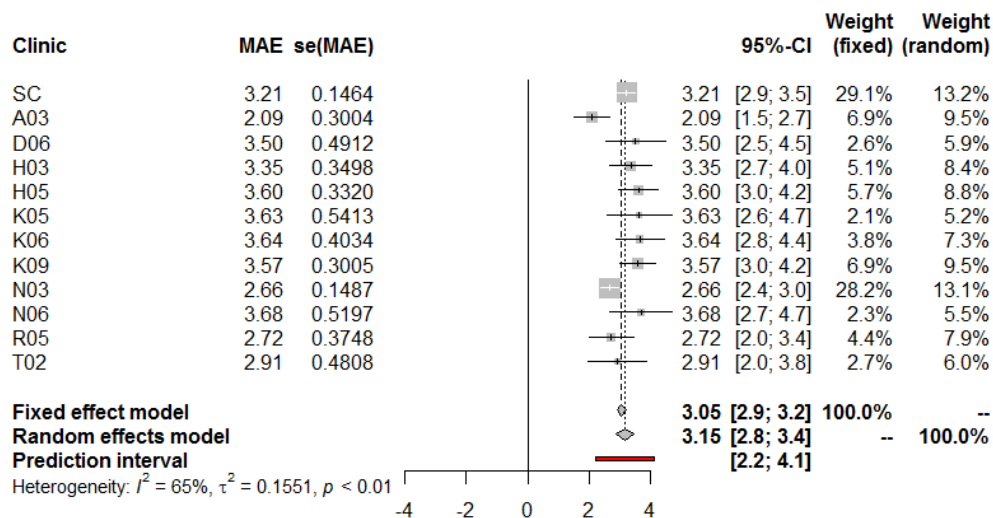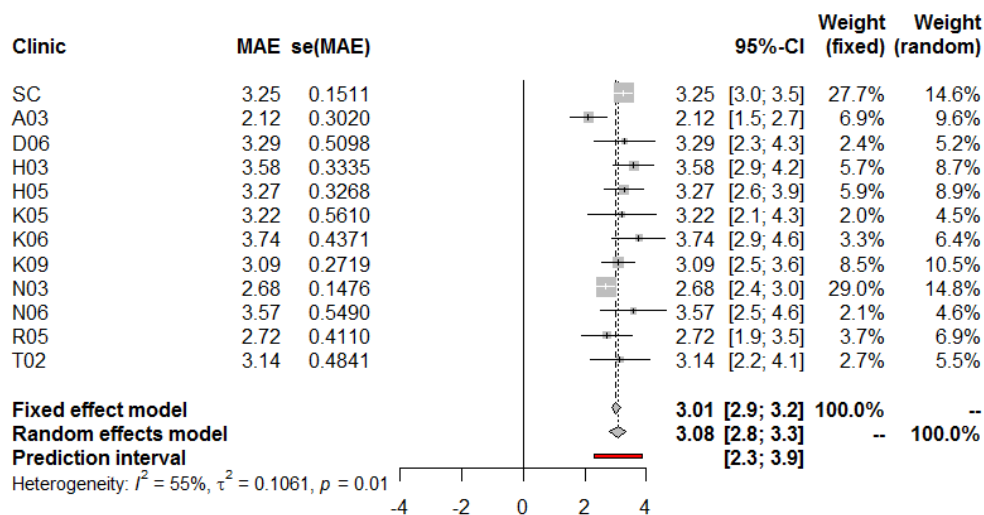## Appendix 1. Choosing the analytical model

