**Supplementary Table 1:** Detailed Characteristics of Included Studies

### Backs-Dermott 2010

|  |  |
| --- | --- |
| **Study details** | **Sponsorship source:** Canadian Institutes of Health Research  **Country:** Canada  **Setting:** Community setting  **Year of recruitment:** Not reported |
| **Methods** | **Type of study:** Model development study  **Source of data:** Prospective longitudinal cohort study  **Method used for model development:** Differential Function Analysis  **Method used for internal validation:** Not reported  **External validation:** Not done  **Handling of missing data:** Not reported  **Evaluation of clinical utility:** Not assessed |
| **Sample size** | **Total number of participants (Number with event):** 49 (29)  **Number of candidate predictor parameters:** 11  **Number of predictors in final model:** 5  **Number of events per candidate predictor parameter (EPP):** Not applicable |
| **Population** | **Inclusion criteria:**   * Female * Aged 18 - 65 * Diagnosis of DSM-IV-TR current Major Depressive Episode (MDE) or MDE within the past 8 week   **Exclusion criteria:**   * Ever experienced a manic or mixed episode * Meeting criteria for a psychotic disorder, or ever experienced 2 or more psychotic symptoms * Meeting criteria for depression with psychotic features * Meeting criteria for substance abuse disorder or dependence |
| **Baseline characteristics** | **Mean age (SD):** Relapse group: 43.1 (10.87); Stable remitted group: 43.65 (11.72)  **Gender (% Female):** 100 |
| **Start-point (diagnosis of depression and remission)** | **Depression:** Diagnosis of DSM-IV-TR current Major Depressive Episode (MDE) or MDE within the past 8 weeks  **Remission:** "per [Frank 1991](#REF-Frank-1991) criteria":  1) reported less than 2 symptoms of depression on the SCID-I for at least 2 weeks; and 2) scored ≤ 13 on the BDI-II |
| **End-point (diagnosis of relapse/recurrence)** | **Relapse** within 12 months: meeting current criteria for MDE according to SCID-I |
| **Timing (length of follow-up)** | 12 months |
| **Notes** |  |

### Berlanga 1999

|  |  |
| --- | --- |
| **Study details** | **Sponsorship source:** Not reported  **Country:** Mexico  **Setting:** Secondary care (outpatients)  **Year of recruitment:** 1994 - 1996 |
| **Methods** | **Type of study:** Model development study  **Source of data:** Post-RCT\* prospective follow-up study  **Method used for model development:** Logistic regression (multivariable analysis with a stepwise backward method in which variables that were significant in the univariable analysis were introduced into the model)  **Method used for internal validation:** Not reported  **External validation:** Not done  **Handling of missing data:** Not reported  **Evaluation of clinical utility:** Not assessed |
| **Sample size** | **Total number of participants (Number with event):** 42 (18)  **Number of candidate predictor parameters:** Not reported  **Number of predictors in final model:** 3  **Number of events per candidate predictor parameter (EPP):** Unclear |
| **Population** | **Inclusion criteria:**   * Between 18 and 65 years old * DSM-IV criteria for diagnosis of major depressive disorder * Scoring at least 18 points on the first 17 items of the 21-item version of the Hamilton Rating Scale for Depression (HAM-D)   **Exclusion criteria:**   * Psychotic symptoms * Substantial suicide risk * If any other situation required hospitalisation |
| **Baseline characteristics** | **Mean age (SD):** Recurrence group: 34.8 (11.1); No-recurrence group: 37.2 (11.2)  **Gender (% Female):** Recurrence group: 83; No-recurrence group: 71 |
| **Start-point (diagnosis of depression and remission)** | **Depression:** Major depressive disorder according to DSM-IV criteria and at least 18 points on the first 17 items of the 21-item HAM-D  **Remission:** Definition of remission not reported |
| **End-point (diagnosis of relapse/recurrence)** | **Recurrence:** Fulfilling criteria for MDD (clinical interview) per [Frank 1991](#REF-Frank-1991) |
| **Timing (length of follow-up)** | 12 months |
| **Notes** | \*The RCT compared the clinical efficacy and tolerance of the antidepressants nefazodone and fluoxetine. A 'washout period' of at least 3 weeks free of antidepressant medication was a requisite for all participants |

### Johansson 2015

|  |  |
| --- | --- |
| **Study details** | **Sponsorship source:** Not reported  **Country:** Sweden  **Setting:** Secondary care (psychiatric outpatients)  **Year of recruitment:** Not reported |
| **Methods** | **Type of study:** Model development study  **Source of data:** Prospective cohort study  **Method used for model development:** Logistic regression (the 2 predictor variables were chosen which showed the strongest independent correlations with relapse/recurrence)  **Method used for internal validation:** Not reported  **External validation:** Not done  **Handling of missing data:** Not reported  **Evaluation of clinical utility:** Not assessed |
| **Sample size** | **Total number of participants (Number with event):** 51 (31)  **Number of candidate predictor parameters:** 4 (based on univariable analysis)  **Number of predictors in final model:** 2  **Number of events per candidate predictor parameter (EPP):** 7.75 |
| **Population** | **Inclusion criteria:**   * Outpatients with a primary diagnosis of depressive episode or recurrent depressive disorder (ICD-10 criteria) * At least 18 years of age * In remission   **Exclusion criteria:**   * Psychotic features * Diagnosis of bipolar disorder * Received ECT for the index period |
| **Baseline characteristics** | **Mean age (SD):** 47 (SD = 17)  **Gender (% Female):** 71 |
| **Start-point (diagnosis of depression and remission)** | **Depression:** ICD-10 criteria for depressive episode or recurrent depressive disorder  **Remission:** determined by psychiatrist at discharge and confirmed by structured clinical interview   * Partial remission defined as not fulfilling the criteria of DSM-IV depressive episode but having more than minimal symptoms (i.e. Montgomery–Asberg depression rating scale—self rating scale (MADRS-S) score > 9) * Full remission is defined as not fulfilling the criteria of DSM-IV depressive episode and showing only minimal symptoms (i.e. MADRS-S < 10) |
| **End-point (diagnosis of relapse/recurrence)** | **Relapse/recurrence:** (per [Frank 1991](#REF-Frank-1991))   * Relapse defined as having a depressive episode within 2 months of discharge * Recurrence defined as having a depressive episode after a period of recovery (at least 2 months after discharge)   Relapse/recurrence and current depressive status established using the sections Mood Episodes and Mood Disorders from The Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I) |
| **Timing (length of follow-up)** | 12-14 months |
| **Notes** |  |

