Online Supplement

Supplement 1. Statistical details of variable selection and IPD network meta-regression

Supplement 2. Data description

Figure S1. PRISMA flowchart for selection of studies

Table S1: Data availability for depression severity by week

Figure S2. Pooled, aggregated data from all the three studies, for 24-item HAM-D at different time points

Supplement 3. Parameter estimates for prognostic factors and effect modifiers

Table S2. Selected prognostic factors (PFs) and effect modifiers (EFs) for change in depression severity and dropout for any reason

Table S3. Parameter estimates for the IPD network meta-regression model for depression severity

Table S4. Parameter estimates for the IPD network meta-regression model for dropouts for any reason

Table S5. Parameter estimates for the IPD network meta-regression model for deterioration

Supplement 4. Examination of inconsistency

Supplement 5. Prediction model in Excel

Supplement 1. Statistical details of variable selection and IPD network meta-regression

1 Individual-patients network meta-analysis model

We used a one-stage, repeated-measurements, individual participant data network meta-analysis (IPD-NMA) model. The model jointly synthesizes information on multiple time points per patient, while stochastically imputing missing outcomes at the same step.

<u>*First, we describe the imputation part of the model*</u>. Assume patient i in study j received treatment $treat_{ij}$. For each patient we measure the outcome at different time points after study's initiation. Let us denote with k the timepoint of an observation (where in our analysis $k = 1,2,4,5,6$ for six different time points, weeks 2, 4, 6, 8, 10 and 12), and let us group a patient's observations in the following vector: $\boldsymbol{sev_{ij}} = (s_{ij1}, s_{ij2}, s_{ij3}, s_{ij4}, s_{ij5}, s_{ij6})$. We assume that the (repeated) observations from each patient follow a multivariate normal distribution, i.e. $\boldsymbol{sev_{ij}} \sim N(\boldsymbol{M_{ij}}, \boldsymbol{\Sigma})$, where $\boldsymbol{M_{ij}} = (\mu_{ij1}, \mu_{ij2}, \mu_{ij3}, \mu_{ij4}, \mu_{ij5}, \mu_{ij6})$. This means that all missing observations are imputed assuming that the missing observations are missing at random, i.e. that we can predict the unobserved outcomes for people dropping out based solely on observed data they have provided, without any extra assumption. The variance-covariance matrix $\boldsymbol{\Sigma}$ is assumed to be common for all studies, and it incorporates the correlation between the patient-level observations at different time points. This way missing outcome observations are imputed using the rest of the observations for each patient, as well as the correlations between the multiple measurements, as observed in other patients. For the inverse of $\boldsymbol{\Sigma}$ we employ a Wishart prior distribution $W(\boldsymbol{V}, n)$, where $n = 6$ and the scale matrix was chosen to have an "autoregressive" form, because observations closer in time are expected to be more highly correlated:

$$V = \begin{pmatrix} 1 & . & . & . & . & . \\ 0.5^1 & 1 & . & . & . & . \\ 0.5^2 & 0.5^1 & 1 & . & . & . \\ 0.5^3 & 0.5^2 & 0.5^1 & 1 & . & . \\ 0.5^4 & 0.5^3 & 0.5^2 & 0.5^1 & 1 & . \\ 0.5^5 & 0.5^4 & 0.5^3 & 0.5^2 & 0.5^1 & 1 \end{pmatrix}$$

<u>*Second, we describe our IPD NMA model.*</u> The model is set in a Bayesian background. First we choose a reference treatment for the network, denoted by T_{ref}. For μ_{ijk} we assume that:

$$\mu_{ijk} = \begin{cases} u_{jk} \quad , \quad if\ treat_{ij} = T_{ref} \\ u_{jk} + \Delta_{treat_i,k} \quad , \quad if\ treat_{ij} \neq T_{ref} \end{cases}$$

In this equation u_{jk} corresponds to the mean outcome for patients in study j who receive the reference treatment, at timepoint k. Note that we employ the fixed effects assumption, since we did not have enough studies to estimate heterogeneity. In order to model the time dependence of the mean outcome, we assume that $u_{jk} = u_{j1} + \beta_0(k-1)$. This is motivated by Figure S2, where there is a clear time trend in the patients' outcome over time. We assume that $u_{j1} \sim N(m_u, s_u^2)$. $\Delta_{treat_i,k}$ corresponds to the estimated relative effects of $treat_i$ vs. the reference at time point k. The consistency equations required for a NMA are automatically incorporated in the

model, for each time point. Note that the imputation models for the missing outcomes, as well as the estimation model are fit simultaneously.