The median absolute error (MAE) for the four different modelling approaches was as follows: LASSO regression 3.0, ridge regression 3.0, SVM 1.6, artificial neural network 2.8. The results from the internal-external cross validation (leaving one institution out at a time) are presented below, for the four different models. Based on these results we decided to use SVM.
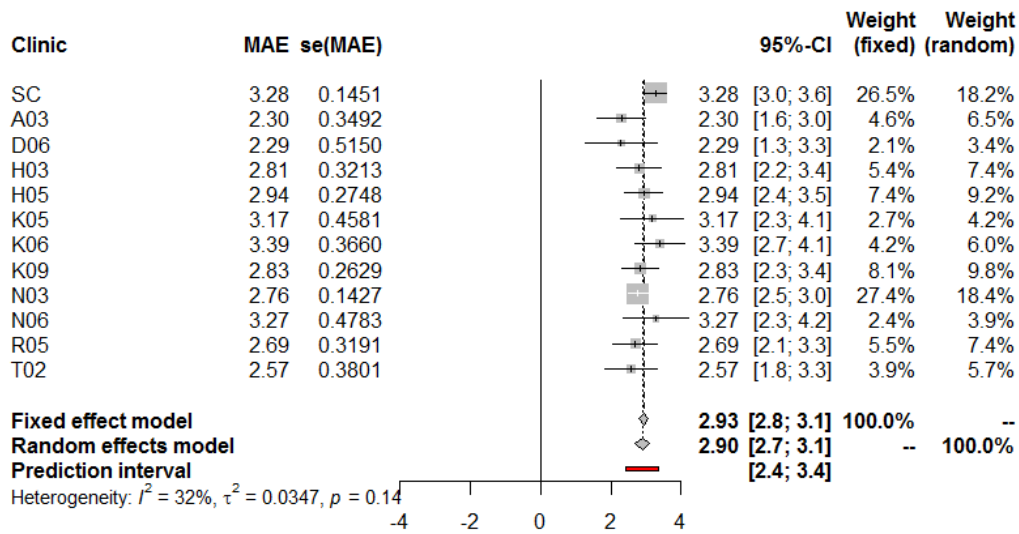
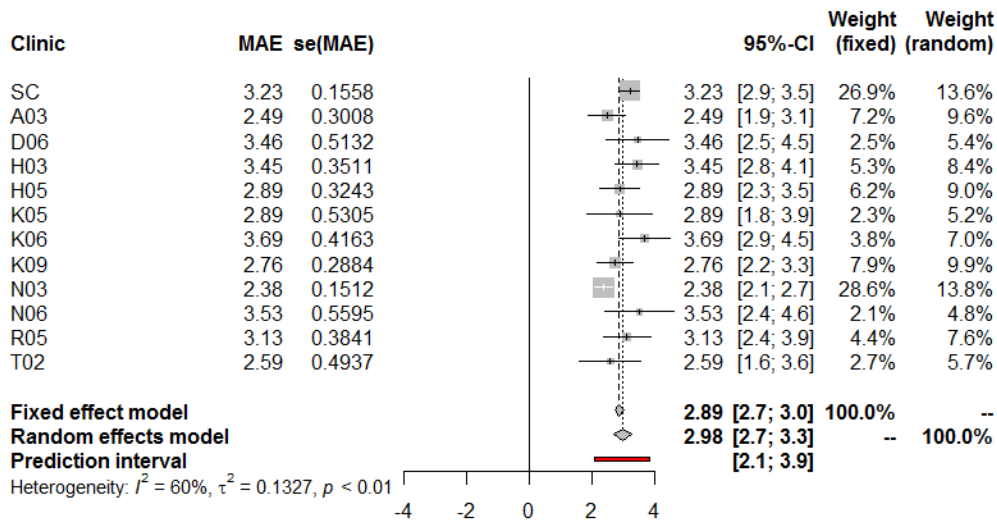**eFigure 1**. Internal-external validation using LASSO regression

