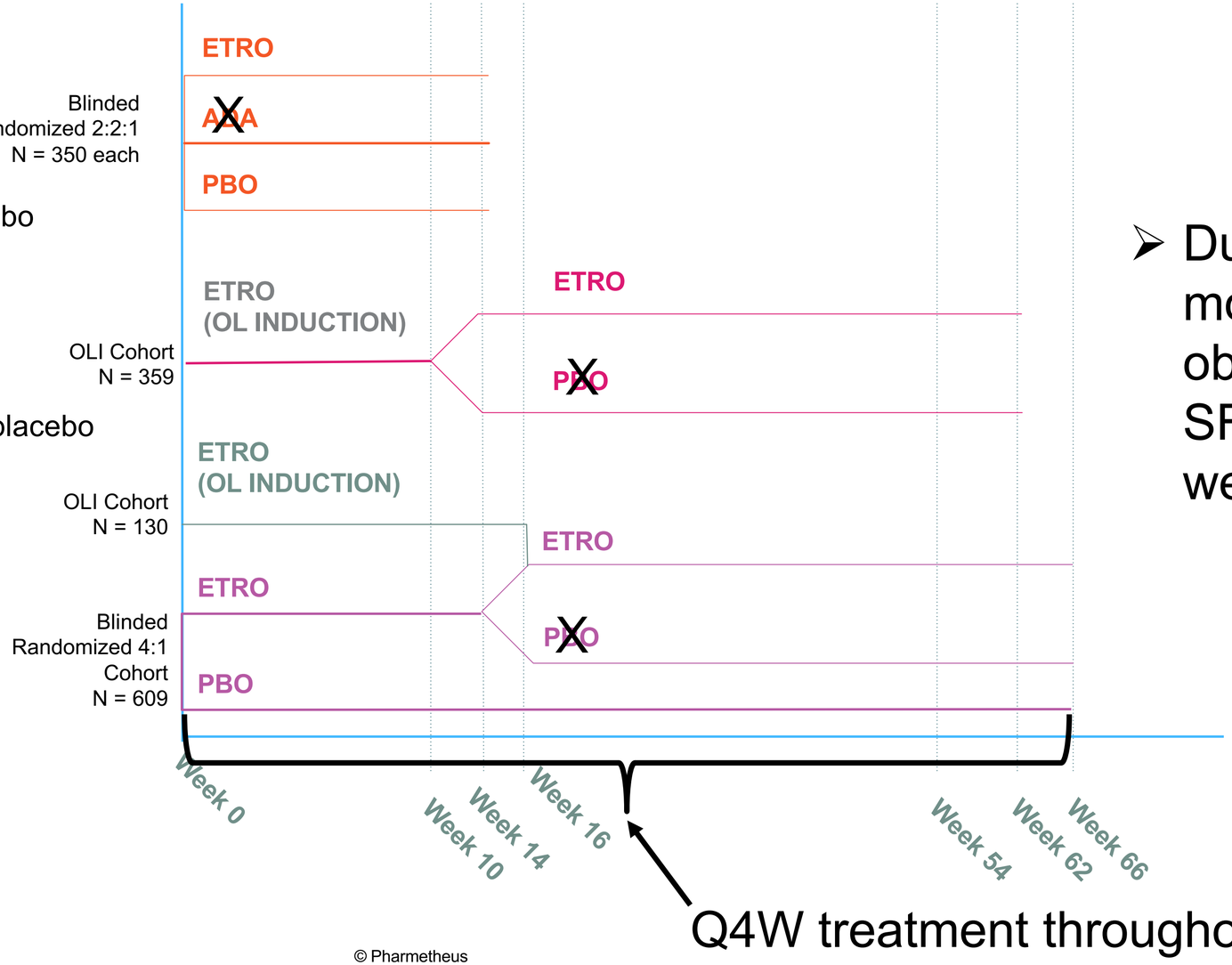
LAUREL

NCT02165215
(GA29102)
1 maintenance trial: Etrolizumab vs placebo

UC TNF-EXP

HICKORY

NCT02100696
(GA28950)
1 induction/maintenance trial:
Etrolizumab vs placebo