The above description of the model pertains to the continuous outcome, for which we had multiple observations per patient, at different time points. For the dichotomous outcome we only analysed the outcome at 12 weeks. For these outcomes the model was changed as follows:

$$logit(p_{ij}) = \begin{cases} u_j \quad , \quad if\ treat_i = T_{ref} \\ u_j + \Delta_{treat_i} \quad , \quad if\ treat_i \neq T_{ref} \end{cases}$$

where p_{ij} denotes the probability of having an event for the outcome under examination.

2 Individual-patients network meta-regression model

We extended the IPD NMA model described above to also include the covariates identified in the variable selection procedure. The model jointly synthesized information on multiple time points per patient, while stochastically imputing missing covariates and missing outcomes at the same step.

<u>*First, we describe the imputation part of the model.*</u> For each patient covariate we used a study-specific distribution for the imputation. This way we stochastically impute the unreported patient covariates, taking into account uncertainty about missing outcomes. This assumes that covariates are missing at random. We use univariate distributions (separate distribution for each covariate), with study-specific parameters. The distributions we used depended on the nature of the covariate (continuous vs. binary). E.g. for patient i in study j we assumed for the continuous covariate `IDS anxiety factor` $IDS_{ij} \sim N(\mu.IDS_j, s.IDS_j)$. For the binary covariate `Neglect` we used a Bernoulli distribution, etc. For the missing outcome values we used the same strategy as described in the IPD NMA model, i.e. $\boldsymbol{sev_{ij}} \sim N(\boldsymbol{M_{ij}}, \boldsymbol{\Sigma})$, where $\boldsymbol{M_{ij}} = (\mu_{ij1}, \mu_{ij2}, \mu_{ij3}, \mu_{ij4}, \mu_{ij5}, \mu_{ij6})$.

<u>*Second, we describe our primary meta-analysis model.*</u> The model is set in a Bayesian background. First we choose a reference treatment for the network, denoted by T_{ref}. Assume patient i in study j received treatment $treat_{ij}$. For this patient we have vectors $\boldsymbol{PF_i}$ and $\boldsymbol{EM_i}$. For μ_{ijk} we assume that:

$$\mu_{ijk} = \begin{cases} u_{jk} + \boldsymbol{\beta}\ \boldsymbol{PF_i} \quad , \quad if\ treat_{ij} = T_{ref} \\ u_{jk} + \boldsymbol{\beta}\ \boldsymbol{PF_i} + \boldsymbol{\gamma_{treat_i}} \boldsymbol{EM_i} + \Delta_{treat_i,k} \quad , \quad if\ treat_{ij} \neq T_{ref} \end{cases}$$

In this equation u_{jk} corresponds to the mean outcome for patients in study j who receive the reference treatment, at timepoint k, with the PFs equal to their mean values (for all PFs). The parameters of the model can be interpreted as follows:

$\boldsymbol{\beta}$: vector of regression coefficients for PFs. This vector quantifies the prognostic value of the covariates

$\boldsymbol{\gamma_{treat_i}}$: vector of regression coefficients for EMs. This vector quantifies the effect modification (treatment-covariate interaction), for each of the covariates, for each comparison vs. the reference

$\Delta_{treat_i,k}$: the estimated relative effects of $treat_i$ vs. the reference at time point k, for mean values of the EMs.

The consistency equations required for a NMA are automatically incorporated in the model, both at the

$\Delta_{treat_i,k}$ as well as the $\boldsymbol{\gamma_{treat_i}}$, for each time point. Note that the imputation models for the missing covariates and missing outcomes, as well as the estimation model are fit simultaneously.