eFigure 2. Internal-external validation using ridge regression

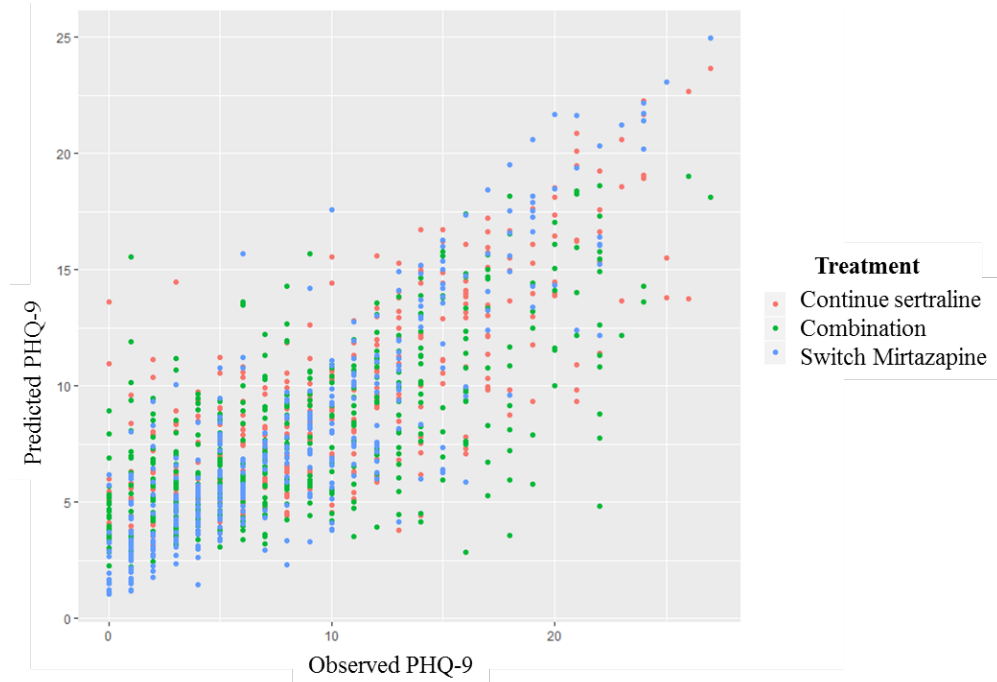**eFigure 3.** Internal-external validation using SVM (the same figure as Figure 1)

Clinic	MAE	se(MAE)		95%-CI	Weight (fixed)	Weight (random)
SC	3.28	0.1451		3.28 [3.0; 3.6]	26.5%	18.2%
A03	2.30	0.3492		2.30 [1.6; 3.0]	4.6%	6.5%
D06	2.29	0.5150		2.29 [1.3; 3.3]	2.1%	3.4%
H03	2.81	0.3213		2.81 [2.2; 3.4]	5.4%	7.4%
H05	2.94	0.2748		2.94 [2.4; 3.5]	7.4%	9.2%
K05	3.17	0.4581		3.17 [2.3; 4.1]	2.7%	4.2%
K06	3.39	0.3660		3.39 [2.7; 4.1]	4.2%	6.0%
K09	2.83	0.2629		2.83 [2.3; 3.4]	8.1%	9.8%
N03	2.76	0.1427		2.76 [2.5; 3.0]	27.4%	18.4%
N06	3.27	0.4783		3.27 [2.3; 4.2]	2.4%	3.9%
R05	2.69	0.3191		2.69 [2.1; 3.3]	5.5%	7.4%
T02	2.57	0.3801		2.57 [1.8; 3.3]	3.9%	5.7%
Fixed effect model				**2.93 [2.8; 3.1]**	**100.0%**	--
Random effects model				**2.90 [2.7; 3.1]**	--	**100.0%**
Prediction interval				**[2.4; 3.4]**		

Heterogeneity: $I^2 = 32\%$, $\tau^2 = 0.0347$, $p = 0.14$

eFigure 3. Internal-external validation using ANN

Clinic	MAE	se(MAE)		95%-CI	Weight (fixed)	Weight (random)
SC	3.23	0.1558		3.23 [2.9; 3.5]	26.9%	13.6%
A03	2.49	0.3008		2.49 [1.9; 3.1]	7.2%	9.6%
D06	3.46	0.5132		3.46 [2.5; 4.5]	2.5%	5.4%
H03	3.45	0.3511		3.45 [2.8; 4.1]	5.3%	8.4%
H05	2.89	0.3243		2.89 [2.3; 3.5]	6.2%	9.0%
K05	2.89	0.5305		2.89 [1.8; 3.9]	2.3%	5.2%
K06	3.69	0.4163		3.69 [2.9; 4.5]	3.8%	7.0%
K09	2.76	0.2884		2.76 [2.2; 3.3]	7.9%	9.9%
N03	2.38	0.1512		2.38 [2.1; 2.7]	28.6%	13.8%
N06	3.53	0.5595		3.53 [2.4; 4.6]	2.1%	4.8%
R05	3.13	0.3841		3.13 [2.4; 3.9]	4.4%	7.6%
T02	2.59	0.4937		2.59 [1.6; 3.6]	2.7%	5.7%
Fixed effect model				**2.89 [2.7; 3.0]**	**100.0%**	--
Random effects model				**2.98 [2.7; 3.3]**	--	**100.0%**
Prediction interval				**[2.1; 3.9]**		

