A comparison of the three year course between chronic depression and depression with multiple vs. few prior episodes

Elke Humer[a], Krisztina Kocsis-Bogar[a], Thomas Berger[b], Johanna Schröder[c], Christina Späth[d], Björn Meyer[e,f], Steffen Moritz[c], Wolfgang Lutz[g], Thomas Probst[a,1], Jan Philipp Klein[d,1,*]

[a] Department for Psychotherapy and Biopsychosocial Health, Danube University Krems, Krems, Austria
[b] Department of Clinical Psychology and Psychotherapy, University of Bern, Bern, Switzerland
[c] Department of Psychiatry and Psychotherapy, University Medical-Center Hamburg-Eppendorf, Hamburg, Germany
[d] Department of Psychiatry and Psychotherapy, Lübeck University, 23538 Lübeck, Germany
[e] Research Department, Gaia AG, Hamburg, Germany
[f] Department of Psychology, City, University of London, United Kingdom
[g] Department of Psychology, University of Trier, Trier, Germany

ARTICLE INFO

ABSTRACT

This study tested the hypothesis that chronic depression (CD) is more similar to depression with multiple prior episodes (ME) than to depression with few prior episodes (FE). Data from participants ($n = 1013$) with mild to moderate depressive symptoms (Patient Health Questionnaire [PHQ-9] score 5 – 14) who took part in a randomized control trial of an internet intervention for depression (EVIDENT trial) were re-analyzed. The MINI-interview was conducted to diagnose CD ($n = 376$). If CD was not diagnosed, the self-reported number of depressive episodes was used to categorize participants as having episodic depression with up to five (FE, $n = 422$) or more than five (ME, $n = 215$) prior episodes. Over a three-year period, participants were assessed repeatedly regarding the course of depression (PHQ-9, QIDS), quality of life (SF-12) and therapeutic progress (FEP-2). At baseline, most scores were different between CD and FE but comparable between CD and ME. Time to remission did not differ between CD and ME but was longer in CD compared to FE. Results suggest that ME closely resembles CD and that CD differs from FE.

1. Introduction

Major depressive disorder is one of the most frequent mental disorders with a lifetime prevalence of 15% (Wittchen et al., 2011). Chronic depression (CD) develops in approximately 20 to 30% of depressed individuals (Jobst et al., 2016; Murphy and Byrne, 2012). Definitions of CD vary among diagnostic manuals and studies (see Jobst et al., 2016). In DSM-5 the diagnostic entity "persistent depressive disorder" was introduced as a new category that aims to combine dysthymic disorder (including double depression) and chronic major depression (American Psychiatric Association, 2013), as these forms of depression could not be differentiated satisfactorily based on their symptomatology and course, and because both illnesses respond to the same treatments (Ildirli et al., 2015; Jobst et al., 2016). The current ICD-10 (WHO, 2010) does not offer a distinct diagnosis of CD in a similar way as DSM-5, with only the category dysthymia corresponding to CD.

Currently, "chronicity" is defined by the duration of depressive symptoms, with a threshold of two years (American Psychiatric Association, 2013). Some clinicians, however, also label highly recurrent depression as "chronic". For example, Barnhofer et al. (2009) combined patients with more than three episodes and patients with CD in their trial on mindfulness-based cognitive therapy for CD. Research results indeed indicate similarities between recurrent depression and CD. A chronic course is predicted by similar risk factors as a recurrent course of depression (Hoertel, 2017; ten Have et al., 2018). Even more similarities can be found between CD and depression with multiple prior episodes (ME). CD is characterized by more psychiatric co-morbidities, a more difficult treatment course, more submissive and hostile interpersonal styles, an earlier onset of depression, and more childhood adversities than non-chronic depression (Bird et al., 2018; Klein et al., 2015; Köhler et al., 2019; Nelson et al., 2017), whereby the

latter two (onset, childhood adversities) are also characteristics of ME compared to depression with few prior episodes (FE) (Bockting et al., 2005; Ma and Teasdale, 2004).

Previous research suggests that CD shows more similarities to ME than to FE; however, to the best of our knowledge individuals with CD have never been directly compared to those with ME vs. FE. Using data from a randomized controlled trial, we compared these three groups (CD, ME and FE) with regard to differences in clinical variables (depressive symptoms, quality of life, well-being, symptom distress, interpersonal problems, and incongruence) at baseline (cross-sectional) and over the course of three years (longitudinal). Based on the literature cited above, we hypothesized that participants with the diagnosis of CD closely resemble (at baseline and over the course of three years) those with ME but differ from those with FE.