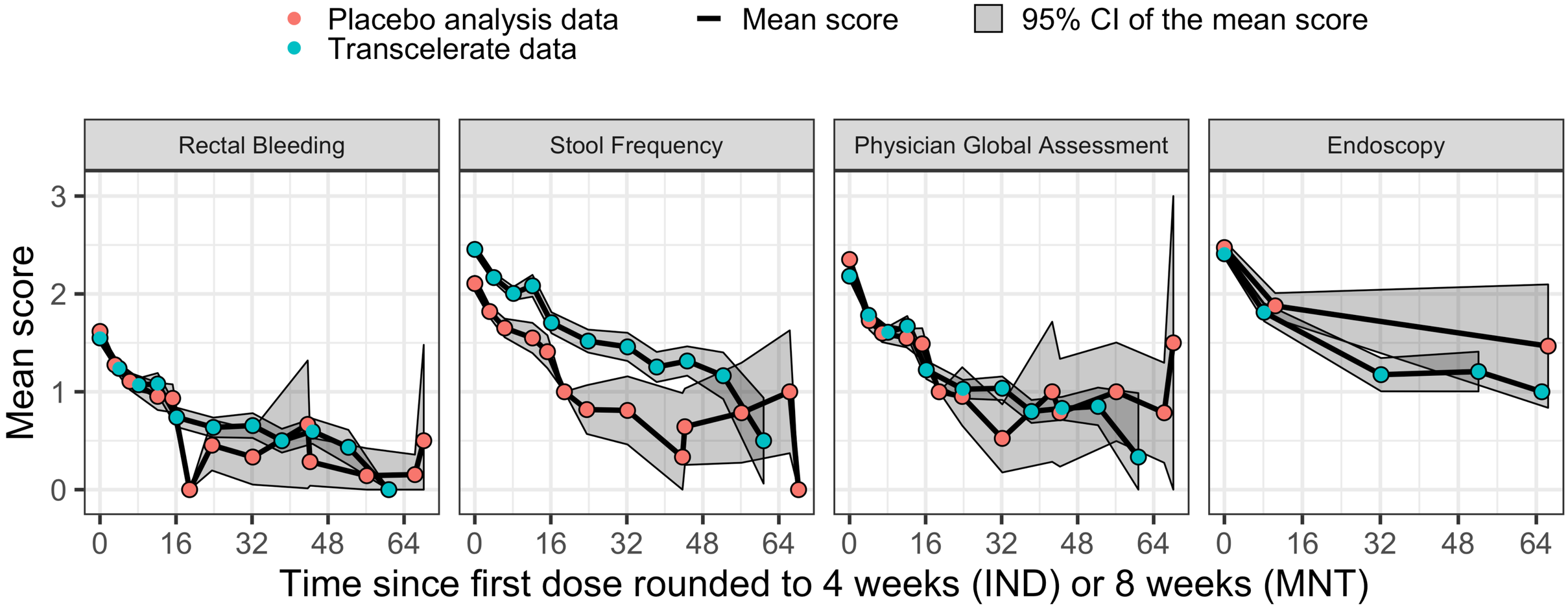


- During treatment more regular observations of SF, RB, and PGA were available

External evaluation data consisted of placebo data from five trials of various drug companies: “TransCelerate” data

