

1 **The development and validation of a prognostic model to predict relapse in**
2 **adults with remitted depression in primary care: secondary analysis of pooled**
3 **individual participant data from multiple studies**
4

5 Andrew S Moriarty^{*1 2 3}, Lewis W Paton^{1 2}, Kym IE Snell^{4,5}, Lucinda Archer^{4,5},
6 Richard D Riley^{4,5}, Joshua EJ Buckman^{6 7}, Carolyn A Chew-Graham³, Simon
7 Gilbody^{1 2}, Shehzad Ali^{2 8}, Stephen Pilling⁶, Nick Meader⁹, Bob Phillips^{1 10}, Peter A
8 Coventry², Jaime Delgadoillo¹¹, David A Richards^{12 13}, Chris Salisbury¹⁴, Dean
9 McMillan^{1 2}

10

- 11 1. Hull York Medical School, University of York
12 2. Department of Health Sciences, University of York
13 3. School of Medicine, Keele University, Keele, UK
14 4. Institute of Applied Health Research, University of Birmingham, UK
15 5. National Institute for Health and Care Research (NIHR) Birmingham Biomedical
16 Research Centre, UK
17 6. CORE Data Lab, Research Department of Clinical, Educational & Health Psychology,
18 University College London, London, UK.
19 7. iCope - Camden & Islington NHS Talking Therapies for anxiety and depression
20 services - Camden & Islington NHS Foundation Trust, London, UK
21 8. Department of Epidemiology and Biostatistics, Schulich School of Medicine &
22 Dentistry, Western University, London, ON, Canada
23 9. Newcastle University, UK
24 10. Centre for Reviews and Dissemination, University of York, UK
25 11. Department of Psychology, University of Sheffield, UK
26 12. Faculty of Health and Life Sciences, University of Exeter, UK
27 13. Department of Health and Caring Sciences, Western Norway University of Applied
28 Sciences, Inndalsveien 28, 5063 Bergen, Norway
29 14. Centre for Academic Primary Care, University of Bristol, UK

30 **What is already known on this topic**

31 Relapse contributes to the morbidity and burden associated with depression and,
32 while there is robust research confirming predictors of relapse, individualised risk
33 prediction is a challenge.

34

35 **What this study adds**

36 We found that it is not possible to accurately predict individualised risk of relapse
37 using prognostic factors that are routinely collected and available in primary care.
38 We found evidence to suggest that relationship status (not being in a relationship) is
39 associated with increased risk of relapse and warrants confirmatory prognostic factor
40 research.

41

42 **How this study might affect research, practice or policy**

43 Future prognosis research in this area should focus on exploring the feasibility of
44 routinely measuring and documenting additional prognostic factors in primary care
45 (for example, adverse childhood events, relationship status and social support) and
46 including these in prognostic models. Until we can more accurately identify
47 individuals at increased risk of relapse, commonly used acute-phase treatments
48 could be optimised to better prepare for and mitigate the risk of relapse and there is
49 a need for brief, scalable relapse prevention interventions that could be provided
50 more widely.

51

52 **Key words:** Relapse; Recurrence; Depression; Prognosis

53 **ABSTRACT**

54

55 **Background**

56 Relapse of depression is common and contributes to the overall associated morbidity
57 and burden. We lack evidence-based tools to estimate an individual's risk of relapse
58 after treatment in primary care, which may help us more effectively target relapse
59 prevention.

60

61 **Objective**

62 Develop and validate a prognostic model to predict risk of relapse of depression in
63 primary care.

64

65 **Methods**

66 Multilevel logistic regression models were developed, using individual participant
67 data from seven primary care-based studies (n=1244), to predict relapse of
68 depression. The model was internally validated using bootstrapping and
69 generalisability was explored using internal-external cross-validation.

70

71 **Findings**

72 Residual depressive symptoms [Odds ratio (OR): 1.13 (95% CI: 1.07-1.20), $p < 0.001$]
73 and baseline depression severity [OR: 1.07 (1.04-1.11), $p < 0.001$] were associated
74 with relapse. The validated model had low discrimination [C-statistic 0.60 (0.55-
75 0.65)] and miscalibration concerns [calibration slope 0.81 (0.31-1.31)]. On secondary
76 analysis, being in a relationship was associated with reduced risk of relapse [OR:
77 0.43 (0.28-0.67), $p < 0.001$]; this remained statistically significant after correction for
78 multiple significance testing.

79

80 **Conclusions**

81 We cannot currently predict risk of depression relapse with sufficient accuracy in a
82 primary care setting, using routinely recorded measures. Relationship status
83 warrants further research to explore its role as a prognostic factor for relapse.

84

85 **Clinical implications**

86 Until we can accurately stratify patients according to risk of relapse, a universal
87 approach to relapse prevention may be most beneficial, either during acute phase
88 treatment or post-remission. Where possible, this could be guided by the presence
89 or absence of known prognostic factors (e.g. residual depressive symptoms) and
90 targeted towards these.

91 **1. BACKGROUND**

92 Depression is the leading cause of disability worldwide(1); the vast majority of
93 adults seeking treatment for depression are managed in primary care(2). Relapse is
94 common, with around half of people experiencing a relapse within one year of
95 reaching remission(3). This high relapse rate contributes to the overall morbidity and
96 burden associated with depression(4).

97
98 An ability to predict an individual patient's risk of relapse after an episode of
99 depression might assist clinicians in targeting relapse prevention interventions
100 toward those at greatest risk. Well-established prognostic factors associated with
101 increased risk of relapse are: residual depressive symptoms; previous depressive
102 episodes; childhood maltreatment; comorbid anxiety; neuroticism; younger age of
103 first onset, and rumination(5). While the presence or absence of these prognostic
104 factors can help refine estimates of overall prognosis to particular sub-groups, they
105 do not effectively aid risk-stratification at the individual level. Sub-grouping methods
106 have been used to predict average risk of relapse for groups of people with different
107 combinations (or profiles) of prognostic factors(6). However, individualised outcome
108 prediction is best shaped using multiple prognostic factors in combination, in the
109 form of multivariable prognostic models(7).