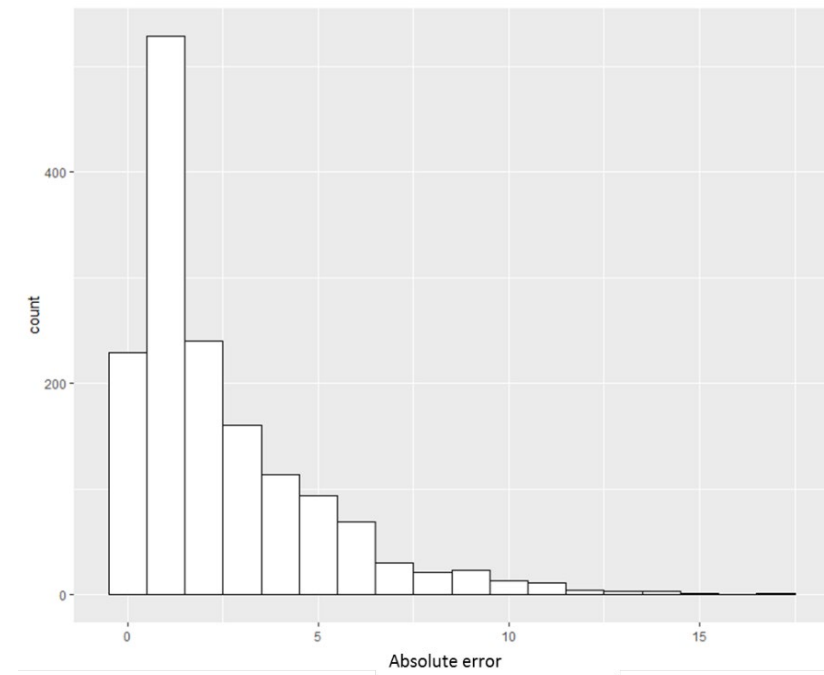Heterogeneity: $I^2 = 60\%$, $\tau^2 = 0.1327$, $p < 0.01$

2

Appendix 2. Apparent performance of the developed full model

eFigure 5. Observed versus predicted PHQ-9 at week 9 using the developed model
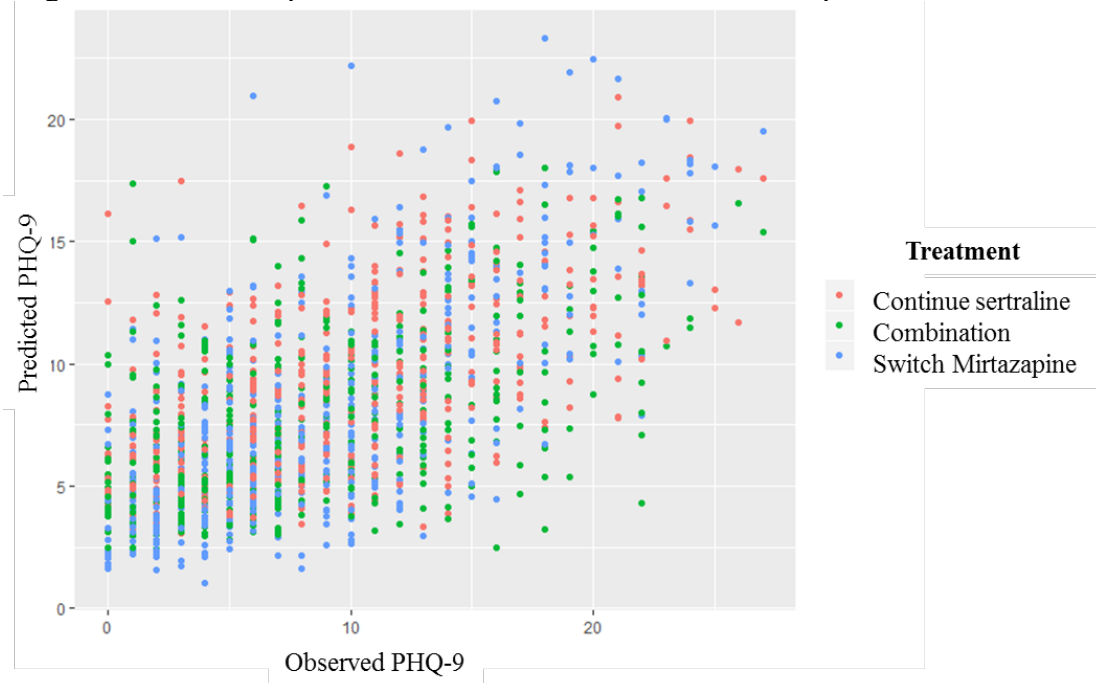