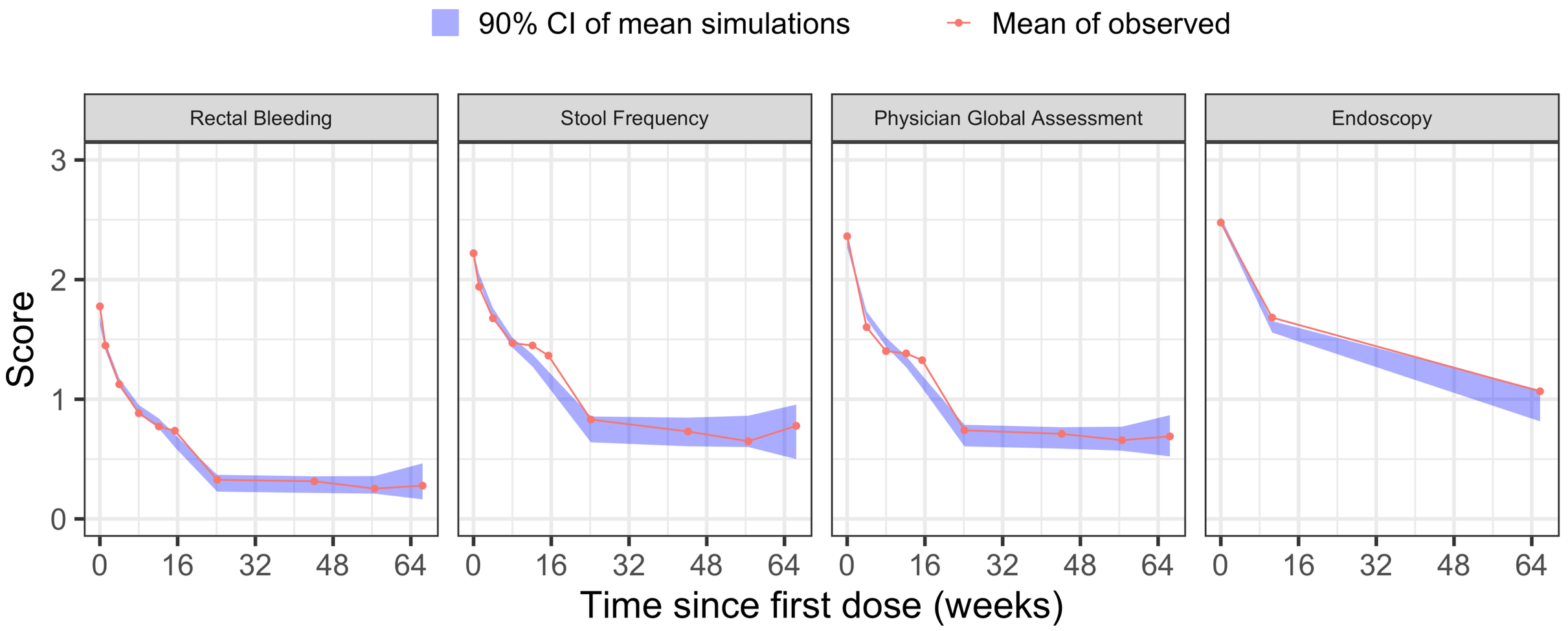
Study NCT	Company	Design	Anti-TNF therapy status	N subjects
NCT00385736	Abbvie	8 week induction	Only Naïve	222
NCT00408629	Abbvie	8 week induction + 44 week maintenance	Naïve and experienced	256
NCT00410410	BMS	12 week induction + 40 week maintenance	Mostly Naïve, but some experienced	135
NCT00787202	Pfizer	8 week induction	Mostly Naïve, but some experienced	46
NCT00853099	Abbvie	8 week induction + 44 week maintenance	Only Naïve	96