### Judd 2016

|  |  |
| --- | --- |
| **Study details** | **Sponsorship Source:** Not reported  **Country:** US  **Setting:** Secondary care (academic centres)  **Year of Recruitment:** 1978-1981 |
| **Methods** | **Type of study:** Model development study  **Source of data:** Prospective cohort study (the National Institute of Mental Health Collaborative Depression Study)  **Method used for model development:** Forward and backward selection of pre-selected predictors using stepwise mixed-model logistic regression  **Method used for internal validation:** Not reported  **External validation:** Not done  **Handling of missing data:** Multiple imputation  **Evaluation of clinical utility:** Not assessed |
| **Sample size** | **Total number of participants (Number with event):** 188 (58)\*  514 SCL-90 assessments (73 with relapse)  **Number of candidate predictor parameters:** 17  **Number of predictors in final model:** 12  **Number of events per candidate predictor parameter (EPP):** 4.29 (17 candidate predictors to 73 "relapses") |
| **Population** | **Inclusion criteria:**   * White * IQ > 70 * Speak English * Entered the National Institute of Mental Health Collaborative Depression Study in an active major depressive episode   **Exclusion criteria:**   * Lifetime bipolar disorder or schizophrenia |
| **Baseline characteristics** | **Mean age (SD):** 37.8 (14.4)  **Gender (% female):** 58.5 |
| **Start-point (diagnosis of depression and remission)** | **Depression:** Major depression, assessed by Research Diagnostic Criteria based on Schedule for Affective Disorders and Schizophrenia interviews (no lifetime bipolar disorder, schizoaffective disorder or schizophrenia)  **Remission:** Psychiatric status rating of 1 (asymptomatic, returned to usual self with no symptoms of the episode) for at least 8 weeks |
| **End-point (diagnosis of relapse/recurrence)** | **Relapse (within 6 months):** 2 consecutive weeks of psychiatric status ratings at threshold for defining episode of major or minor/dysthymic depression |
| **Timing (length of follow-up)** | 6 months |
| **Notes** | \*There were 514 SCL-90 assessments taken from 188 participants. 73 of these assessments (from 58 participants) were identified as having relapsed |

### Klein 2018

|  |  |
| --- | --- |
| **Study details** | **Sponsorship source:** Not reported  **Country:** The Netherlands  **Setting:** Primary care  **Year of recruitment:** Development data:2010 - 2013; Validation data: 2009 - April 2015 |
| **Methods** | **Type of study:** Model development study with external validation  **Source of data:** Prospective data from 2 pragmatic RCTs  **Method used for model development:** Cox proportional hazards regression (backward selection at P < 0.05)  **Method used for internal validation:** Bootstrapping; shrinkage determined for all statistics  **External validation:** Separate RCTs formed development and validation datasets  **Handling of missing data:** Multiple imputation  **Evaluation of clinical utility:** Not assessed |
| **Sample size** | **Total number of participants (Number with event):** Development dataset: 235 (104); Validation dataset: 205 (116)  **Number of candidate predictor parameters:** 8  **Number of predictors in final model:** 4  **Number of events per candidate predictor parameter (EPP):** 13 |
| **Population** | **Inclusion criteria:**   * Aged 18 to 65 years * Experienced at least 2 episodes of major depressive disorder (the last one within 2 years) * Remitted according to DSM-IV criteria and HRSD < 10   **Exclusion criteria:**   * Mania/hypomania * Psychotic or bipolar disorder (past or present) * Alcohol/drug abuse * Primary diagnosis of an anxiety disorder * Organic brain damage |
| **Baseline characteristics** | **Mean age (SD):** Development dataset: 46.8 (10.6); Validation dataset: 48.3 (9.9)  **Gender (% female):** Development dataset: 74.5; Validation dataset: 66.5 |
| **Start-point (diagnosis of depression and remission)** | **Depression:** DSM-IV criteria  **Remission:** Assessed using SCID-I and HRSD score ≤ 10 |
| **End-point (diagnosis of relapse/recurrence)** | **Recurrence (time to) within 2 years:** assessed using SCID-I |
| **Timing (length of follow-up)** | 2 years |
| **Notes** |  |