The above description of the model pertains to the continuous outcome, for which we had multiple observations per patient, at different time points. For the dichotomous outcome we only analysed the outcome at 12 weeks. For these outcomes the model was changed as follows:

$$logit(p_{ij}) = \begin{cases} u_j + \boldsymbol{\beta}\, \boldsymbol{PF_i} \quad , \quad if\ treat_i = T_{ref} \\ u_j + \boldsymbol{\beta}\, \boldsymbol{PF_i} + \boldsymbol{\gamma_{treat_i}} \boldsymbol{EM_i} + \Delta_{treat_i} \quad , \quad if\ treat_i \neq T_{ref} \end{cases}$$

where p_{ij} denotes the probability of having an event for the outcome under examination. Also note that:

- For the dichotomous outcome of dropout we focused at 12 weeks, and thus we excluded Schramm et al. (2015) because its duration was 8 weeks. There were no missing values for this outcome.

3 Estimating inconsistency

We used the design-by-treatment inconsistency model to estimate inconsistency in our network (1,2). This model "bends" the consistency equations by adding a number of inconsistency factors (IFs) to the model. In our network two inconsistency factors were in principle needed per time point. The model for the primary outcome can be written as

$$\mu_{ijk} = \begin{cases} u_{jk} + \boldsymbol{\beta}\, \boldsymbol{PF_i} \quad , \quad if\ treat_{ij} = T_{ref} \\ u_{jk} + \boldsymbol{\beta}\, \boldsymbol{PF_i} + \boldsymbol{\gamma_{treat_i}} \boldsymbol{EM_i} + \Delta_{treat_i,k} + IF_{jk} \quad , \quad if\ treat_{ij} \neq T_{ref} \end{cases}$$

where $IF_{1k} = 0$ for all k, because study 1 (Keller) is a three-armed study (consistent by definition). The third study (Lundbeck) only provides observations at $k = 4$ (8 weeks), so that $IF_{3k} = 0$ for $k \neq 4$. An IF significantly different than 0 would point to significant inconsistency. A similar model can be written for the dichotomous outcome.

For the first primary outcome (efficacy) we focused at 8 weeks (where all studies provided information), and thus only two IFs were relevant, i.e. IF_{24} and IF_{34}. For the second primary outcome (dropout) only one IF was needed, as all analyses were performed at 12 weeks.

4 Selection of variables to be included in the meta-regression model

Here we describe the strategy we used to choose which patient covariates (variables) to include in our IPD network meta-regression model. We began by rescaling all continuous candidate covariates by subtracting the mean and dividing with the standard deviation of each covariate. This was done to facilitate all analyses described below. For example, we rescaled observations regarding the continuous variable `IDS` (IDS anxiety factor at baseline) by subtracting 12 and dividing by 5. These numbers correspond to the mean and standard deviation across all studies. Next, we performed multiple imputations on the missing data (both outcome and variable data) using the `ice` command in Stata. We generated 10 multiply imputed datasets. As a first exploratory step we performed a stepwise variable selection on each of the imputed datasets, by (linearly) regressing the outcome at week 12 vs. all the covariates. We combined estimates of the imputed datasets with the Rubin's rules, using the `mim` command in Stata. We took this first step in order to do a first screening,

aiming to narrow down the number of covariates.

After having identified the most important covariates in the first step, we proceeded to the second step. In this step, in order to decide which of these covariates or combinations thereof to include in the regression model as PF or EM, we fitted a penalized regression model using the `glmnet` package in R. We explored first and second order combinations of the covariates, and their interactions with the treatment. We fitted the model separately in each multiply imputed dataset, and we kept the terms that were chosen by the penalized regression model in all multiply imputed datasets.

Penalized regression models offer a means for identifying the most important covariates, dropping the rest out of the model. In order to decide which terms to include in the model, we performed a 10-fold cross-validation (CV). The dataset was randomly partitioned into 10 equal sized subsamples. A single subsample was retained as the CV data for testing the model, and the remaining 9 subsamples were used as training data. The CV process was repeated 10 times, with each of the 10 subsamples used exactly once as the CV data. The 10 results from the folds were then combined to produce a single model.

5 Fitting the models

We fitted all IPD NMA and meta-regression models in OpenBUGS, using 2 independent chains per model. For all model parameters we used vague distributions. For the coefficients of PFs and EMs we used $N(0,10^2)$. For the imputation of dichotomous covariates we used a uniform $U(0,1)$ for the study-specific probabilities (i.e. used for imputing the binary covariates such as `Prior medication` and `Neglect`). For the rest of the continuous covariates as well as the IFs we used $N(0,10^2)$ for the means and $U(0,4)$ for the standard deviations. We performed 50,000 iterations, and we discarded the first 10,000 samples. Convergence was confirmed using the Brooks-Gelman-Rubin criterion (3). After fitting the model to the data and estimating the model parameters we used the model to make predictions about patients, given the values of the PFs and EMs. This is used in the interactive webpage and also the excel file.