All subscores look similar except for SF that was lower for the analysis data

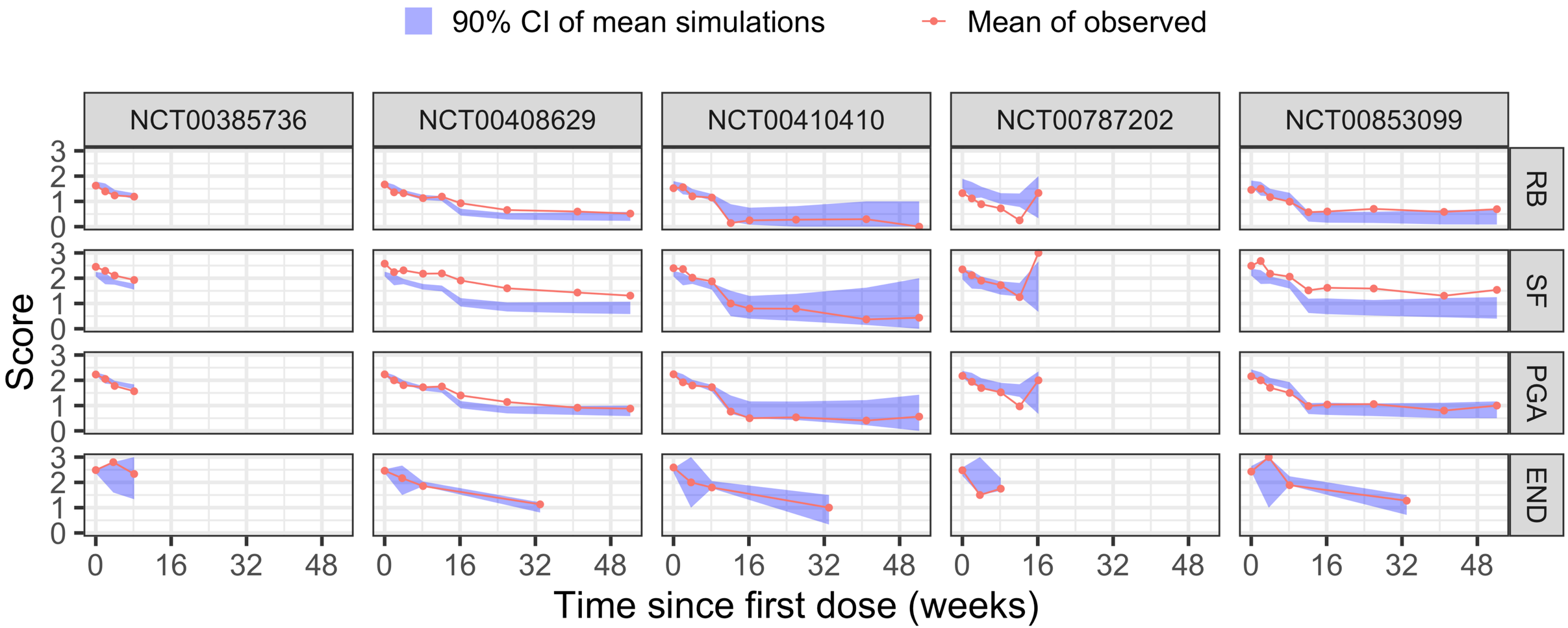