### Mocking 2021

|  |  |
| --- | --- |
| **Study details** | **Sponsorship source:** Not reported  **Country:** US  **Setting:** Community setting  **Year of recruitment:** 2011-2014 |
| **Methods** | **Type of study:** Model development study  **Source of data:** Cross-sectional study comparing people with remitted recurrent MDD (rrMDD) with never depressed controls (rrMDD population followed up for 2.5 years)  **Method used for model development:** Cox proportional hazards regression  **Method used for internal validation:** Repeated double cross validation (rdCV), with bootstrapping 100 times to test random subsamples of 2/3 in and 1/3 out, and by permutation analysis  **External validation:** Not done  **Handling of missing data:** Not reported  **Evaluation of clinical utility:** Not assessed |
| **Sample size** | **Total number of participants (Number with event):** 62 (35)  **Number of candidate predictor parameters:** 399 intracellular and plasma metabolites (number of parameters unclear)  **Number of predictors in final model:** Unclear  **Number of events per candidate predictor parameter (EPP):** Unclear |
| **Population** | **Inclusion criteria:**  ≥2 episodes of MDD according to DSM-IV  Stable remission – not meeting SCID criteria for MDD and HAM-D<8  Aged 35-65  **Exclusion criteria:**  Current diagnoses of alcohol and/or drug dependence, psychotic or bipolar symptoms, predominant anxiety or severe personality disorder. Also standard MRI-exclusion criteria, history of severe head trauma or neurological disease, or severe general physical illness. All participants had to be without psychoactive medication for ≥4weeks. |
| **Baseline characteristics** | **Mean age (SEM):** Males: 54 (1.4); Females: 53 (1.2)  **Gender (% female):** 66.1 |
| **Start-point (diagnosis of depression and remission)** | **Depression:** DSM-IV criteria  **Remission:** Assessed using SCID-I and HRSD score ≤ 10 |
| **End-point (diagnosis of relapse/recurrence)** | **Recurrence:** ≥5 depressive symptoms lasting at least 2 weeks according to the DSM-IV criteria (SCID). |
| **Timing (length of follow-up)** | 2.5 years |
| **Notes** |  |

### Pintor 2009

|  |  |
| --- | --- |
| **Study details** | **Sponsorship source:** Not reported  **Country:** Spain  **Setting:** Secondary care (outpatients)  **Year of recruitment:** 2001 - 2005 |
| **Methods** | **Type of study:** Model development  **Source of data:** Prospective cohort study  **Method used for model development:** Logistic regression  **Method used for internal validation:** Not reported  **External validation:** Not done  **Handling of missing data:** Not reported  **Evaluation of clinical utility:** Not assessed |
| **Sample size** | **Total number of participants (Number with event):** 43 (18)  **Number of candidate predictors:** Not reported  **Number of predictors in final model:** 3  **Number of events per candidate predictor parameter (EPP):** Unclear |
| **Population** | **Inclusion criteria:**   * Experienced a depressive episode according to DSM-IV (SCID) * Aged 30 - 65   **Exclusion criteria:**   * Alcohol or drug dependence * Current or history of severe psychiatric disorders except MDD * Severe physical health disorders * Body weight > 150% of ideal weight * Taking antiepileptics * Needle phobia * Pregnant |
| **Baseline characteristics** | **Mean age (SD):** Relapsed group: 50.67 (8.04); Non-relapsed group: 51.88 (8.54)  **Gender (% female):** Relapsed group: 50; Non-relapsed group: 56 |
| **Start-point (diagnosis of depression and remission)** | **Depression:** SCID-IV diagnosis for unipolar major depressive episode (first or recurrent)  **Remission:** identified using Hamilton Depression Rating Scale (HDRS-21); “[Frank 1991](#REF-Frank-1991) criteria were applied” (does not describe exactly how) |
| **End-point (diagnosis of relapse/recurrence)** | **Presence versus absence of relapse over 2-year follow-up:** identified using Hamilton Depression Rating Scale (HDRS-21); “[Frank 1991](#REF-Frank-1991) criteria were applied” (does not describe exactly how) |
| **Timing (length of follow-up)** | 2 years |
| **Notes** |  |

### Ruhe 2019

|  |  |
| --- | --- |
| **Study details** | **Sponsorship source:** Not reported  **Country:** The Netherlands  **Setting:** Primary care  **Year of recruitment:** Not reported |
| **Methods** | **Type of study:** Model development study  **Source of data:** Prospective cohort study  **Method used for model development:** Machine learning support vector machine (SVM); data-driven model (classification-based algorithm)  **Method used for internal validation:** "Leave-one-out" validation procedure  **External validation:** Not done  **Handling of missing data:** Mean imputation  **Evaluation of clinical utility:** Not assessed |
| **Sample size** | **Total number of participants (Number with event):** 64 (35)  **Number of candidate predictors:** Not reported  **Number of predictors in final model:** 4  **Number of events per candidate predictor parameter (EPP):** Unclear |
| **Population** | **Inclusion criteria:**   * Voluntarily free of anti-depressants for past weeks * Between 35 and 65 years old * 2 or more previous episodes of MDD   **Exclusion criteria:**   * Alcohol or drug dependence * Primary anxiety disorder * Psychotic or bipolar disorder * Received ECT within 2 months of assessment * History of head trauma, neurological disease or severe physical illness |
| **Baseline characteristics** | **Mean age (SD):** 53.4 (7.7)  **Gender (% female):** 65.8 |
| **Start-point (diagnosis of depression and remission)** | **Depression:** Recurrent MDD: 2 or more MDD episodes according to the SCID-I  **Remission:** ≤ 7 on the HDRS) for ≥ 8 weeks and not fulfilling the criteria for a current MDD episode (SCID-I) |
| **End-point (diagnosis of relapse/recurrence)** | **Recurrence:** MDD according to SCID-I. |
| **Timing (length of follow-up)** | **Median follow up:** 233 days (IQR 92 - 461) |
| **Notes** |  |