110
111 A systematic review of prognostic models, undertaken by our group, identified
112 twelve studies of relapse prediction models(8). The majority were at high overall risk
113 of bias (the most significant limitations being inadequate sample size, inappropriate
114 handling of missing data, and calibration or discrimination not reported). The
115 developed models either demonstrated insufficient predictive performance on
116 reported validation by the study authors, or they could not be feasibly implemented in
117 a primary care setting due to the large number and type of included predictors. We
118 concluded that we currently lack evidence-based tools to assist clinicians with risk
119 prediction of depressive relapse in any clinical setting, and that new models are
120 required to give accurate risk predictions in primary care settings.

121
122 **2. OBJECTIVE**

123 To develop and validate a prognostic model, for use in clinical primary care
124 settings, to predict risk of relapse in adults with remitted depression(9).

125

126 **3. METHODS**

127 The methods align with the PROGNosis RESearch Strategy (PROGRESS)
128 recommendations(7) and the study is reported according to the Transparent
129 Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis
130 (TRIPOD-Cluster) guidance(10). A Patient Advisory Group (PAG) contributed to this
131 study, including selecting predictors, definition of outcome, target patient population
132 and clinical application. The study was registered prospectively (ClinicalTrials.gov:
133 NCT04666662). Further methodological details are available in our protocol
134 paper(9).

135

136 **3.1. Source of data, participants and setting**

137 We formed the “PREDICTR” dataset from combined individual participant
138 data (IPD) from UK primary care-based studies(9), identified through a literature
139 search and a review of the NIHR trials registry. Authors were asked to share data if
140 studies included adult patients (18 years and over) with depression, measured using
141 the Patient Health Questionnaire (PHQ-9) at a minimum of three time-points (to
142 identify depression, remission, relapse/no relapse). We excluded studies in patient
143 groups with significant psychiatric comorbidity and feasibility studies. The
144 PREDICTR dataset is derived from all arms (control and intervention) of six RCTs of
145 primary care-based interventions for depression (CADET, CASPER Plus, COBRA,
146 Healthlines Depression, REEACT, and REEACT-2) and one observational cohort
147 study (WYLOW) (Supplemental Table 1.1, Figure 1.1).

148

149 **3.2. Start-point (remission)**

150 Participants were in remission at the point of prediction. Participants must
151 have had case-level depression at baseline [PHQ-9 score of 10 or more(11)], and (at
152 4 months after trial baseline): i) a post-treatment PHQ-9 score below the established
153 cut-off of 10 [consistent with clinical recovery(3,12)] and ii) an improvement of ≥ 5
154 points on the PHQ-9 since depression diagnosis [which aligns with an established
155 reliable change index to identify those with “reliable improvement”(13)].

156

157 **3.3. End-point/outcome (relapse)**

158 We coded participants as relapsed if they fulfilled the following criteria within
159 6-8 months post-remission: i) PHQ-9 score above the diagnostic cut-off (10 or more)
160 and ii) ≥ 5 points greater than their symptom score at the time of remission. This is
161 also consistent with an established criteria for reliable and clinically significant
162 deterioration(13).

163

164 **3.4. Predictors**

165 We identified predictors *a priori*, following a literature review and consensus
166 within the multidisciplinary research team and the PAG. We included predictors that
167 would currently be routinely available in primary care settings at the intended
168 moment of prediction.

169

170 3.4.1. Predictors in primary analysis

171 The following variables have robust evidence for their role as relapse
172 predictors(5,14) and were included in the model:

- 173 • Residual depressive symptoms (PHQ-9 score at remission (0-9); continuous
174 variable);
- 175 • Previous episodes of depression (dichotomous predictor (0=no previous
176 episodes, 1=one or more previous episodes);
- 177 • Comorbid anxiety (measured using the GAD-7(15) in six of the seven studies
178 and the Clinic Interview Schedule-Revised (CIS-R)(16) in REEACT (see
179 Supplemental Table 1.2). These measures were combined to create a
180 composite score (z-score), modelled as a continuous predictor;
- 181 • Baseline severity of depressive symptoms (continuous predictor; PHQ-9 score
182 at baseline (pre-treatment));
- 183 • RCT Intervention: to control for the presence of interventions within the RCTs,
184 we coded the presence or absence of an effective intervention (based on the
185 results of the RCT) as a dichotomous variable. This predictor was intended to
186 control for the intervention as part of the model building process only; when
187 making predictions in real-world primary care this predictor would always be
188 set to zero (i.e., no experimental intervention present).

189

190 3.4.2. Exploratory predictors

191 These are less well-evidenced predictors of relapse and were included as part
192 of an exploratory secondary analysis: age; gender; ethnicity; relationship status;
193 multimorbidity (two or more long-term physical or mental health conditions, excluding
194 depression and comorbid anxiety); employment status (unemployment being those
195 of working age who do not have a job and are actively seeking one); and current
196 antidepressant use(5,14,17).

197

198 **3.5. Sample size**

199 We used the *pmsampsize* package (18) to calculate the required minimum
200 sample size of 722, with 145 events (see protocol for details(9)); our actual sample
201 size (n=1244; 261 events) exceeded this.

202

203 **3.6. Statistical analysis methods**

204

205 3.6.1. Data pre-processing

206 Anonymised data were transferred and stored securely and harmonised in
207 line with the pre-specified harmonisation procedure (Supplemental Table 1.3).

208

209 3.6.2. Data integrity checks (risk of bias)

210 Data were summarised and checked against publications for key features
211 such as number of participants (total and in each study arm), demographics, primary
212 outcomes of study, relapse rates and missing data. Validity of data values were
213 checked on data inspection and irregularities clarified through communication with
214 the original authors. Risk of bias assessment was undertaken using the participants,
215 predictors and outcome domains of PROBAST(19).