REFERENCES

1. Higgins JPT, Jackson D, Barrett JK, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. ResSynthMeth. 2012;3(2):98–110.

2. White IR, Barrett JK, Jackson D, Higgins JPT. Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. ResSynthMeth. 2012;3(2):111–25.

3. Brooks SP, Gelman A. General Methods for Monitoring Convergence of Iterative Simulations. J Comput Graph Stat. 1998 Dec 1;7(4):434–55.

Supplement 2. Data description

Figure S1. PRISMA flowchart for selection of studies

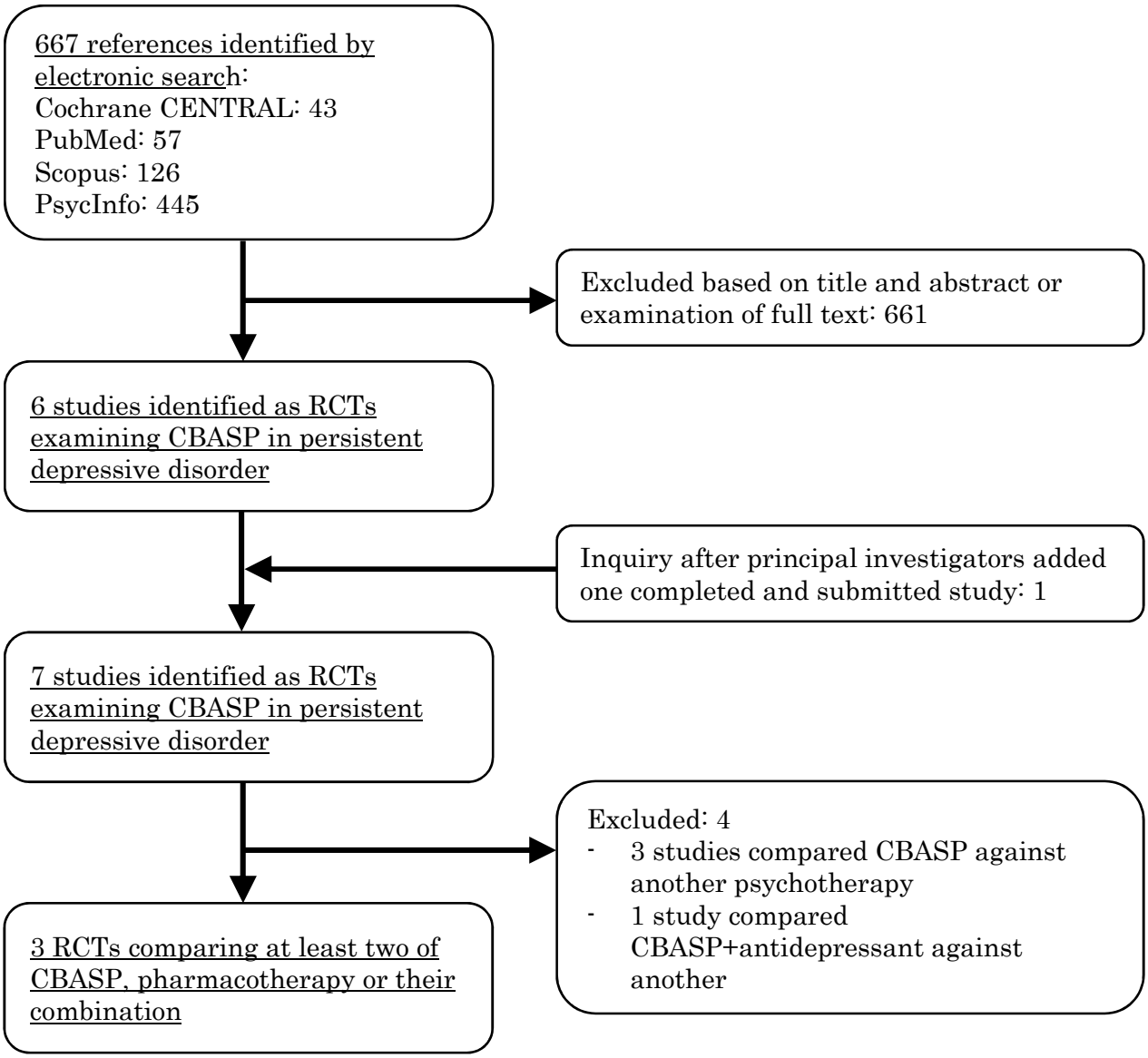
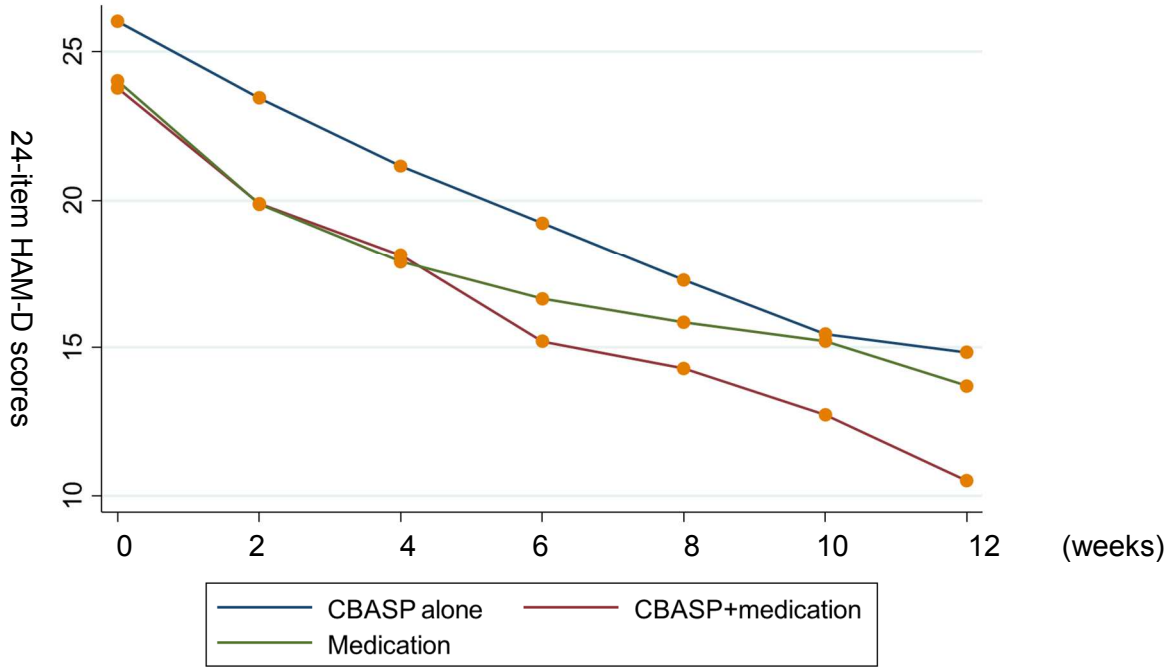