Evaluation of the suitability of the solution

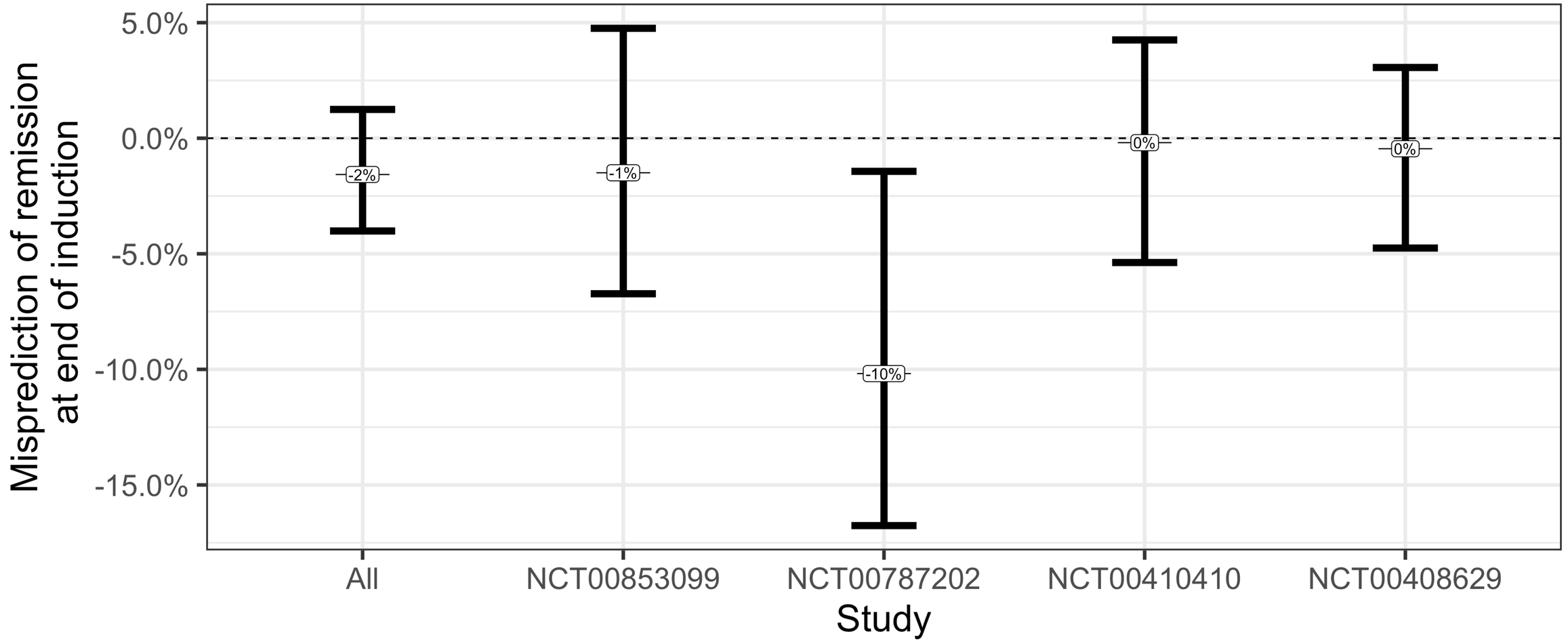
The analysis subscore data were well predicted by the model