216

217 3.6.3. Missing data

218 Missing data were handled using multiple imputation with chained equations
219 (MICE), under a missing at random assumption(20). Missing values were imputed
220 based on the values of other predictors and the outcome, using linear models for
221 continuous predictors (residual symptoms, severity, comorbid anxiety) and logistic

222 models for binary predictors [number of previous episodes, RCT Intervention,
223 outcome (relapse/no relapse)]. Imputation was undertaken for each study separately,
224 preserving the clustering of participants within studies and any between-study
225 heterogeneity in predictor effects and outcome prevalence. Each imputed dataset
226 was then analysed separately using the same statistical methods, and the estimates
227 were combined using Rubin's rules, to produce an overall estimate and measure of
228 uncertainty of each regression coefficient and model performance measures(10). We
229 used thirty imputations, based on the maximum percentage of participants with one
230 or more missing values across all individual studies(20).

231

232 3.6.4. Model development (primary analysis)

233 Multilevel multivariable logistic regression models were built to model the
234 relationship of the predictors with the binary outcome (relapse/no relapse), forcing in
235 all predictors. Model parameters were estimated via unpenalized maximum
236 likelihood estimation, and then penalised post-estimation using a uniform shrinkage
237 factor. The modelling preserved the clustering of participants within studies, with a
238 random effect on the intercept, a random intervention effect, and allowing for
239 between-study correlation in these effects. We explored non-linear relationships in
240 the continuous variables using multivariable fractional polynomials (MFPs)(7).
241 Predictive performance statistics (C-statistic for discrimination; calibration slope and
242 calibration-in-the-large) were calculated for the final developed model, first within
243 each cluster in turn and then pooled using random effects meta-analysis to
244 summarise the model's performance across clusters with estimates of the pooled
245 average and 95% confidence intervals. Prediction intervals were constructed to
246 estimate the model's likely performance in new but similar settings(10). Calibration
247 was also assessed visually by producing calibration plots with smooth calibration
248 curves.

249

250 3.6.5. Model validation

251 The optimism of the developed model was measured using non-parametric
252 bootstrapping. One hundred bootstrap samples (each stratified by study) were
253 produced from the original dataset. Within each bootstrap sample, the same
254 modelling procedures were used as for model development. The model estimated

255 using each bootstrap sample was then applied in both the same bootstrap sample
256 ('apparent performance') and in the original (imputed) dataset ('test performance').
257 Each time, average performance measures were calculated by pooling within-study
258 statistics using meta-analysis, as above.

259

260 Optimism was calculated as the difference between apparent and test
261 performance; this process was repeated one hundred times and the average
262 difference between the bootstrap (apparent) and test performance for each
263 performance statistic provided the estimate of overall optimism for that statistic.
264 Optimism-adjusted performance statistics (C-statistic, calibration slope and
265 calibration-in-the-large) were subsequently derived. The uniform shrinkage factor (in
266 this study, the optimism-adjusted calibration slope) was applied to all of the original
267 estimated beta coefficients (to shrink them toward zero to address overfitting) to
268 produce a penalised logistic regression model. Finally, the intercept was re-
269 estimated (whilst constraining the penalised predictor effects at their shrunken value)
270 to ensure overall calibration was maintained. This formed the final model.
271 Generalisability of the model and between-study heterogeneity in model
272 performance was assessed using internal-external cross-validation (IECV)(21) (see
273 Supplemental Materials 4 for procedure).

274

275 3.6.6. Sensitivity analysis

276 To understand the impact of including a composite measure of comorbid
277 anxiety calculated from both GAD-7 and CIS-R, a sensitivity analysis was performed
278 measuring predictive performance statistics when omitting REEACT and using only
279 GAD-7 as the measure of comorbid anxiety, rather than a z-score.

280

281 3.6.7. Secondary (exploratory) analyses

282 Univariable analyses were performed to evaluate the unadjusted association
283 between each predictor variable and the outcome variable. Where univariable
284 analysis found statistically significant associations (after accounting for multiple
285 significance testing), the model was refit using all of the original included predictors
286 plus the additional exploratory predictor, to explore the impact on model predictive
287 performance (using only studies in which the exploratory predictor was available).

288

289 **3.7. Ethics approval**

290 The University of York's Health Sciences Research Governance Committee
291 confirmed that this study was exempt from full ethical approval as it entailed the
292 secondary analysis of anonymised data from studies that had already received
293 ethical approval.

294

295 **4. RESULTS**

296 **4.1. Summary of data**

297 Table 1 presents the descriptive statistics for the IPD. Supplemental Table 2.1
298 summarises the missing data. Risk of bias and concerns around applicability were
299 low (Supplemental Table 2.4).

300

301 [Table 1]

Table 1: Descriptive statistics for IPD

Variable	Study							
	CADET	CASPER Plus	COBRA	Healthlines Depression	REEACT	REEACT-2	WYLOW	Combined PREDICTR IPD dataset
Total number included in analysis (remission)	158	101	169	110	221	159	326	1244
Number relapsed at 6-8 months (%)	32 (20)	28 (27.7)	19 (11.2)	24 (21.8)	34 (15.4)	17 (10.7)	107 (32.8)	261 (21)
Number with one or more previous episodes of depression (%)	112 (70.8)	82 (81.9)	104 (61.5)	89 (80.9)	154 (69.7)	159 (100)	124 (38)	824 (66.2)
Mean PHQ-9 at baseline (SD)	17.35 (4.20)	15.33 (3.39)	17.27 (4.05)	16 (3.70)	15.96 (3.92)	15.77 (3.54)	15.67 (4.17)	16.17 (3.98)
Mean PHQ-9 at remission (SD)	4.70 (2.69)	5.19 (2.50)	4.08 (2.56)	5.97 (2.30)	4.50 (2.61)	4.53 (2.62)	3.52 (2.31)	4.40 (2.60)
Mean GAD-7 (SD)	12.52 (4.83)	8.25 (4.85)	12.41 (4.96)	11.80 (4.44)	GAD-7 not used in REEACT (CIS-R anxiety measure**)	12.97 (4.44)	13.62 (4.63)	12.43 (4.92)
Mean age (SD)	43.2 (12.9)	71.8 (5.16)	45.3 (13.9)	Not applicable (age is categorical*)	40.3 (13.1)	43.3 (14.7)	41.7 (13.8)	45.1 (15.63)