### Van Loo 2015

|  |  |
| --- | --- |
| **Study details** | **Sponsorship source:** Not reported  **Country:** USA  **Setting:** Community setting  **Year of recruitment:** 1988 - 1997 |
| **Methods** | **Type of study:** Model development study with external validation  **Source of data:** Prospective longitudinal data\*  **Method used for model development:** Elastic net penalised Cox proportional hazards regression  **Method used for internal validation:** 10-fold cross-validation and shrinkage of beta-coefficients  **External validation:** Temporal validation  **Handling of missing data:** Single imputation  **Evaluation of clinical utility:** Not assessed |
| **Sample size** | **Total number of participants (Number with event):** Development dataset: 194 (45); Validation dataset: 133 (57)  **Number of candidate predictor parameters:** 81 candidate predictors (number of parameters unclear)  **Number of predictors in final model:** 26  **Number of events per candidate predictor parameter (EPP):** Unclear |
| **Population** | **Inclusion criteria:**   * Female twins * DSMIII MD episode in the previous year   **Exclusion criteria:**   * Not listed. |
| **Baseline characteristics** | **Mean age (SD):** Development dataset: 30.7 (7.1); Validation dataset: 32.4 (7.1)  **Gender (% female):** 100 |
| **Start-point (diagnosis of depression and remission)** | **Depression:** DSM-III MD episode in previous year (self-report and confirmed by SCID)  **Remission:** No longer meeting criteria according to SCID |
| **End-point (diagnosis of relapse/recurrence)** | **Recurrence:** first episode meeting DSM-III-R criteria after a period of not meeting the criteria (remission or recovery) for at least 4 months  **Time to recurrence:** Number of months between initial interview and recurrence |
| **Timing (length of follow-up)** | Development dataset: median follow-up 5.5 years; Validation dataset: median follow-up 6.1 years |
| **Notes** | \*Data from prospective longitudinal studies of Caucasian female-female twin pairs (Virginia Adult Twin Study of Psychiatric and Substance Use Disorder) |

### Van Loo 2018

|  |  |
| --- | --- |
| **Study details** | **Sponsorship source:** Not reported  **Country:** USA  **Setting:** Community setting  **Year of recruitment:** 1988 - 1997 |
| **Methods** | **Type of study:** Model development study  **Source of data:** Longitudinal cohort study\*  **Method used for model development:** Cox proportional hazards model with elastic net penalised regression analysis  **Method used for internal validation:** Random split "test" sample.The final model was selected based on minimal prediction error as assessed in 10-fold cross-validation  **External validation:** Not done  **Handling of missing data:** Multiple imputation by chained equations  **Evaluation of clinical utility:** Not reported |
| **Sample size** | **Total number of participants (Number with event): Total sample (men and women):** 653\*\*  **Number of candidate predictor parameters:** 70 predictors (number of parameters unclear)  **Number of predictors in final model:** 24  **Number of events per candidate predictor parameter (EPP):** Unclear |
| **Population** | **Inclusion criteria:**   * Episode of MD in year prior to baseline interview   **Exclusion criteria:**   * No MD episode in year prior to baseline interview * Those who reported an interval of 60 days or less between the offset of their last MD episode at baseline and their first depressive episode during the follow-up |
| **Baseline characteristics** | **Mean age (SD):** 35 (8.8)  **Gender (% female):** 34.6 |
| **Start-point (diagnosis of depression and remission)** | **Depression:** A diagnosis of MD in the year prior to baseline interview was based on the DSM-III-R criteria as assessed by the Structured Clinical Interview for DSM-III-R  **Remission:** All participants reported a period of > 60 days of (partial) remission or recovery |
| **End-point (diagnosis of relapse/recurrence)** | **Recurrence:** First reported episode meeting DSM-III-R criteria in the year prior to follow-up interview  **Time to recurrence:** Time at risk for recurrence (follow-up) was defined as the interval between the offset of MD in the year prior to baseline interview and the onset of MD in the year prior to follow-up interview |
| **Timing (length of follow-up)** | 5 years |
| **Notes** | \*Virginia Adult Twin Study of Psychiatric and Substance Use Disorders (VATSPSUD), a population-based longitudinal study of male–male and male–female Caucasian twin pairs  \*\*This was the full sample size, including men and women. There were also separate analyses in women (n = 226) and in men (n = 427). The male cohort was split further into a training sample (n = 277) and a test sample (for external validation) |