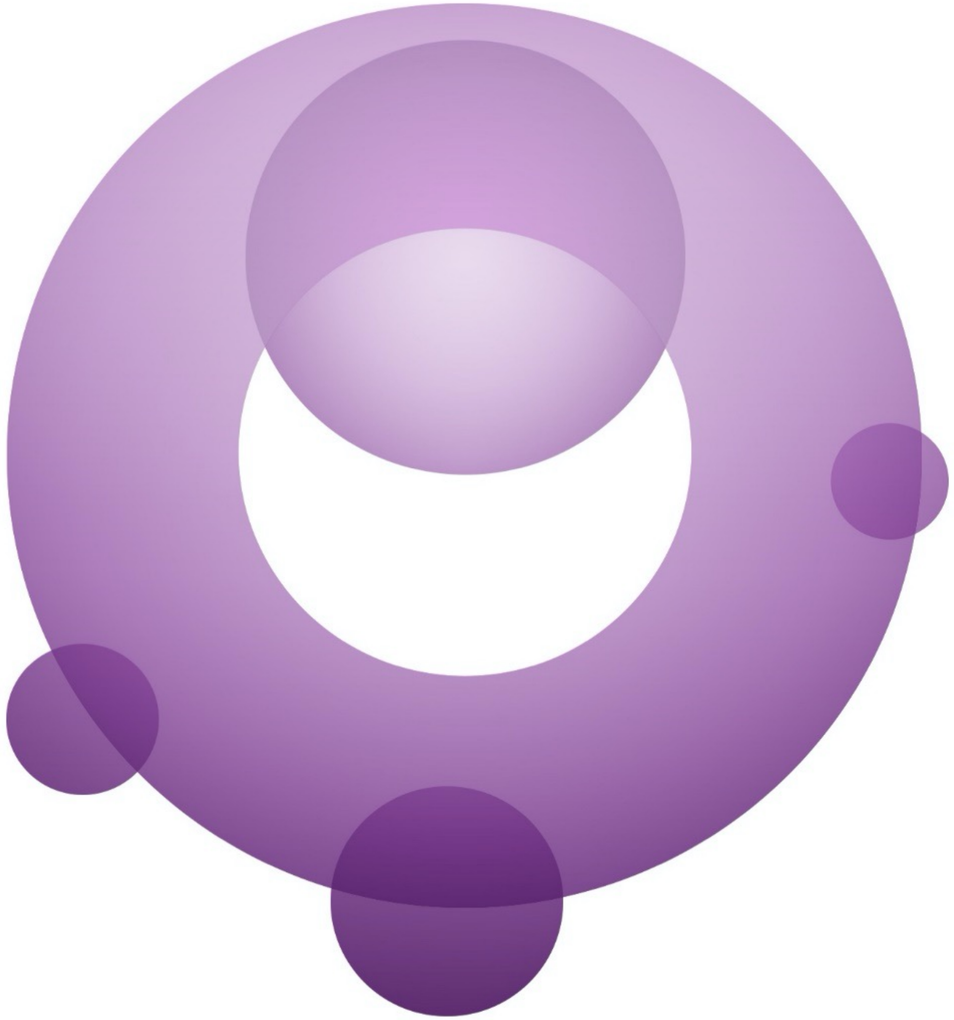


The external data were mostly well predicted by the model, though SF was underpredicted for several studies



The derived key endpoint ”remission” at end of induction in the external data was well captured by the model, except for study NCT00787202





MIDD applications

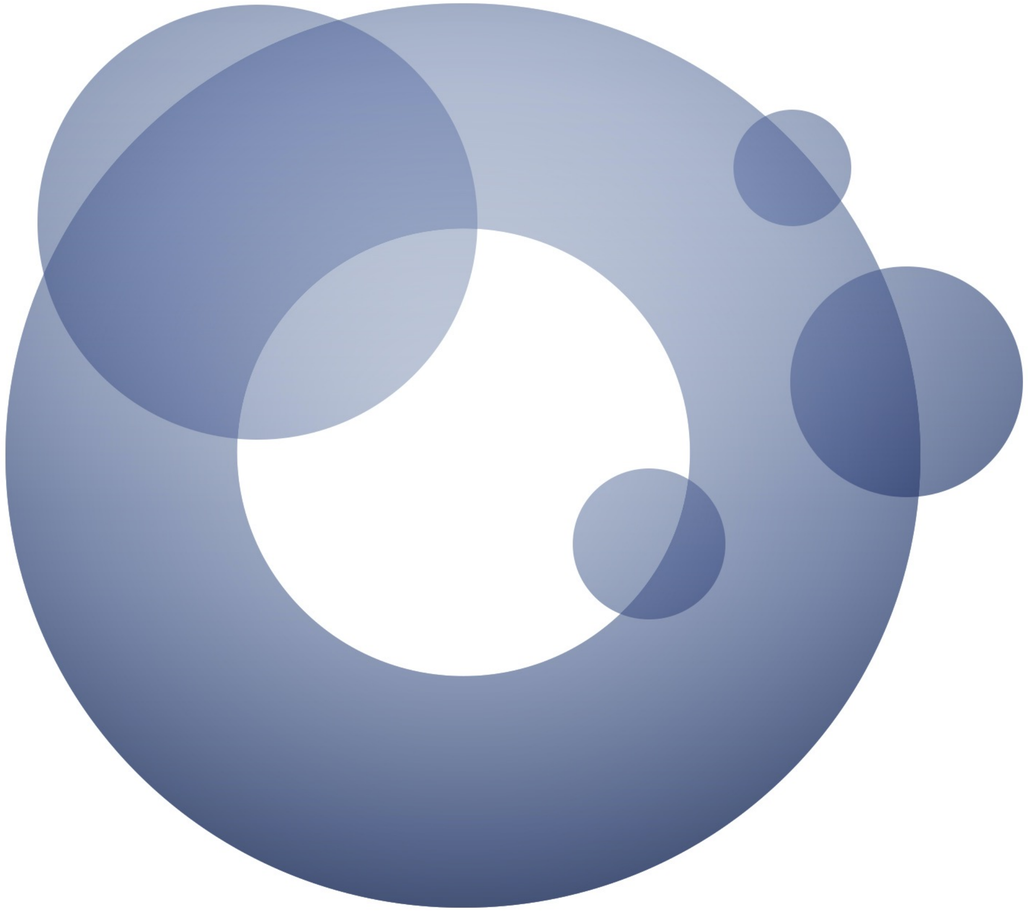
The proposed IRT model can be used to improve MIDD in UC