Gender: Number Female (%)	113 (71.5)	65 (64.3)	100 (59.2)	80 (72.7)	148 (67)	102 (64.2)	190 (58.3)	798 (64.1)
Ethnicity*: Number White (%)	137 (86.7)	100 (99)	165 (97.6)	107 (97.2)	216 (97.7)	154 (97.5)	292 (94.2)	1171 (95.4)
Employment*: Number Employed (%)	133 (82.8)	Not collected	132 (78.1)	105 (95.4)	213 (96.3)	149 (93.7)	214 (65.6)	946 (82.9)
Relationship status*: Number In a relationship (%)	69 (43.7)	Not collected	106 (62.7)	Not collected	140 (75.7)	93 (58.9)	Not collected	408 (60.9)
Multimorbidity*: Number with multimorbidity (%)	88 (55.7)	86 (85.1)	96 (56.8)	Not collected	Not collected	Not collected	87 (29.7)	357 (49.5)
Current antidepressant use: Number on antidepressants at remission (%)	113 (71.5)	34 (41)	128 (75.7)	98 (90.7)	Not collected	55 (41.4)	130 (40.2)	558 (57.3)

303 *See Supplemental Table 2.2 for further detail on categorical predictors

304 **See Supplemental Table 2.3 for summary statistics for CIS-R anxiety subscale in REEACT

305 **4.2. Univariable analysis**

306 Table 2 presents the results from univariable multilevel models. Residual
307 symptoms [OR: 1.13 (1.07-1.20)] and severity [OR: 1.07 (1.04-1.11)] were
308 statistically significantly associated with relapse; number of previous episodes and
309 comorbid anxiety were not.

310

311 [Table 2]

Table 2: Univariable and multivariable associations between outcome and predictors (primary and secondary analysis)

Predictor	Univariable analysis*			Primary analysis** (multivariable) Seven studies (n=1244)		Secondary analysis** (multivariable) Four studies (n=707)****	
	Unadjusted odds ratio (95% CI)	p-value	Number of participants (N)	Regression beta coefficient (95% CI)***	p-value	Regression beta coefficient (95% CI)	p-value
Predictors in primary analysis							
Number of previous episodes	1.19 (0.84 to 1.72)	0.319	1117	0.13 (-0.24 to 0.50)	0.500	-0.15 (-0.66 to 0.37)	0.582
Residual symptoms	1.13 (1.07 to 1.20)	<0.001	1244	-2.11 (-3.07 to -1.14)	<0.001	0.07 (-0.01 to 0.15)	0.081
Baseline severity of depressive symptoms	1.07 (1.04 to 1.11)	<0.001	1244	0.09 (0.04 to 0.13)	<0.001	0.07 (0.01 to 0.13)	0.020
Comorbid anxiety (z-score)	1.04 (0.90 to 1.20)	0.589	1239	-0.13 (-0.30 to 0.03)	0.936	-0.12 (-0.37 to 0.14)	0.363
RCT intervention	0.99 (0.60 to 1.66)	0.981	1244	0.03 (-0.59 to 0.65)	0.936	-0.40 (-0.84 to 0.04)	0.076
Predictors in secondary analysis							
Relationship status (in a relationship)	0.43 (0.28 to 0.67)	<0.001	670			-0.79 (-1.23 to -0.34)	0.001
Comorbid anxiety (GAD-7)	1.00 (0.97 to 1.03)	0.943	1019				
Gender (female)	0.87 (0.62 to 1.10)	0.196	1244				
Ethnicity (white)	1.59	0.138	1227				

		(0.86 to 2.93)			
Age (continuous)		1.01 (1.00 to 1.02)	0.198	1133	
Age – categorical (years old)	<40	Reference category		1243	
	40-49	1.16 (0.80 to 1.68)	0.433		
	50-59	1.31 (0.88 to 1.95)	0.180		
	60-69	0.85 (0.51 to 1.43)	0.543		
	70+	1.93 (1.04 to 3.59)	0.037		
Employment (employed)		0.76 (0.52 to 1.11)	0.161	1141	
Multimorbidity		1.31 (0.90 to 1.90)	0.158	721	
Current antidepressant medication		0.97 (0.70 to 1.35)	0.853	974	

313 * Univariable associations (unadjusted) between predictors and relapse within 6-8 months, after accounting for clustering by study

314 **Beta coefficients before shrinkage (predictors adjusted for other variables in column)

315 *** Intercept (baseline risk): -1.55 (95% CI: -2.12 to -1.00); Standard deviation of random effect on intercept: 0.43 (95% CI: 0.14 to 1.38); Standard
316 deviation of random effect on slope (RCT Intervention): 0.49 (95% CI: 0.11 to 2.16); Correlation between random effects: -0.23 (-0.93 to 0.84)

317 ****Secondary analysis included only studies with relationship status available (n=707)

318 **4.3. Model development and apparent predictive performance**

319 Table 2 presents the results of multivariable, multilevel logistic regression
320 analysis for the primary analysis. The developed model, prior to shrinkage, had a
321 pooled apparent performance of: C-statistic 0.62 (95% CI: 0.57-0.67), calibration
322 slope of 0.95 (95% CI: 0.54-1.36), and calibration-in-the-large of 0.03 (95% CI: -0.49-
323 0.54). See Supplemental Materials 3 for within-study performance statistics
324

325 **4.4. Internal validation, shrinkage and final equation**

326 Optimism-adjusted performance statistics, after bootstrapping, were: C-statistic
327 0.60, calibration slope 0.81, and calibration-in-the-large 0.03. The final model (Table
328 3) was produced by multiplying the original beta regression coefficients (from Table
329 2) by 0.81 (the optimism-adjusted calibration slope), and re-estimating the intercept
330 to ensure calibration-in-the-large.