eFigure 6. Histogram of prediction errors (i.e. predicted vs. observed) using the developed model
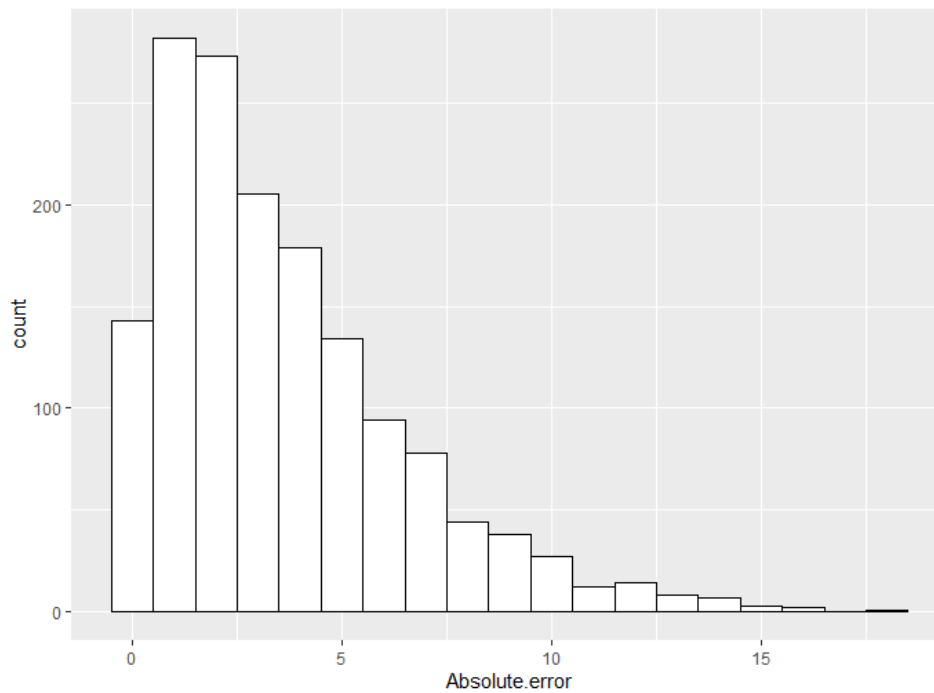
Appendix 3. Performance in leave-one-patient-out cross-validation

eFigure 7. Leave-one-patient-out cross-validation, observed vs. predicted outcomes at week 9.

eFigure 8. Histogram of prediction errors (i.e. predicted vs. observed) in leave-one-patient-out cross-validation

Appendix 4. Characteristics of the obtained subgroups

Table 1. Further details of baseline clinical characteristics

Clinical characteristics	Total sample (n=1544)	Continuing sertraline was the best (n=161)	Combining with mirtazapine was the best (n=678)	Switching to mirtazapine was the best (n=705)
PHQ-9 item scores at baseline, mean (SD)				
Item 1 (anhedonia)	2.5 (0.6)	2.6 (0.5)	2.5 (0.7)	2.5 (0.6)
Item 2 (depressed mood)	2.5 (0.6)	2.6 (0.6)	2.5 (0.6)	2.5 (0.6)
Item 3 (sleep disturbance)	2.4 (0.8)	2.4 (0.8)	2.4 (0.8)	2.3 (0.9)
Item 4 (anergia)	2.6 (0.6)	2.7 (0.4)	2.6 (0.6)	2.6 (0.6)
Item 5 (appetite disturbance)	2.0 (1.0)	2.0 (0.9)	2.0 (1.0)	2.0 (1.0)
Item 6 (low self-esteem)	2.2 (0.9)	2.3 (0.7)	2.2 (0.8)	2.1 (0.9)
Item 7 (concentration difficulties)	2.0 (0.9)	2.1 (0.8)	2.0 (0.9)	1.9 (0.9)
Item 8 (psychomotor symptoms)	1.6 (1.0)	1.8 (0.9)	1.7 (1.0)	1.6 (1.0)
Item 9 (suicidal ideations)	0.9 (0.8)	1.0 (0.8)	0.9 (0.9)	0.8 (0.8)
PHQ-9 item scores at week 1, mean (SD)				
Item 1 (anhedonia)	2.1 (0.8)	2.2 (0.7)	2.2 (0.7)	2.0 (0.8)
Item 2 (depressed mood)	2.1 (0.8)	2.2 (0.7)	2.3 (0.7)	1.9 (0.8)
Item 3 (sleep disturbance)	2.1 (0.9)	2.0 (0.9)	2.1 (0.9)	2.0 (1.0)
Item 4 (anergia)	2.4 (0.7)	2.6 (0.6)	2.5 (0.7)	2.3 (0.8)
Item 5 (appetite disturbance)	1.8 (1.0)	1.8 (1.0)	1.7 (1.0)	1.9 (1.0)
Item 6 (low self-esteem)	1.8 (1.0)	2.3 (0.7)	2.0 (0.9)	1.6 (0.9)
Item 7 (concentration difficulties)	1.6 (1.0)	1.8 (0.9)	1.8 (1.0)	1.4 (1.0)