- Currently, decisions during drug development are mainly based on remission status at the end of treatment
  - Only 1 observation of 1 or 0 per subject
  - Difficult to impute missing data (e.g. interim analysis)
  - No possibility to extrapolate remission to other times

- Using the proposed model would leverage all key efficacy data
  - Longitidunal data of 0, 1, 2, 3
  - Allow predictions of missing individual subscore data
  - Possibility to simulate remission at unobserved time points

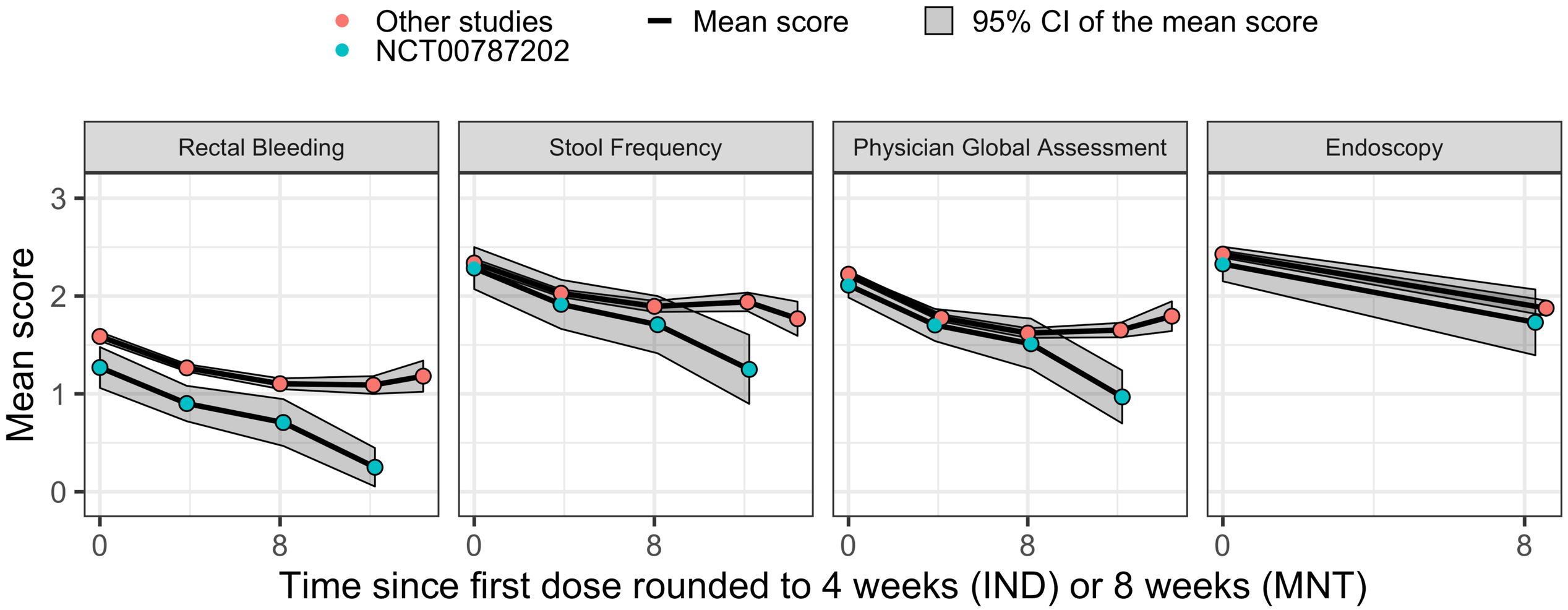
Key messages

- Remission in UC is a key clinical parameter, but it binarizes a large quantity of longitudinal integer score data
- An item response theory (IRT) model with bounded integer item models is proposed as a more powerful alternative to the binarized approach
- The IRT model described the analysis data and an external source of data adequately
- The IRT model can be used to improve MIDD in UC