331

332 [Table 3]

Table 3: Summary of model's predictive performance for primary, sensitivity and secondary analyses

Measure of predictive performance*	Primary analysis			Secondary analysis (exploring relationship status as a predictor)	
	Development (apparent performance)	Internal validation (optimism-adjusted performance)	Internal-external cross validation	Without relationship status	With relationship status
				Apparent performance	Apparent performance
Number of participants (number of clusters)	1244 (7)	1244 (7)	1244 (7)	707 (4)	707 (4)
C-statistic (95% CI*)	0.62 (0.57 to 0.67)	0.60	0.60 (0.55 to 0.65)	0.60 (0.54 to 0.66)	0.63 (0.57 to 0.70)
Calibration slope (95% CI*)	0.95 (0.53 to 1.36)	0.81	0.81 (0.31 to 1.31)	0.94 (0.37 to 1.51)	0.96 (0.56 to 1.36)
Calibration-in-the-large (95% CI*)	0.03 (-0.49 to 0.54)	0.03	0.00 (-0.61 to 0.61)	0.01 (-0.25 to 0.27)	0.01 (-0.20 to 0.23)

334 *Pooled predictive performance across studies (within-study performance statistics all available in Supplemental Materials); 95% CI presented,
 335 where applicable (not for optimism-adjusted statistics)

336 Final shrunken model: Intercept (baseline risk): -1.49; Shrunken beta coefficients: Number of previous episodes: 0.11, Residual symptoms: -
337 1.71; Severity: 0.07; Comorbid anxiety: -0.11; RCT Intervention: 0.02
338 To calculate the risk of relapse within 6-8 months, using this model, one would use the following formula: $\exp(\text{person's risk score}) \div [1 +$
339 $\exp(\text{person's risk score})]$
340 Where person's risk score (linear predictor) = $-1.49 + 0.11 (\text{Number of previous episodes}^1) - 1.71 (\text{Residual symptoms}^2) + 0.07 (\text{Severity}^3) -$
341 $0.11 (\text{comorbid anxiety}^4) + 0.02 (\text{RCT Intervention}^5)$

¹ No previous episodes = 0; One or more previous episodes = 1

² $X^{\wedge} -0.5 - 0.43$ (where $X = (\text{residual_symptoms}+1)$) – this is the adjustment for non-linear transformation and mean-centring

³ Severity – 16.17

⁴ Comorbid_anx_zscore + 0.118

⁵ This would be zero when applied in clinical practice, outside the context of an RCT

342 **4.5. Internal-external cross-validation**

343 Generalisability of the model was assessed using IECV(21). Calibration plots
344 were compared for each validation in each of the different studies (Figure 1). These
345 demonstrate inadequate calibration in most studies and significant heterogeneity in
346 predictive performance across clusters. For example, WYLOW study shows severe
347 miscalibration, with estimated risks generally too low, whereas in the COBRA study
348 estimated risks are generally too high. In some studies, calibration was generally
349 excellent (e.g., Healthlines Depression)

350

351 [Figure 1: Calibration plots for internal-external cross-validation within each study]

352 **4.6. Sensitivity analysis**

353 We removed REEACT and repeated the modelling process on the remaining six
354 studies, to assess the impact of using z-scores to model comorbid anxiety. This did
355 not change the study conclusions (see Supplemental Material 5 for analysis).

356

357 **4.7. Secondary analysis**

358 On univariable multilevel logistic analysis, relationship status was a highly
359 statistically significant predictor (after adjusting the significance level to account for
360 multiple significance testing using the Bonferroni correction). To further explore
361 relationship status as a predictor of relapse, we repeated the model development
362 procedures used in the primary analysis for the studies that included relationship
363 status (CADET, COBRA, REEACT, and REEACT-2). We conducted these analyses
364 both with and without the relationship status variable to provide a direct comparison
365 (see Supplemental Material 6). Relationship status remained a statistically significant
366 relapse predictor after adjusting for other prognostic factors (previous episodes,
367 residual symptoms, severity, and comorbid anxiety).

368

369 **5. DISCUSSION**

370 We developed a model for predicting depression relapse in adults with remitted
371 depression in primary care. Generally, the model had suboptimal predictive
372 performance, with heterogeneous calibration across clusters on IECV and C-statistic
373 below that required for acceptable discrimination. We would not recommend
374 implementation in its current form, though calibration was promising in a subset of
375 studies. Secondary analysis found a statistically significant association between
376 relationship status and relapse.

377

378 **5.1. Findings in the context of the literature**

379 The performance of this model was in line with the performance measures for
380 previous relapse prediction models(8). Residual symptoms were associated with
381 relapse, which is consistent with existing literature(5,14). Residual symptoms are
382 also associated with a more chronic depression course and poorer psychosocial
383 functioning(22) and, as such, are an established treatment target in depression. The

384 pre-existing evidence for severity as a prognostic factor for relapse is more equivocal
385 than for residual symptoms(5). Residual symptoms are more likely in people with
386 more severe initial depressive illness(23) and so the presence of residual symptoms
387 may be a mediator of the relationship between baseline severity and relapse.

388

389 The lack of association between previous episodes and relapse in our study is
390 not consistent with the consensus view(5,14). It is possible that this is because it was
391 limited to a binary variable (“no previous episodes” or “one or more previous
392 episodes”) in our study. However, the prognostic effect of previous episodes on
393 relapse is known to be strongest when comparing any number of previous episodes
394 to no previous episodes rather than being related to a specific number of previous
395 episodes(5). An alternative explanation may be that previous episodes is most
396 strongly associated with recurrence (which occurs over a longer time period than
397 relapse) and therefore our 6-8 month follow-up was not sufficient to detect this
398 association(5). Comorbid anxiety is recognised as a predictor of relapse(5); in
399 particular, higher levels of anxiety at baseline have been found to predict a shorter
400 time to relapse after treatment(24). In this study, we used anxiety at baseline (when
401 depressed), using GAD-7 and the CIS-R anxiety subscale as measures of comorbid
402 anxiety. It may be that an isolated measure of anxiety symptom severity at a single
403 time-point is a crude measure and less important than knowing an individual’s history
404 of comorbid anxiety.