### Van Loo 2020

|  |  |
| --- | --- |
| **Study details** | **Sponsorship source:** Funding for NESDA reported in paper  **Country:** Netherlands (NESDA); USA (VATSPSUD)  **Setting:** Primary care, secondary care and community setting (NESDA); Community setting (VATSPSUD)  **Year of recruitment:** 2004 - 2007 (NESDA); 1988-1997 (VATSPSUD) |
| **Methods** | **Type of study:** External validation study using NESDA data (internal validation also performed on VATSPSUD data)  **Source of data:** 2 longitudinal cohort studies\*  **Method used for model development:** Not applicable  **Method used for internal validation:** Random split sample of VATSPSUD data used in [Van Loo 2018](#STD-Van-Loo-2018)\*  **External validation:** Logistic regression using NESDA dataset\*\*  **Handling of missing data:** Multiple imputation by chained equations  **Evaluation of clinical utility:** Not done |
| **Sample size** | **Total number of participants (Number with event):** NESDA Test sample (n = 1925); VATSPSUD Test sample (n = 2301). Number with event not clear  **Number of candidate predictor parameters:** Not applicable  **Number of predictors in final model:** 24  **Number of events per candidate predictor parameter (EPP):** Not applicable |
| **Population** | **For external validation (NESDA):**  **Inclusion criteria:**   * Dutch general population, primary care, and specialised mental health care, aged 18 – 65 at baseline assessment   **Exclusion criteria:**   * No MD episode in year prior to baseline interview. * Those who reported an interval ⩽ 60 days between the offset of their last MD episode at baseline and their first depressive episode during the follow-up   **For internal validation (VATSPSUD):**  Female-female twins (n = 757) and male-male/male-female twins (n = 1544) from the VATSPSUD study (only those not included in the original training sample used to develop the prediction model in [Van Loo 2018](#STD-Van-Loo-2018)) |
| **Baseline characteristics** | **Mean age (SD):** NESDA Test sample: 42 (12.4); VATSPSUD Test sample: 34.9 (8.6)  **Gender (% female):** NESDA Test sample: 68.6; VATSPSUD Test sample: 53.2 |
| **Start-point (diagnosis of depression and remission)** | **Depression:**Lifetime episode of MD at baseline  **Remission:**Not described |
| **End-point (diagnosis of relapse/recurrence)** | **Recurrence:** Any episode of MD during follow-up  **Time to recurrence:** Follow-up to 9 years |
| **Timing (length of follow-up)** |  |
| **Notes** | \*Two independent test samples from Virginia Adult Twin Study of Psychiatric and Substance Use Disorders (VATSPSUD) and the Netherlands Study of Depression and Anxiety (NESDA)  \*\*External validation performed on NESDA cohort |

### Wang 2014

|  |  |
| --- | --- |
| **Study details** | **Sponsorship source:** Not reported  **Country:** USA  **Setting:** Community setting  **Year of recruitment:** 2001 - 2005 |
| **Methods** | **Type of study:** Model development study with external validation  **Source of data:** Prospective longitudinal dataset\*  **Method used for model development:** Logistic regression with combined forward and backward selection (compared C-statistic with and without each predictor, then used Net Reclassification Improvement to examine if the predictor could correctly reclassify participants into appropriate categories)  **Method used for internal validation:** Application of heuristic shrinkage factor  **External validation:** Geographical validation  **Handling of missing data:** Single imputation  **Evaluation of clinical utility:** Not assessed |
| **Sample size** | **Total number of participants (number with event):** Development dataset: 1518 (362); Validation dataset: 1195 (307)  **Number of candidate predictor parameters:** Not reported  **Number of predictors in final model:** 24  **Number of events per candidate predictor parameter (EPP):** Unclear |
| **Population** | **Inclusion criteria:**   * Current or lifetime MDE * Remitted from MDE for at least 2 months * Went to health professionals (councillors and/or medical doctors) for help to improve mood, were hospitalised for depression, or went to emergency room because of depression   **Exclusion criteria:**   * Lifetime manic or hypomanic episodes |
| **Baseline characteristics** | **Mean age (SEM):** Development dataset: 45.38 (0.37); Validation dataset: 45.37 (0.41)  **Gender (% Female):** Development dataset: 77.4%; Validation dataset: 74.9% |
| **Start-point (diagnosis of depression and remission)** | **Depression:** DSM-IV  **Remission:** “Having remitted from recent depressive episode for at least 2 months” |
| **End-point (diagnosis of relapse/recurrence)** | **Recurrence, within 3 years:** Meeting DSM-IV diagnostic criteria for MDE |
| **Timing (length of follow-up)** | 3 years |
| **Notes** | \*Data from the US National Epidemiological Survey on Alcohol and Related Conditions |

Supplementary Table 2 Summary of final predictors and predictive performance of prognostic models