2. Methods

This study is based on a multicenter (five sites in Germany) randomized, controlled and assessor-blinded trial comparing the internet-based intervention "Deprexis" with the control condition care-as-usual (CAU). The trial was conducted in compliance with the Declaration of Helsinki and approved by the Ethics Committee of the German Psychological Association (SM 04_2012) and registered with ClinicalTrials.gov (NCT01636752). The full protocol (Klein et al., 2013), results regarding the effectiveness of the intervention (Klein et al., 2016, 2017b) as well as results for moderators of treatment effects (Probst et al., 2020) have been published previously.

2.1. Participants

The recruitment process has been described in detail elsewhere (Klein et al., 2017a). Inclusion criteria were age between 18 and 65, internet access, the ability to communicate in German, and self-reported mild to moderate depressive symptoms on the Patient Health Questionnaire-9 (PHQ-9; (Kroenke et al., 2001). Mild depressive symptoms were defined as PHQ-9 scores between 5 and 9, whereas moderate depressive symptoms were PHQ-9 scores from 10 to 14. Exclusion criteria were acute suicidality, lifetime diagnosis of bipolar disorder or schizophrenia (Klein et al., 2016; 2017b).

In total, $n = 1013$ participants (Fig. 1) were randomized (ratio 1:1) into the control condition (care-as-usual, CAU) or the intervention condition (CAU plus Deprexis). Randomization into the intervention or control condition was stratified by PHQ-9 (PHQ-9 < 10 vs. PHQ-9 ≥ 10), as participants with PHQ-9 < 10 (mild depressive symptoms) received unguided Deprexis and participants with PHQ-9 ≥ 10 (moderate depressive symptoms) received guided Deprexis. Based on the Mini International Neuropsychiatric Interview (MINI), a structured interview conducted at baseline (see measures), participants were categorized as having dysthymia or not. Dysthymia was diagnosed in $n = 376$ (37%) participants (CD group). Among CD-participants, 198 (53%) reported a history of up to five depressive episodes and 178 (47%) of more than five previous episodes. In 115 CD-participants dysthymia was accompanied by a current major depressive disorder, thus meeting the criteria of double depression according to DSM-IV (American Psychiatric Association, 1994). Participants without a MINI dysthymia diagnosis were categorized according to their self-reported number of depressive episodes as either non-chronic with a history of up to five depressive episodes (few episodes; FE group, $n = 422$, 42%) or non-chronic with more than five depressive episodes (multiple prior episodes; ME group, $n = 215$, 21%).

2.2. Intervention

All participants were permitted to use any form of treatment, including medication and psychotherapy for depression or other conditions. This constituted CAU. Participants in the control condition received solely CAU during the first 12-months. Following a naturalistic design approach, CAU was not influenced by the investigators but monitored during the course of the study (Klein et al., 2017). The control group participants were offered access to a psychological internet intervention (Deprexis) 12 months after the baseline assessment.

Participants in the intervention condition received CAU as well as immediate access to the internet intervention (Deprexis). Immediately after randomization, they were offered access to the 12-week CBT-based internet intervention Deprexis. Briefly, the Deprexis program consists of ten modules, including therapeutic techniques such as cognitive restructuring, behavioral activation, acceptance and mindfulness, relaxation exercises, problem-solving, and positive psychology interventions (Meyer et al., 2009).

Details on the treatment utilization (antidepressant medication, inpatient psychiatric treatment, outpatient psychiatric treatment) in the CAU as well as Deprexis group have been reported previously (Klein et al., 2016, 2017b) and details on adherence to Deprexis were published by Fuhr et al. (2018).

2.3. Measures

2.3.1. Mini International Neuropsychiatric Interview (MINI)

The MINI 6.0 (Sheehan et al., 1998) is a structured interview to assess DSM-IV (American Psychiatric Association, 1994) and ICD-10 (World Health Organization, 2010) psychiatric disorders. The MINI was administered by trained raters via telephone. Raters were mostly degree-educated psychologists but also included advanced graduate students majoring in psychology or medicine. Before they were permitted to rate trial participants, raters completed face-to-face training and were required to demonstrate adequate interrater reliability on an audiotaped interview.