405

406 Marital status (being single) is a risk factor for developing depression(17). A
407 recent study also identified marital status (being single or no longer married) as
408 being associated with a worse prognosis (more depressive symptoms) at 3-4 months
409 (but not beyond 3-4 months)(25). While marital status is not an established predictor
410 of relapse(5,17), our systematic review of prognostic models identified low-quality
411 evidence of an association between relapse and marital status(8). The lack of
412 association between relapse and other exploratory predictors (age, gender, ethnicity,
413 employment status and multimorbidity) is consistent with other findings from the
414 literature(5,14).

415

416 The prognostic factor research cited here is based on longitudinal studies,
417 which examine the predictive value of specific variables based on sample-level

418 trends. As our systematic review (8) and the current study demonstrate, combining
419 these well-evidenced relapse predictors to produce individualised predictions of
420 relapse risk remains challenging and calibration remains suboptimal for the purpose
421 of personalised decision-making(7).

422

423 **5.2. Strengths and limitations**

424 This study was conducted according to best practice recommendations for
425 methodology and reporting, with a sufficient sample size to produce precise risk
426 estimates(26). By pre-selecting predictors with a robust, pre-existing evidence base,
427 we aimed to mitigate the risk of overfitting and additional uncertainty associated with
428 more data-driven approaches for predictor selection. While cohort studies and RCTs
429 are recommended sources of data for prognostic model development(7), participants
430 may differ from the general population in important ways and results should be
431 interpreted with this in mind. For example, the majority of participants in this study
432 were white, which limited our ability to explore ethnicity as a predictor. Some
433 potentially useful predictors were not included (e.g., neuroticism, childhood
434 maltreatment and rumination(5)) as they were not coded for in our cohorts (these are
435 also not routinely measured in general practice settings). There was a risk of
436 selection bias given the way studies were selected for inclusion. There was
437 heterogeneity in IPD study populations (e.g., CASPER Plus included older adults),
438 settings and treatment dose. WYLOW followed-up patients after low-intensity CBT
439 whereas interventions in other studies (like COBRA) were more intensive and
440 delivered over a greater number of sessions. While interventions were controlled for
441 in the analysis, this heterogeneity could explain some of the observed miscalibration.
442 Finally, outcomes (remission and relapse) were defined according to PHQ-9 (which
443 is less optimal than diagnostic interview) and over a time-period (6-8 months)
444 necessitated by the IPD data collection points [although one that is aligned with
445 established definitions of relapse(5)].

446

447 **5.3. Implications for future research**

448 The strong statistically significant association between relationship status and
449 relapse in our study warrants further confirmatory prognostic factor research going
450 forwards, to better understand the association between marital/relationship and
451 relapse and how this relates to its better understood association with depression

452 more generally. Further research is needed to better understand whether other
453 relapse predictors can be captured and recorded in an acceptable and valid way by
454 primary care health professionals. For example, we know that routinely asking
455 people about childhood maltreatment is not harmful(27), therefore it may be feasible
456 to explore the use of childhood maltreatment and adversity as part of routine relapse
457 risk assessment in primary care. There have been some efforts to develop clinically
458 useful and valid brief instruments to measure rumination(28), which could be
459 explored in a primary care setting. If there were robust evidence supporting the
460 clinical utility of measuring and documenting additional relapse predictors, health
461 professionals might then adopt this as routine practice. Better data linkage and
462 systems integration across health and wellbeing services may also help address this
463 problem.

464

465 In this study, we modelled the outcome of relapse as a binary outcome, which
466 enabled us to incorporate criteria for reliable and significant change to define the
467 outcome. It would be informative as part of future work to model the outcome on its
468 continuous scale (i.e., build a linear model that estimates the outcome value, rather
469 than the probability of a binary outcome), with dichotomisation done post-prediction.
470 We focused on outcome occurrence by 6-8 months, but other time-points may be of
471 interest.

472

473 As the IECV in this study demonstrated, there was heterogeneity in the
474 external performance and generalisability of the model was not guaranteed. The
475 studies from which IPD were drawn were generally comparable in terms of baseline
476 demographic variables such as age, gender and ethnicity. Our IECV used the mean
477 intercept and predictor-outcome associations to estimate performance. An
478 alternative approach that could be considered in future work is local recalibration or
479 intercept selection, where similarities in the outcome frequency or baseline
480 characteristics (for example, mean age or proportion female) of a new population of
481 interest is used to guide the intercept when applying the model in a different
482 context(29). If the predictive performance of relapse prediction models can be
483 improved in the future through recalibration or updating, clinical usefulness (using
484 net benefit analysis) must be considered prior to implementation.

485

486 **6. CLINICAL IMPLICATIONS**

487 We cannot currently predict an individual's risk of relapse with a high degree of
488 accuracy, and existing relapse risk prediction models are unlikely to be suitable to
489 guide the provision of relapse prevention. There are different approaches to
490 prevention: universal approaches, which target whole populations; selective
491 approaches, which target higher-risk groups; and indicated approaches, directed at
492 individuals. In the absence of sufficiently accurate relapse risk prediction tools, we
493 argue that a universal approach to relapse prevention of depression in primary care
494 is currently warranted. This is likely to require a systems approach to mitigating the
495 risk of and improving the management of relapse for all patients. This could mean
496 targeting treatment at known prognostic factors (for example, focussing on reducing
497 residual symptoms) or providing interventions as part of treatment during the acute
498 phase of depression that target mechanisms of relapse (for example, prioritising
499 relapse prevention planning)(30). Longer term, brief, inexpensive and scalable
500 relapse prevention interventions are likely to be required for use in primary care.