Table S1: Data availability for depression severity by week

	study		
Severity at week	**Keller**	**Kocsis**	**Schramm**
0	✓	✓	✓
2	✓	✓	-
4	✓	✓	-
6	✓	✓	-
8	✓	✓	✓
10	✓	✓	-
12	✓	✓	-

Figure S2. Pooled, aggregated data from all the three studies, for 24-item HAM-D at different time points



Supplement 3. Parameter estimates for prognostic factors and effect modifiers

Table S2. Selected prognostic factors (PFs) and effect modifiers (EFs) for change in depression severity, dropout for any reason and deterioration

Primary outcomes	**Prognostic factors**	**Effect modifiers**	
		CBASP vs COMB	**MEDS vs COMB**
Depression severity	1. IDS anxiety factor 2. HAM-D 3. Prior medication 4. $(\text{HAM-D})^2$ 5. Neglect × HAM-D 6. HAM-D × Prior medication 7. IDS anxiety factor × Prior medication	1. $(\text{HAM-D})^2$ 2. HAM-D × IDS anxiety factor	1. HAM-D 2. HAM-D × Prior medication
Dropout for any reason	1. HAM-D 2. Age 3. Prior medication 4. $(\text{HAM-D})^2$ 5. $(\text{Age})^2$ 6. Chronic MDD × HAM-D 7. Dysthymia × HAM-D 8. Age × Marital status single 9. Age × Chronic MDD 10. Marital status married ×Prior medication	1. $(\text{Age})^2$ 2. Age × Chronic MDD	1. $(\text{HAM-D})^2$ 2. Age × HAM-D
Secondary outcome			
Deterioration	1. HAM-D 2. IDS anxiety factor 3. $(\text{Social function})^2$ 4. $(\text{HAM-D})^2$ 5. $(\text{IDS anxiety factor})^2$ 6. IDS anxiety factor × HAM-D 7. Marital status married * HAM-D	1. $(\text{HAM-D})^2$	1. HAM-D

IDS anxiety factor: Anxiety/arousal factor score of Inventory of Depressive Symptomatology Self-Report at baseline (continuous)
HAM-D: 24-item Hamilton Rating Scale for Depression score at baseline (continuous)
Prior medication: Prior treatments with antidepressants (dichotomous)
Neglect: Emotional or physical neglect (dichotomous)
Chronic MDD: Chronic major depression (dichotomous)
Dysthymia: Dysthymic disorder (dichotomous)
Marital status married: married/de facto/in a relationship (dichotomous)

CBASP: Cognitive-Behavioral Analysis System of Psychotherapy
MEDS: Antidepressants
COMB: Cognitive-Behavioral Analysis System of Psychotherapy + Antidepressants

Table S3. Parameter estimates for the IPD network meta-regression model for depression severity

	Description of the parameter	mean	Lower 95% CrI	Upper 95% CrI
Regression coefficients for time	β_0	-0.23	-0.25	-0.20
Regression coefficients for prognostic factors	β_1 IDS anxiety factor	0.04	-0.07	0.15
	β_2 HAM-D	0.37	0.19	0.55
	β_3 Prior medication	0.06	-0.11	0.21
	β_4 $(HAM\text{-}D)^2$	-0.03	-0.07	0.02
	β_5 Neglect*HAM-D	0.05	-0.09	0.19
	β_6 Prior medication*HAM-D	0.11	-0.06	0.30
	β_7 IDS anxiety factor*Prior medication	0.06	-0.05	0.19
Regression coefficients for effect modification CBASP vs combination	$(HAM\text{-}D)^2$	-0.05	-0.14	0.04
	IDS anxiety factor* HAM-D	0.16	0.04	0.28
Regression coefficients for effect modification medication vs combination	HAM-D	0.04	-0.16	0.23
	Prior medication * HAM-D	0.08	-0.13	0.27

Where:

- Neglect is the emotional or physical neglect (binary variable)
- IDS anxiety factor is IDS anxiety/arousal factor at week 0 (continuous variable)
- HAM-D is 24-item HAM-D score at week 0 (continuous variable)
- Prior medication represents history of prior medication (binary variable)

All covariates have been standardized before the analyses, so that it is not straightforward to interpret the values of these covariates.

Table S4. Parameter estimates for the IPD network meta-regression model for dropouts for any reason

	Description of the parameter	mean	Lower 95% CrI	Upper 95% CrI
Regression coefficients for prognostic factors	β_1 HAM-D	0.11	-0.27	0.50
	β_2 Age	-0.16	-0.41	0.08
	β_3 Prior medication	-0.07	-0.44	0.30
	β_4 $(HAM\text{-}D)^2$	-0.09	-0.28	0.08
	β_5 $(Age)^2$	0.20	0.04	0.36
	β_6 Chronic MDD*HAM-D	0.42	0.01	0.84
	β_7 Dysthymia*HAM-D	0.43	0.03	0.84
	β_8 Age*Marital status married	-0.16	-0.54	0.21
	β_9 Age*diagnos1	-0.22	0.59	0.15
	β_{10} Marital status single*Prior medication	-0.38	-0.78	0.03
Regression coefficients for effect modification CBASP vs combination	$(Age)^2$	-0.39	-0.75	-0.07
	Age * Chronic MDD	-0.87	-1.64	-0.17
Regression coefficients for effect modification medication vs combination	$(HAM\text{-}D)^2$	-0.13	-0.38	0.12
	Age * HAM-D	0.09	-0.17	0.34