|  | | Predictive performance | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| Internal validation | | External validation | |  |
| Study | Predictors included in final model | Calibration | Discrimination | Calibration | Discrimination | Other performance statistics presented |
| Backs-Dermott 2010 | 'Psychosocial’ predictors: Life stress; Cognitive-Personality Vulnerability Factors; Social support; and Coping style:   * Interpersonal marked difficulties (Short Life Events and Difficulties Scale, SLEDS); * Perceived social support from a significant other (Multidimensional Scale of Perceived Social Support, MSPSS) * Perceived social support from friends (MSPSS) * Emotion-oriented coping (Coping Inventory for Stressful Situations, CISS); * Avoidance-oriented coping (CISS) | Not reported | Not reported | Not applicable | Not applicable | The DFA was significant:  Wilk's Lambda = 0.69, *x*2 (5) = 16.35, *P* = 0.006  Standardised discriminant function coefficients:   * MSPSS (Significant Other): 0.48; * MSPSS (Friends): 0.35; * CISS (Emotion-Oriented Coping): 0.67; * CISS (Avoidance-Oriented Coping): −0.58; * Presence of interpersonal severe difficulties: −0.63 |
| Berlanga 1999 | 'Personality and clinical predictors':   * Elevated EPQ (Eysenck Personality Questionnaire) score on the neuroticism subscale * Short duration of treatment of the index episode * A slow onset of response to treatment of the index episode | Not reported | Not reported | Not applicable | Not applicable | Combination of 3 variables predicted recurrence of depression in 90% of cases.  Threshold not specified  Sensitivity: 89%  Specificity: 92%  Positive Predictive Value: 89%  Negative Predictive Value: 92% |
| Johansson 2015 | * Number of previous episodes (0/1/2/3 or more) * Having a partner (yes/no) | Not reported | Not reported | Not applicable | Not applicable | Sensitivity: 90%  Specificity: 60%  Overall accuracy: 78%  (Threshold not defined)  Measure of overall model fit: Nagelkerke’s R2= 0.45  R2= 2.97 (Hosmer and Lemeshow), 0.33 (Cox and Snell)  Model Х2 = 20.66 (df = 2, *P* < 0.001) (compared with constant-only model)  Final model presented with regression coefficients and intercept:   * Intercept = −0.68 * Partner Beta coefficient = −2.14 (0.02 to 0.64) *P* = 0.01 * Previous episodes Beta coefficient = 1.19 (1.55 to 7.06) *P* = 0.00 |
| Judd 2016 | 12 SCL-90 items in final model:   * Feeling blocked in getting things done * Feeling pushed to get things done * Feeling tense or keyed up * Having ideas/beliefs others do not share * Feeling inferior to others * Feeling low in energy or slowed down * Feeling very self-conscious with others * Headaches * Crying easily * Feelings being easily hurt * Worrying too much about things * Trouble concentrating | Not reported | Not reported | Not applicable | Not applicable | Predictive statistics for ‘experiencing any one or more of the 12 symptoms most predictive of relapse at a moderate or worse level of severity for the past week’:  Sensitivity: 80.8%  Specificity: 51.2%  Positive Predictive Value: 21.5%; Negative Predictive Value: 94.2% |
| Klein 2018 | * Number of previous MDEs (life-chart of SCID-I), categorised as less than 3, 3 or 4, and 5 or more; * Number of residual depressive symptoms (Inventory of Depressive Symptomatology, continuous) * Severity of the last MDE (SCID-I), mild or moderate *v*. severe * Treatment in RCT also included as a non-significant predictor | Calibration slope = 0.81 | Harrell’s C-statistic = 0.56 | Calibration slope = 0.56 | Harrell’s C-statistic = 0.59 | Total risk score calculated from final model ‘scores’: low (< 35), moderate (35–50), high (> 50)  Cut-off score 35 or more (37% risk of recurrence):  Sensitivity: 52%  Specificity: 69%  PPV: 59%  NPV: 63%  Cut-off score 50 or more (71% risk of recurrence):  Sensitivity: 16%  Specificity: 95%  PPV: 72%  NPV: 57% |
| Mocking 2021 | Predictors were all metabolites (peripheral blood metabolomics) known to be core features of the cell danger and integrated stress response (CDR and ISR) pathways.  80% of the metabolic predictors of recurrence in both males and females belonged to 6 pathways: (1) phospholipids, (2) sphingomyelins, (3) glycosphingolipids, (4) eicosanoids, (5) microbiome, and (6) purines. | Not reported | Females: C-statistic = 0.90 (95% CI 0.69 to 1.0)  Males: C-statistic = 0.99 (95% CI 0.90 to 1.0) | Not applicable | Not applicable |  |
| Pintor 2009 | * Corticotrophin-releasing factor test (net area under cortisol curve (NAUCC), cut-off point of 251.24 μg/ml/min) * Previous suicide attempt * Stress during follow-up | Not reported | Not reported | Not applicable | Not applicable | Nagelkerke’s R2= 0.797  Sensitivity: 89%  Specificity: 92%  Hosmer-Lemeshow Goodness-of-fit test: χ2= 2.23, df = 8, *P* = 0.97 |
| Ruhe 2019 | Best classifier included 4 predictors:   * Number of previous episodes in last 10 years * Age of onset * CTQ-physical abuse subscale-score * CTQ-physical abuse of 8 or more | Not reported | Not reported | Not applicable | Not applicable | Results for ‘best classifier’:  Sensitivity: 71.4  Specificity: 79.3 |
