Supplemental Materials

Pre-processing details:
The data provided to us was already pre-processed. They were denoised[24], corrected for intensity heterogeneity[36], and normalized into the range 0-100 [5]. Next, for each patient, the FLAIR scans were co-registered to the structural T1w scan using a 6-parameter rigid registration and a mutual information objective function[9]. The T1w scans were then registered to an average template defining stereotaxic space. FLAIR volumes are re-sampled onto a 1 mm isotropic grid using the concatenated contrast-to-T1 and T1-to-stx transformations[8].

Statistics and Results on Low Efficacy Drugs:

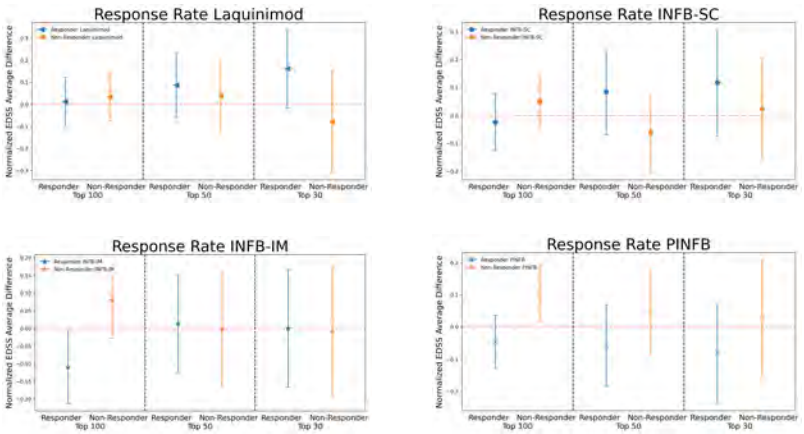

Fig. 1: Treatment response estimates for the lower efficacy treatments in RRMS trials. Comparisons are shown between responders and non-responders at three different uncertainty levels. Negative values indicate a reduction in average EDSS change on the drug as compared to placebo.

Treatment	N Patients	Subtype	$EDSS_{t_0}$	Age	% Male
Placebo (BRAVO)	317	RRMS	2.6	37.6	0.29
Placebo (DEFINE)	138	RRMS	2.5	37.5	0.22
Laquinimod	295	RRMS	2.6	36.6	0.34
INFB-IM	326	RRMS	2.6	37.7	0.32
Ocrelizumab	572	RRMS	2.8	37.2	0.35
INFB-SC	554	RRMS	2.7	37.1	0.30
Dimethyl Fumarate	241	RRMS	2.3	37.6	0.23
Peginterferon	840	RRMS	2.4	37.3	0.31
Natalizumab	293	SPMS	5.6	46.5	0.40
Placebo (ASCEND)	302	SPMS	5.6	47.1	0.37

Table 1: Dataset statistics.

Architectures for Encoder-Decoder and LSTM Models:

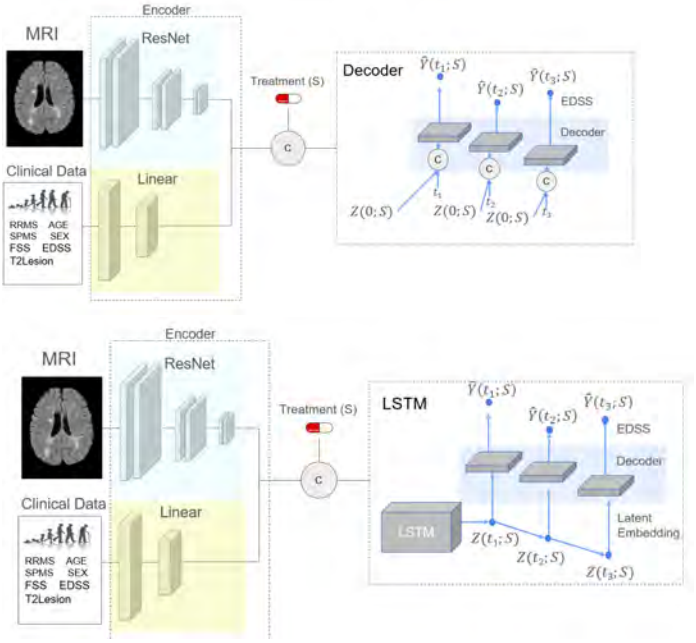

Fig. 2: (Top) Fixed time-point Encoder-Decoder EDSS regressor. The number of days to predict in the future is used as an additional concatenated feature in latent space before predicting EDSS. (Bottom) The LSTM model uses the discretized temporal information to project the latent space forward in time before predicting EDSS.