Item 8 (psychomotor symptoms)	1.2 (1.0)	1.2 (0.9)	1.4 (1.0)	1.1 (0.9)
Item 9 (suicidal ideations)	0.7 (0.8)	0.9 (0.7)	0.8 (0.8)	0.7 (0.8)
PHQ-9 item scores at week 3, mean (SD)				
Item 1 (anhedonia)	1.7 (0.8)	1.8 (0.8)	1.8 (0.8)	1.5 (0.8)
Item 2 (depressed mood)	1.6 (0.8)	1.9 (0.8)	1.8 (0.8)	1.4 (0.7)
Item 3 (sleep disturbance)	1.8 (1.0)	2.0 (0.9)	1.8 (1.0)	1.8 (1.0)
Item 4 (anergia)	2.0 (0.8)	2.3 (0.7)	2.1 (0.8)	1.8 (0.8)
Item 5 (appetite disturbance)	1.5 (1.1)	1.6 (1.0)	1.5 (1.0)	1.5 (1.0)
Item 6 (low self-esteem)	1.5 (1.0)	2.0 (0.9)	1.7 (0.9)	1.2 (0.9)
Item 7 (concentration difficulties)	1.3 (1.0)	1.6 (0.9)	1.4 (1.0)	1.1 (0.9)
Item 8 (psychomotor symptoms)	0.9 (0.9)	1.1 (0.9)	1.1 (1.0)	07 (0.8)
Item 9 (suicidal ideations)	0.5 (0.7)	0.7 (0.8)	0.6 (0.8)	0.4 (0.6)
BDI-II scores at week 1				
Item 1	1.5 (0.7)	1.6 (0.7)	1.7 (0.7)	1.3 (0.7)
Item 2	1.5 (0.7)	1.9 (0.6)	1.7 (0.7)	1.3 (0.7)
Item 3	1.4 (0.8)	1.7 (0.7)	1.6 (0.8)	1.2 (0.8)
Item 4	1.6 (0.7)	1.7 (0.7)	1.7 (0.8)	1.4 (0.7)
Item 5	1.1 (0.8)	1.5 (0.7)	1.3 (0.8)	0.9 (0.7)
Item 6	1.2 (1.0)	1.7 (1.0)	1.4 (1.0)	1.0 (1.0)
Item 7	1.6 (0.9)	1.8 (0.8)	1.6 (0.9)	1.4 (0.8)
Item 8	1.5 (0.9)	1.9 (0.7)	1.7 (0.9)	1.3 (0.9)
Item 9	0.8 (0.7)	1.0 (0.6)	0.8 (0.7)	0.7 (0.7)
Item 10	1.3 (0.9)	1.4 (0.8)	1.5 (0.9)	1.1 (0.9)
Item 11	1.2 (0.7)	1.6 (0.8)	1.2 (0.8)	1.0 (0.7)
Item 12	1.4 (0.8)	1.5 (0.7)	1.5 (0.8)	1.2 (0.7)