| Van Loo 2015 | Recent depressive episode:   * Loss of interest (HR 1.10) * Appetite loss (HR 1.02) * Weight loss (HR 1.05) * Weight gain (HR 0.99) * Insomnia (HR 1.07) * Concentration difficulties (HR 1.07) * Feeling anxious, nervous, worried (HR 1.03) * Feeling tense, jumpy, shaky (HR = 1.06); * Sum of 9MD criteria (HR 1.02)   Current state:   * SCL past 30 days (HR 1.03)   Psychiatric history (lifetime):   * Age at first depression (HR 1.06) * Number of MD episodes ≥ 6 (HR 1.05) * Duration of most severe MD episode 1–3 months (HR 0.98) * Duration of most severe MD episode ≥ 3 months (HR 1.03) * Early anxiety (HR 1.06)   Family history:   * GAD co-twin (HR 1.06)   Personality:   * Extraversion (HR 1.02)   Adverse life events (early):   * Parental loss childhood/adolescence (HR 1.03) * Disturbed family environment (HR 1.02) * Sum of lifetime traumas 3–4 (HR 1.06) * Childhood sexual abuse (severe) (HR 1.04)   Adverse life events (recent):   * Number of stressful life events in past year (HR 1.03)   Social and economic environment:   * Marital status (HR 1.03) * Low marital satisfaction (HR 1.04) * Problems with relatives (HR 1.02) * Financial problems (HR 1.15) | Not reported. | AUC = 0.79 | Not reported. | AUC = 0.61 | Comparable KM-curves for the 2 lowest risk groups was used as evidence that the model is well-calibrated for those at lower risk but less so for higher-risk patients |
| Van Loo 2018 | Recent depressive episode:   * Loss of interest (HR 1.11) * Appetite gain (HR 1.01) * Weight loss (HR 1.02) * Feeling restless (HR 1.02) * Fatigue (HR 1.04) * Hypersomnia (HR 1.04) * Feeling irritable/angry (HR 1.06) * Feeling tense (HR 1.04) * Cardio-respiratory panic symptoms (HR 1.11) * Sum of 9 MD criteria (HR 1.05)   Current state:   * SCL last 30 days (HR 1.06)   Psychiatric history (lifetime):   * Early anxiety (HR 1.15) * History of GAD (HR 1.76) * 2–3 MD episodes lifetime (HR 1.02) * ⩾ 6 MD episodes lifetime (HR 1.14) * History of alcohol dependence (HR 1.03)   Family history:   * MD mother (HR 1.09)   Early adverse life events:   * Childhood sexual abuse (HR 1.19) * Traumas ⩾ 5 (HR 1.13)   Recent adverse life events:   * Number of stressful life events in past year (HR 1.01)   Social and economic environment:   * No partner (HR 1.03) * Low marital satisfaction (HR 1.13) * Support from relatives (HR 0.99) * Problems with relatives (HR 1.03) | Not reported | AUC in the male training sample = 0.785  AUC in male test sample = 0.710 | Not applicable | Not applicable | KM-curves for the low-risk group in both training and test data were very similar, indicating good discrimination and calibration for participants with lower risk for depression. The KM-curves for the intermediate and high-risk groups were more similar in the test data than in the training data, which indicated that the model was less well- calibrated for higher risk patients |
| Van Loo 2020 | As for Van Loo 2018 | Not reported | Predicting MD over 0–1 year:  AUC = 0.73 (95% CI 0.69 to 0.76)a | Not reported | Predicting MD over 0–2 years:  AUC = 0.68 (95% CI: 0.66 to 0.71)  Predicting MD over 0 -9 years:  AUC = 0.72 (95% C: 0.69 to 0.75) | – |
| Wang 2014 | * Female sex * Age (continuous); * Married/common-law * Divorced/separated/single * White * Had MDD last year * 2 depressive episodes * 3 + depressive episodes * Lifetime GAD or specific phobia * Avoidant personality disorder   Depressive symptoms in MDE:   * Difficulties in concentration * Wanted to eat more * Felt guilty * Took medication for low mood * SF-12 physical disability scores (53.9 to 57.8; 43.3 to 53.8; 0 to 43.2) * SF-12 mental disability scores (48.4 to 54.5; 37.7 to 48.3; 0 to 37.6) * Experience of racial discrimination * Ever physically attacked/beaten/injured); by spouse, partner, or anyone else (abuse) (Experience of sexual assault) * Before 18, parents/caregiver swear, insult, or say hurtful things to you (Almost never/sometimes; fairly often/very often) * Before 10 being left alone/unsupervised by parents/care givers (Almost never/sometimes; fairly often/very often)   Interaction terms:   * Sex × SF-physical * Marital × Abuse * Race × Avoid * SF-physical × Guilty | Not reported | C statistic = 0.75 | Not reported | C statistic = 0.7195 | Model development:  Hosmer-Lemeshow χ2 (8) = 10.48, *P* = 0.23  ‘Excellent calibration’  External validation:  Hosmer-Lemeshow χ2 (8) = 3.51, *P* = 0.90  ‘Excellent calibration’  In the combined development and validation data:  C statistic of 0.7365 and ‘excellent calibration’ (H–L χ2 (8) = 6.22, *P* = 0.62)  Observed risk of recurrence over 3 years = 25.40% (95% CI 23.76% to 27.04%)  Mean predicted risk of recurrence based on the model = 25.34% (95% CI 24.73% to 25.95%).  ‘We visually compared the predicted *v*. the observed risk of recurrence by decile risk groups’ |