2.3.2. Patient Health Questionnaire-9 (PHQ-9)

The PHQ-9 is a self-administered version of the PRIME-MD diagnostic instrument for common mental disorders (Kroenke et al., 2001). The nine items known as PHQ-9 represent the depression module; each item is rated on a 4-point Likert-scale from "0″ (not at all) to "3″ (nearly every day). The PHQ-9 score can thus range between 0 and 27, and remission can be defined as a total score below the threshold for mild depressive symptoms (PHQ < 5) (Nierenberg and DeCecco, 2001). The PHQ-9 has been shown to be a valid and reliable measure of depression that is sensitive to change (Löwe et al., 2004; Nierenberg and DeCecco, 2001). The PHQ-9 represented the primary outcome and was administered at baseline and at 14 follow-ups: 3-months, monthly from 4 to 12 months, 18-months, 24-months, 30-months, 36-months). In our sample, internal consistency was good (α = 0.80 – 0.85).

2.3.3. Quick Inventory of Depressive Symptomatology (QIDS)

The PHQ-9 was complemented by the clinician-rated version of the QIDS via telephone interviews. Depressive symptom severity is assessed by 16 items referring to the last seven days, focusing on the nine DSM-IV criterion symptom domains measuring psychological as well as somatic symptoms (Rush et al., 2003). The QIDS was applied at baseline and at two follow-ups: 3-months and 12-months. The internal consistency of QIDS was satisfactory (α = 0.75).

2.3.4. Short-Form Health Survey (SF-12)

The SF-12 is an abbreviated version of the 36-item "Short-Form Health Survey", containing two subscales measuring physical and mental aspects of health-related quality of life (Ware et al., 1996). The SF-12 is a self-report measure rating the presence and severity of physical and psychological impairment over the course of the last four weeks. The subscale scores range from 0 to 100, where higher scores indicate higher levels of health. The SF-12 was administered at baseline and seven follow-ups: 3-months, 6-months, 12-months, 18-months, 24-months, 30-months, 36-months.
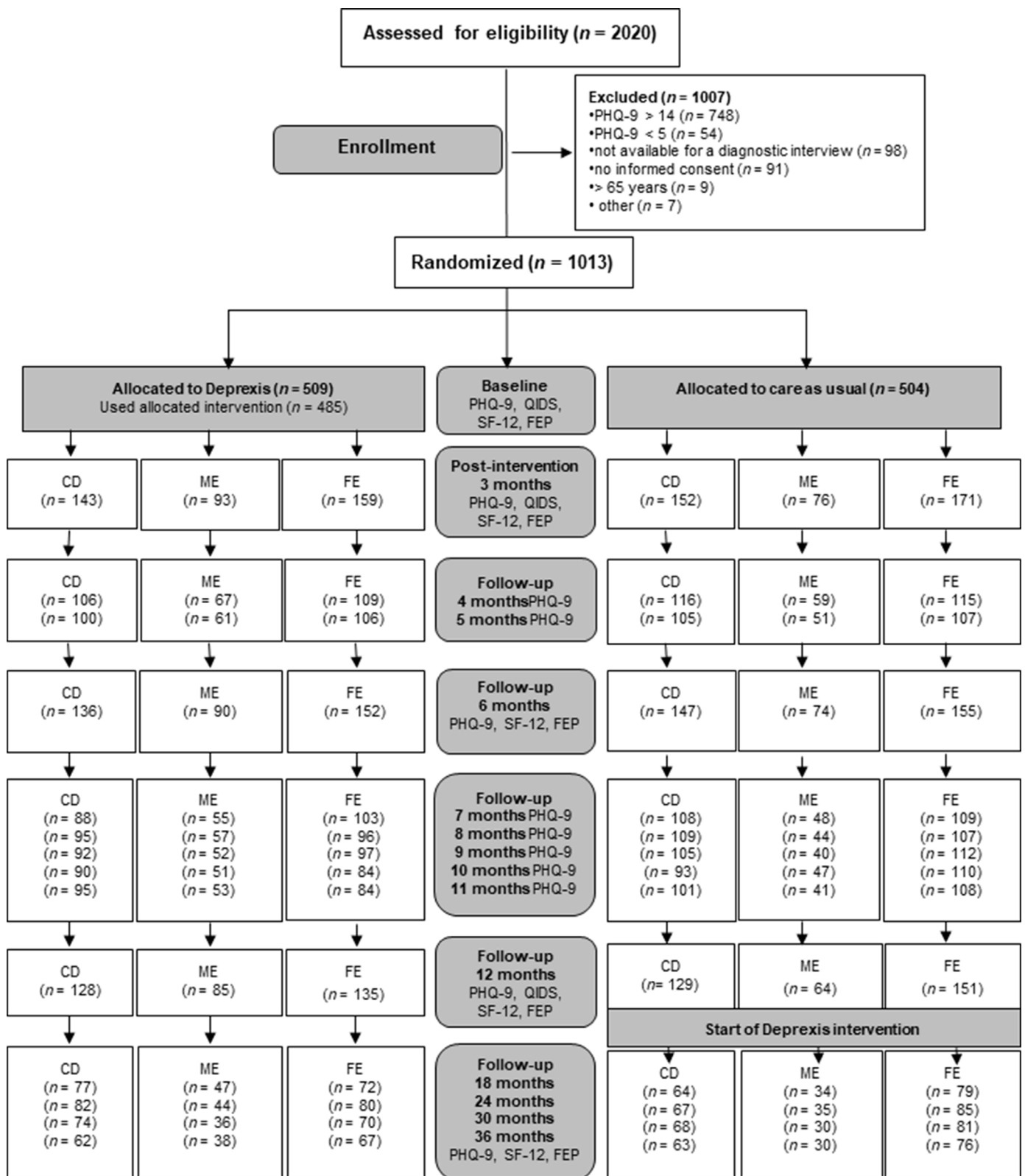
Fig. 1. Flow Diagram of Study. **Deprexis**: internet intervention for depression. **CD**: chronic forms of depression. **ME**: more than five depressive episodes. **FE**: up to five depressive episodes. **PHQ-9**: Patient Health Questionnaire-9. **QIDS**: Quick Inventory of Depressive Symptomatology. **SF-12**: Short-Form Health Survey. **FEP**: Questionnaire for the Evaluation of Psychotherapeutic Progress-2.