Item 13	1.4 (0.8)	1.8 (0.8)	1.5 (0.9)	1.2 (0.8)
Item 14	1.4 (0.8)	1.9 (0.7)	1.5 (0.8)	1.2 (0.8)
Item 15	1.6 (0.7)	1.8 (0.7)	1.7 (0.7)	1.6 (0.6)
Item 16	1.5 (0.9)	1.5 (0.9)	1.5 (0.9)	1.5 (0.9)
Item 17	1.0 (0.9)	1.2 (0.9)	1.1 (0.9)	0.8 (0.8)
Item 18	1.2 (0.8)	1.2 (0.8)	1.0 (0.8)	1.3 (0.8)
Item 19	1.5 (0.7)	1.9 (0.7)	1.6 (0.8)	1.4 (0.7)
Item 20	1.6 (0.7)	1.9 (0.7)	1.7 (0.8)	1.4 (0.7)
Item 21	1.5 (1.1)	1.5 (1.0)	1.7 (1.1)	1.4 (1.0)
BDI-II scores at week 3				
Item 1	1.2 (0.7)	1.6 (0.7)	1.4 (0.7)	1.0 (0.6)
Item 2	1.3 (0.8)	1.7 (0.7)	1.5 (0.7)	1.0 (0.7)
Item 3	1.2 (0.8)	1.6 (0.7)	1.4 (0.7)	1.0 (0.8)
Item 4	1.3 (0.7)	1.5 (0.8)	1.5 (0.8)	1.1 (0.6)
Item 5	0.9 (0.8)	1.4 (0.8)	1.1 (0.8)	0.7 (0.7)
Item 6	1.1 (1.0)	1.6 (1.0)	1.2 (1.0)	0.8 (0.9)
Item 7	1.3 (0.8)	1.9 (0.8)	1.4 (0.8)	1.1 (0.8)
Item 8	1.2 (0.9)	1.7 (0.8)	1.4 (0.9)	1.0 (0.8)
Item 9	0.6 (0.7)	0.8 (0.7)	0.7 (0.7)	0.5 (0.6)
Item 10	1.0 (0.9)	1.4 (1.0)	1.2 (1.0)	0.7 (0.8)
Item 11	1.0 (0.7)	1.4 (0.7)	1.1 (0.7)	0.8 (0.6)
Item 12	1.2 (0.7)	1.5 (0.8)	1.3 (0.8)	1.1 (0.7)
Item 13	1.2 (0.8)	1.6 (0.8)	1.3 (0.8)	1.0 (0.7)
Item 14	1.2 (0.8)	1.7 (0.7)	1.4 (0.8)	1.0 (0.8)
Item 15	1.4 (0.7)	1.6 (0.6)	1.5 (0.7)	1.3 (0.6)
Item 16	1.3 (0.8)	1.6 (0.9)	1.2 (0.8)	1.3 (0.8)
Item 17	0.7 (0.8)	1.2 (0.9)	0.8 (0.8)	0.5 (0.7)

Item 18	1.0 (0.8)	1.1 (0.8)	1.0 (0.7)	1.0 (0.8)
Item 19	1.3 (0.7)	1.7 (0.7)	1.4 (0.7)	1.2 (0.6)
Item 20	1.4 (0.8)	1.8 (0.8)	1.5 (0.8)	1.2 (0.7)
Item 21	1.5 (1.1)	1.6 (1.0)	1.6 (1.1)	1.3 (1.0)
FIBSER item scores at week 1, mean (SD)				
Item 1 (frequency)	2.4 (1.7)	2.7 (1.6)	2.6 (1.7)	2.3 (1.6)
Item 2 (intensity)	2.4 (1.4)	2.9 (1.5)	2.5 (1.5)	2.3 (1.4)
Item 3 (burden)	1.9 (1.2)	2.3 (1.3)	2.0 (1.3)	1.7 (1.1)
FIBSER item scores at week 3, mean (SD)				
Item 1 (frequency)	2.5 (1.7)	2.0 (1.6)	2.9 (1.7)	2.3 (1.6)
Item 2 (intensity)	2.5 (1.5)	2.2 (1.3)	2.9 (1.5)	2.2 (1.3)
Item 3 (burden)	2.0 (1.3)	1.8 (1.1)	2.3 (1.4)	1.78(1.1)

CI: confidence interval. SD: standard deviation.

BDI-II: Beck Depression Inventory, 2nd edition. FIBSER: Frequency, Intensity and Burden of Side Effects Rating. PHQ-9: Patient Health Questionnaire-9

Appendix 4. Snapshots from the web app

For details, please visit https://cinema.ispm.unibe.ch/shinies/sund/.