a. This internal validation used the same data as development data (Van Loo 2018).

**Supplementary Table S3** Detailed Risk of bias and applicability assessment

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Study** | | | | | | | | | | | | | | | | |
| **Backs-Dermott 2010** | **Berlanga 1999** | **Johansson 2015** | **Judd 2016** | **Klein 2018** | | | **Mocking 2021** | **Pintor 2009** | | **Ruhe 2019** | **Van Loo 2015** | | **Van Loo 2018** | **Van Loo 2020** | **Wang 2014** | |
| **Type of study** | | **Dev** | **Dev** | **Dev** | **Dev** | **Dev** | **Val** | | **Dev** | **Dev** | | **Dev** | **Dev** | **Val** | **Dev** | **Val** | **Dev** | **Val** |
|  | **Domain 1: Participants** | | | | | | | | | | | | | | | | | |
|  | **A. Risk of bias** | | | | | | | | | | | | | | | | | |
| 1.1. Appropriate data sources? | | Yes | Yes | Yes | Yes | Yes | Yes | Yes | | | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| 1.2. Appropriate inclusions and exclusion? | | Yes | Yes | Yes | Yes | Yes | Yes | Yes | | | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| **Risk of bias** | | **Low** | **Low** | **Low** | **Low** | **Low** | **Low** | **Low** | | | **Low** | **Low** | **Low** | **Low** | **Low** | **Low** | **Low** | **Low** |
|  | **B. Applicability** | | | | | | | | | | | | | | | | | |
| **Concern about applicability** | | **Low** | **Unclear** | **Low** | **Low** | **Low** | **Low** | **Low** | | | **Low** | **Low** | **Low** | **Low** | **Low** | **Low** | **Low** | **Low** |
|  | **Domain 2: Predictors** | | | | | | | | | | | | | | | | | |
|  | **A. Risk of bias** | | | | | | | | | | | | | | | | | |
| 2.1. Defined and assessed in similar way for all participants? | | Yes | Yes | Yes | Yes | Yes | Yes | Probably yes | | | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| 2.2. Assessments made without knowledge of outcome? | | Probably yes | Probably yes | Yes | Yes | Yes | Yes | No information | | | No information | Yes | Probably yes | Probably yes | No information | No information | Probably yes | Probably yes |
| 2.3. All available at time of model’s intended use? | | Yes | Yes | Yes | Yes | Yes | Yes | Yes | | | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes |
| **Risk of bias** | | **Low** | **Low** | **Low** | **Low** | **Low** | **Low** | **Unclear** | | | **Unclear** | **Low** | **Low** | **Low** | **Unclear** | **High** | **Low** | **Low** |
|  | **B. Applicability** | | | | | | | | | | | | | | | | | |
| **Concern about applicability** | | **Low** | **Low** | **Low** | **Low** | **Low** | **Low** | **Low** | | | **Low** | **Low** | **Low** | **Low** | **Low** | **Low** | **Low** | **Low** |
|  | **Domain 3: Outcome** | | | | | | | | | | | | | | | | | |
|  | **A. Risk of bias** | | | | | | | | | | | | | | | | | |
| 3.1. Determined appropriately? | | Yes | Yes | Yes | Probably yes | Yes | Yes | Probably yes | | | No information | Yes | Yes | Yes | Yes | Yes | Probably yes | Probably yes |
| 3.2. Pre-specified or standard definition? | | Yes | Yes | Yes | Yes | Yes | Yes | Yes | | | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| 3.3. Predictors excluded from outcome definition? | | Yes | Yes | Yes | Yes | Yes | Yes | Yes | | | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| 3.4. Defined and determined similar for all participants? | | Yes | Yes | Yes | Yes | Yes | Yes | Yes | | | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| 3.5. Determined without knowledge of predictors? | | No information | No information | No information | Probably yes | Yes | Yes | Probably yes | | | No information | No information | No information | No information | No information | No information | No information | No information |
| 3.6. Appropriate time interval between predictor assessment and outcome determination? | | Yes | Yes | Yes | Yes | Yes | Yes | Yes | | | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| **Risk of bias** | | **Unclear** | **Unclear** | **Unclear** | **Low** | **Low** | **Low** | **Low** | | | **Unclear** | **Unclear** | **Unclear** | **Unclear** | **Unclear** | **Unclear** | **Unclear** | **Unclear** |
|  | **B. Applicability** | | | | | | | | | | | | | | | | | |
| **Concern about applicability** | | **Low** | **Low** | **Low** | **Low** | **Low** | **Low** | **Low** | | | **Low** | **Low** | **Low** | **Low** | **Low** | **Low** | **Low** | **Low** |
|  | **Domain 4: Analysis** | | | | | | | | | | | | | | | | | |
| 4.1. Reasonable number of participants with outcome? | | Probably yes | No | No | No | Probably yes | Yes | No | | | No | No | No | No | No | Probably yes | No information | Yes |
| 4.2. Predictors handled appropriately? | | Yes | Probably yes | Yes | No | Probably yes | Probably yes | No information | | | No | Probably no | No | No | Probably yes | Probably yes | No | No |
| 4.3. All enrolled participants included in analysis? | | No | No | Yes | No | Yes | Yes | No | | | Yes | No | Probably yes | Probably yes | No | Yes | Yes | Yes |
| 4.4. Missing data handled appropriately? | | No information | No information | No information | Yes | Yes | Yes | No information | | | No information | No | No | No | Yes | Yes | Probably no | Probably no |
| 4.5. Univariable analysis avoided? | | No | No | No | No | Yes | NA | Probably yes | | | No | Yes | Yes | NA | Yes | NA | No | NA |
| 4.6. Complexities in data accounted for? | | Probably yes | Probably yes | Probably yes | Yes | Yes | Yes | No information | | | Probably yes | Yes | Probably yes | Probably yes | Yes | Probably yes | Probably yes | Probably yes |
| 4.7. Relevant performance measures? | | No | No | No | No | Yes | Yes | No | | | No | No | No | No | No | No | No | No |
| 4.8. Overfitting and optimism accounted for? | | No | No | No | No | Yes | NA | No information | | | No | No | Yes | NA | Yes | NA | Yes | NA |
| 4.9. Final model corresponds to multivariable analysis? | | No information | No information | Probably yes | No information | Yes | NA | No information | | | No information | No information | Probably no | NA | Probably yes | NA | No information | NA |
| **Risk of bias** | | **High** | **High** | **High** | **High** | **Low** | **Low** | **High** | | | **High** | **High** | **High** | **High** | **High** | **High** | **High** | **High** |
| **Overall assessment of risk of bias** | | **High** | **High** | **High** | **High** | **Low** | **Low** | **High** | | | **High** | **High** | **High** | **High** | **High** | **High** | **High** | **High** |
| **Overall concern for applicability** | | **Low** | **Unclear** | **Low** | **Low** | **Low** | **Low** | **Low** | | | **Low** | **Low** | **Low** | **Low** | **Low** | **Low** | **Low** | **Low** |