501

502 **FIGURE LEGENDS**

503 Figure 1: Calibration plots for internal-external cross-validation within each study

504

505 **ACKNOWLEDGEMENTS**

506 Thank you to the patient advisory group without whose contributions this study would
507 not have been possible: Joanne Castleton, Penney Mayall, Gillian Payne, Sue Penn,
508 and Emma Williams.

509

510 **COMPETING INTERESTS**

511 None.

512

513 **FUNDING**

514 This report is independent research supported by the National Institute for Health
515 and Care Research (NIHR Doctoral Research Fellowship, Dr Andrew Moriarty, DRF-
516 2018-11-ST2-044). KIES, RDR and LA are supported by funding from the NIHR
517 Birmingham Biomedical Research Centre (BRC). RDR, SG, DAR and CS are NIHR
518 Senior Investigators. The views expressed in this publication are those of the

519 author(s) and not necessarily those of the NHS, the NIHR, or the Department of
520 Health and Social Care.

521

522 **AUTHOR CONTRIBUTIONS**

523 ASM, CCG, SG, SA and DM conceptualised the study and acquired funding. ASM
524 led on data curation, study administration and oversight, and wrote the original draft.
525 ASM, LWP, KIES, LA, RDR, JEJB and NM contributed to methodology and data
526 analysis. SG, JD, DAR and CS were responsible for collecting the data and agreed
527 to share these for the purpose of this study. All authors contributed to the pre-
528 registered protocol and study plan and all authors contributed to and approved the
529 final manuscript.

530

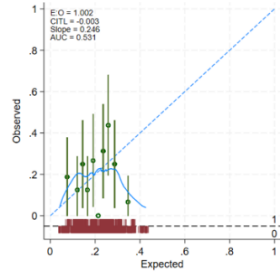
531 **REFERENCES**

- 532 1. World Health Organisation. Weekly Epidemiological Report [Internet]. Vol. 30,
533 World Health Organization. 2017. Available from:
534 <http://www.who.int/wer/2016/wer9130.pdf?ua=1>
- 535 2. Ferenchick EK, Ramanuj P, Pincus HA. Depression in primary care: part 1-
536 screening and diagnosis. *BMJ*. 2019;365:l794.
- 537 3. Ali S, Rhodes L, Moreea O, McMillan D, Gilbody S, Leach C, et al. How
538 durable is the effect of low intensity CBT for depression and anxiety?
539 Remission and relapse in a longitudinal cohort study. *Behaviour Research and*
540 *Therapy* [Internet]. 2017;94:1–8. Available from:
541 <http://dx.doi.org/10.1016/j.brat.2017.04.006>
- 542 4. Gauthier G, Mucha L, Shi S, Guerin A. Economic burden of relapse/recurrence
543 in patients with major depressive disorder. *J Drug Assess* [Internet].
544 2019;8(1):97–103. Available from:
545 <https://doi.org/10.1080/21556660.2019.1612410>
- 546 5. Buckman JEJ, Underwood A, Clarke K, Saunders R, Hollon SD, Fearon P, et
547 al. Risk factors for relapse and recurrence of depression in adults and how
548 they operate: A four-phase systematic review and meta-synthesis. *Clin*
549 *Psychol Rev*. 2018;64(July 2017):13–38.
- 550 6. Saunders R, Cohen ZD, Ambler G, Derubeis RJ, Wiles N, Kessler D, et al. A
551 Patient Stratification Approach to Identifying the Likelihood of Continued
552 Chronic Depression and Relapse Following Treatment for Depression. *J Pers*
553 *Med*. 2021 Dec 1;11(12).
- 554 7. Riley RD, van der Windt D, P C, Moons K. *Prognosis Research in Healthcare:*
555 *Concepts, Methods, and Impact*. First edit. Oxford University Press.; 2019.
- 556 8. Moriarty AS, Meader N, Snell KIE, Riley RD, Paton LW, Dawson S, et al.
557 Predicting relapse or recurrence of depression: Systematic review of
558 prognostic models. *British Journal of Psychiatry*. Cambridge University Press;
559 2022.
- 560 9. Moriarty AS, Paton LW, Snell KIE, Riley RD, Buckman JEJ, Gilbody S, et al.
561 The development and validation of a prognostic model to PREDICT Relapse of