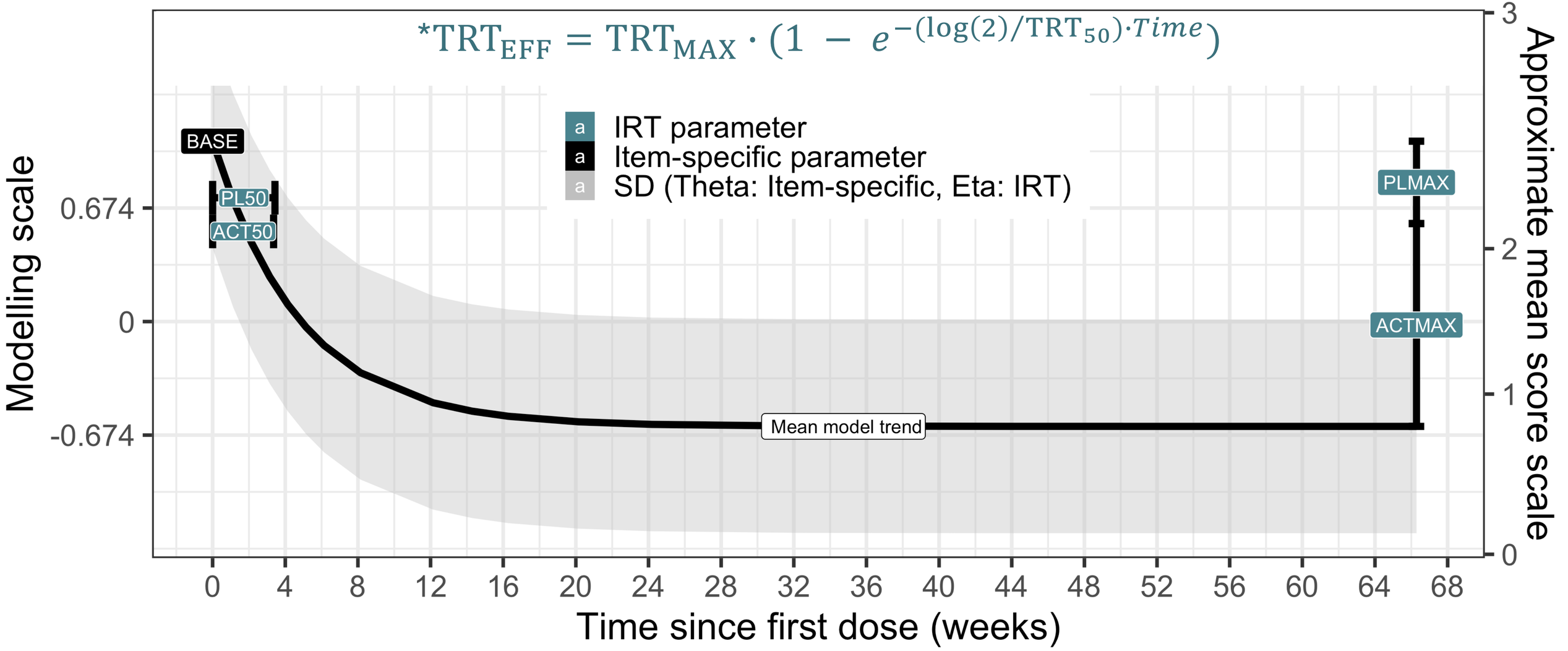


Back-up slides

Underprediction of remission in NCT00787202 is mainly due to lower rectal bleeding scores



The model behavior visualized



Assumptions

Assumption	Consequence of violation	Evaluation method
Each score is impacted the same by the treatment effect (IRT)	The model mispredicts the score for which this assumption doesn't hold	VPC per score for population level
		IPRED vs DV for individual level
Each score adds equally unique information (IRT)	Shared information across a subset of scores increases the weight of those scores that share the information	VPC per score may show misspecifcation for certain score due to increased weight of other scores.
The probability of a non-extreme score is larger than at least one of the nearest adjacent scores (B)	Could pose a problem if the distribution of scores is for some reason not unimodal (e.g. 1 and 3 are much more common than 0 and 2)	VPC per fraction of score over time