Where

- Chronic MDD is a binary variable which denotes whether the primary diagnosis depression type is MDD (chronic).
- Dysthymia is a binary variable which denotes whether the primary diagnosis depression type is Dysthymic Disorder.
- If Chronic MDD = Dysthymia = 0 the primary diagnosis depression type is "recurrent major dep with incomplete interepisode recovery"
- Marital status married is a binary variable which denotes whether marital status is married/defacto/in relationship
- Marital status single is a binary variable which denotes whether marital status is single
- If Marital status married = Marital status single = 0, marital status is widowed/divorced/separated

We removed Schramm (2015) from analyses for dropout because it measured the dropouts at 8 weeks, while all the other studies measured them at 12 weeks.

Table S5. Parameter estimates for the IPD network meta-regression model for deterioration

	Description of the parameter	mean	Lower 95% CrI	Upper 95% CrI
Regression coefficients for prognostic factors	β_1 HAM-D	-0.87	-1.33	-0.43
	β_2 IDS anxiety factor	0.18	-0.08	0.45
	β_3 (Social function)^2	0.06	-0.01	0.13
	β_4 (HAM-D)^2	-0.23	-0.54	0.07
	β_5 (IDS anxiety factor)^2	-0.15	-0.37	0.04
	β_6 IDS anxiety factor*HAM-D	0.32	-0.09	0.75
	β_7 Marital status married*HAM-D	-0.03	-0.43	0.36
Regression coefficients for effect modification CBASP vs combination	(HAM-D)^2	0.18	-0.56	0.93
Regression coefficients for effect modification medication vs combination	HAM-D	0.00	-0.46	0.46

Where:

- HAM-D is 24-item HAM-D score at week 0 (continuous variable)
- IDS anxiety factor is IDS anxiety/arousal factor at week 0 (continuous variable)
- Social function is Global Assessment of Functioning score at week 0 (continuous variable)
- Marital status married is a binary variable which denotes whether marital status is married/defacto/in relationship

We removed Schramm (2015) from analyses from deterioration because its last outcome was at 8 weeks, while all the other studies measured them at 12 weeks.

Odds ratios for deterioration and 95% CrI are as follows. Values larger than 1 indicate that the second treatment in the comparison is favoured, i.e. it corresponds to a smaller probability of deterioration.

Assumed baseline HAM-D score	CBASP vs COMB	CBASP vs MEDS	MEDS vs COMB
High (40)	3.30 (0.70, 16.06)	2.23 (0.52, 9.75)	1.48 (0.44, 5.14)
Moderate (30)	2.64 (1.23, 5.84)	1.78 (0.89, 3.60)	1.48 (0.72, 3.08)
Low (20)	2.11 (1.07, 4.22)	1.42 (0.74, 2.72)	1.48 (0.95, 2.32)

Regarding relative effects, both COMB and MEDS appear to cause less deterioration than CBASP alone, and COMB appears to be better than MEDS. However, there is large uncertainty, as all but two ORs were not statistically significant. For high HAM-D scores there are very few events of deterioration, and thus the estimates are very uncertain.

Comparing the regression coefficients across different baseline HAM-D scores, we can see that for MEDS vs COMB there is no effect modification. I.e. the relative effects for deterioration do not change with changes over the baseline HAM-D (the point estimate was always 1.48).
On the other hand, the comparison CBASP vs COMB and CBASP vs MEDS appears to show weak evidence that for larger values of HAM-D the benefit of COMB over CBASP and of MEDS over CBASP is larger. But again there is large uncertainty in the estimates.

Overall we conclude that there is no strong evidence of effect modification.

Supplement 4. Examination of inconsistency

For the first primary outcome (efficacy) we focused at 8 weeks. At this time point all studies provided information. Two IFs were included in the model. The corresponding estimates were 0.7 (95% CrI -2.0 to 3.3) and 1.5 (-2.9 to 5.9) in the HAMD scale.

For the second primary outcome (dropout) all analyses were performed at 12 weeks. Only one inconsistency factor was included in the model. The corresponding estimate was 0.12 (-0.69 to 0.90) in the logOR scale. Please refer to 3. Estimating inconsistency of Supplement 1 for the statistical models.