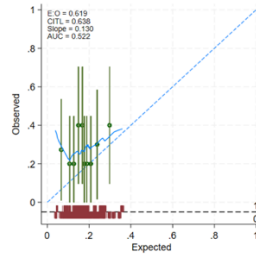
- 562 depression in adult patients in primary care: protocol for the PREDICTR study.
563 *Diagn Progn Res.* 2021 Dec;5(1).
- 564 10. Debray TPA, Collins GS, Riley RD, Snell KIE, Van Calster B, Reitsma JB, et
565 al. Transparent reporting of multivariable prediction models developed or
566 validated using clustered data (TRIPOD-Cluster): explanation and elaboration.
567 *BMJ [Internet].* 2023 Feb 7;e071058. Available from:
568 <https://www.bmj.com/lookup/doi/10.1136/bmj-2022-071058>
- 569 11. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: Validity of a brief
570 depression severity measure. *J Gen Intern Med.* 2001;16(9):606–13.
- 571 12. Clark DM, Layard R, Smithies R, Richards DA, Suckling R, Wright B.
572 Improving access to psychological therapy: Initial evaluation of two UK
573 demonstration sites. *Behaviour Research and Therapy.* 2009 Nov;47(11):910–
574 20.
- 575 13. McMillan D, Gilbody S, Richards D. Defining successful treatment outcome in
576 depression using the PHQ-9: A comparison of methods. *J Affect Disord*
577 *[Internet].* 2010;127(1–3):122–9. Available from:
578 <http://dx.doi.org/10.1016/j.jad.2010.04.030>
- 579 14. Wojnarowski C, Firth N, Finegan M, Delgadillo J. Predictors of depression
580 relapse and recurrence after cognitive behavioural therapy: a systematic
581 review and meta-analysis. *Behavioural and Cognitive Psychotherapy [Internet].*
582 2019 Sep 21 [cited 2020 Mar 22];47(5):514–29. Available from:
583 [https://www.cambridge.org/core/product/identifier/S1352465819000080/type/jo](https://www.cambridge.org/core/product/identifier/S1352465819000080/type/journal_article)
584 [urnal_article](https://www.cambridge.org/core/product/identifier/S1352465819000080/type/journal_article)
- 585 15. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing
586 generalized anxiety disorder: The GAD-7. *Arch Intern Med.*
587 2006;166(10):1092–7.
- 588 16. Lewis G, Pelosi AJ, Araya R, Dunn G. Measuring psychiatric disorder in the
589 community: A standardized assessment for use by lay interviewers. *Psychol*
590 *Med.* 1992;22(2):465–86.
- 591 17. Burcusa SL, Iacono WG. Risk for recurrence in depression. *Clin Psychol Rev.*
592 2007;27(8):959–85.
- 593 18. Ensor J. PMSAMPSIZE: Stata module to calculate the minimum sample size
594 required for developing a multivariable prediction model. *Statistical Software*
595 *Components S458569, Boston College Department of Economics;* 2018.
- 596 19. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al.
597 PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction
598 Model Studies. *Ann Intern Med [Internet].* 2019;170(1):51. Available from:
599 <http://annals.org/article.aspx?doi=10.7326/M18-1376>
- 600 20. White IR, Royston P, Wood AM. Multiple imputation using chained equations:
601 Issues and guidance for practice. *Stat Med.* 2011;30(4):377–99.
- 602 21. Takada T, Nijman S, Denaxas S, Snell KIE, Uijl A, Nguyen TL, et al. Internal-
603 external cross-validation helped to evaluate the generalizability of prediction
604 models in large clustered datasets. *J Clin Epidemiol [Internet].* 2021;137:83–
605 91. Available from: <https://doi.org/10.1016/j.jclinepi.2021.03.025>
- 606 22. Kennedy N, Foy K. The Impact of Residual Symptoms on Outcome of Major
607 Depression. *Curr Psychiatry Rep.* 2005;7:441–6.
- 608 23. Paykel E, Ramana R, Cooper Z, Hayhurst H, Kerr J, Barocka A. Residual
609 symptoms after partial remission: An important outcome in depression.
610 *Psychol Med.* 1995;25(6):1171–80.

- 611 24. Forand NR, Derubeis RJ. Pretreatment anxiety predicts patterns of change in
612 cognitive behavioral therapy and medications for depression. *J Consult Clin*
613 *Psychol.* 2013;81(5):774–82.
- 614 25. Buckman JEJ, Saunders R, Stott J, Arundell LL, O’Driscoll C, Davies MR, et
615 al. Role of age, gender and marital status in prognosis for adults with
616 depression: An individual patient data meta-analysis. *Epidemiol Psychiatr Sci.*
617 2021 Jun 4;30:e42.
- 618 26. Riley RD, Ensor J, Snell KIE, Harrell FE, Martin GP, Reitsma JB, et al.
619 Calculating the sample size required for developing a clinical prediction model.
620 *BMJ* [Internet]. 2020 Mar 18;m441. Available from:
621 <http://www.bmj.com/lookup/doi/10.1136/bmj.m441>
- 622 27. Becker-Blease KA, Freyd JJ. Research participants telling the truth about their
623 lives: The ethics of asking and not asking about abuse. *American*
624 *Psychologist.* 2006 Apr;61(3):218–26.
- 625 28. Marchetti I, Mor N, Chiorri C, Koster EHW. The brief state rumination inventory
626 (BSRI): Validation and psychometric evaluation. *Cognit Ther Res.* 2018 Feb
627 26;42(4):447–60.
- 628 29. Debray TPA, Moons KGM, Ahmed I, Koffijberg H, Riley RD. A framework for
629 developing, implementing, and evaluating clinical prediction models in an
630 individual participant data meta-analysis. *Stat Med.* 2013 Aug 15;32(18):3158–
631 80.
- 632 30. Bockting CL, Hollon SD, Jarrett RB, Kuyken W, Dobson K. A lifetime approach
633 to major depressive disorder: The contributions of psychological interventions
634 in preventing relapse and recurrence. *Clin Psychol Rev* [Internet]. 2015;41:16–
635 26. Available from: <http://dx.doi.org/10.1016/j.cpr.2015.02.003>
636

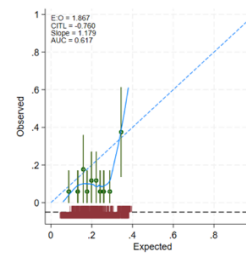
1) CADET



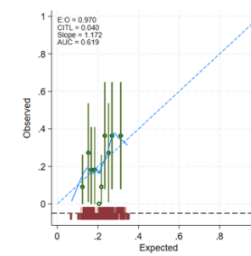
2) CASPER Plus



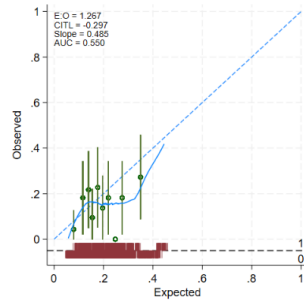
3) COBRA



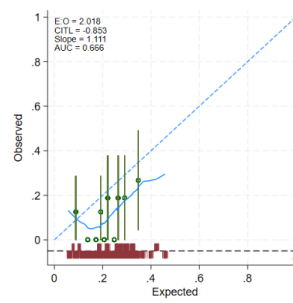
4) Healthlines Depression



5) REEACT



6) REEACT-2



7) WYLOW