2.3.5. Questionnaire for the Evaluation of Psychotherapeutic Progress-2 (FEP-2)

The FEP-2 was developed as a measure of therapeutic progress, which has been shown to be change-sensitive as well as reliable (Lutz et al., 2009; Lutz and Böhnke, 2008). The instrument can also be used as a broad symptom measure covering the following four dimensions: well-being, symptom distress, interpersonal problems and incongruence with respect to approach and avoidance goals. The FEP-2

consists of 40 items that are rated on a 5-point Likert-type scale, with higher scores indicating higher impairment. Reliability is high for the global scale (Cronbach α = 0.96; Retest between rtt = 0.69–0.77) and sensitivity to change has been demonstrated (Lutz et al., 2009). The FEP-2 was administered at baseline and seven follow-ups: 3-months, 6-months, 12-months, 18-months, 24-months, 30-months, 36-months.

2.4. Statistical analyses

Statistical analyses were conducted with SPSS version 25 (Inc, Chicago, IL, USA). All tests were performed two-tailed and the significance value was set to $p < 0.05$. To evaluate differences in sociodemographic characteristics, univariate ANOVAs and chi-square-tests were conducted. Bonferroni corrections were applied for the pairwise-post-hoc tests. To assess whether the three depression groups (CD, ME, FE) differ at baseline and over time in the outcome variables (depressive symptoms, quality of life, well-being, symptom distress, interpersonal problems, and incongruence), linear multilevel models were conducted. One multilevel model was conducted for each outcome variable (PHQ-9, QIDS, SF-12, and FEP-2 scales), as the number of repeated assessments differed among the dependent variables (i.e., PHQ-9: 15 time points; FEP-2, SF-12: 8 time points; QIDS: 3 time points) as described above. All models had two-levels (repeated measures nested within individuals) and were performed with the full maximum likelihood method. Fixed effects were time and depression group (CD, ME, FE with CD being the reference) as well as the interaction between time and depression group. The time variable was calculated for each participant individually as days of the follow-up assessment after baseline assessment (coded as 0), since the timing of the follow-up assessments varied between participants. As random term, the random intercept was included, the random slope was not added as the models did not converge with a random intercept and a random slope.

We also conducted a time to event analysis to evaluate differences between the depression groups (CD, ME, FE with CD being the reference) in time to remission from depressive symptoms as measured with the PHQ-9 (the first time a PHQ-9 score < 5 was reached was defined as the onset of remission). Participants who did not achieve remission within 36 months or dropped out were censored. No missing data were substituted in the analysis. The *Kaplan-Meier method* was used to estimate time to remission (in days) and differences among the three depression groups were compared using the log-rank, Breslow and Taron-Ware tests.

The multilevel models and the event analysis were conducted for the total sample as well as for the intervention (Deprexis) and control condition (CAU) separately. This was done to explore whether the results for the total sample can be replicated in both conditions.

3. Results

3.1. Baseline comparisons between the three depression groups

3.1.1. Sociodemographic variables

Significant differences among the three depression groups were observed in age ($F(2, 1012) = 3.65$, $p = 0.026$) and gender ($\chi^2 = 8.18$, $p = 0.017$) but not in marital status and education (Table 1). Post-hoc tests for age showed that the CD group was older than the FE group ($p = 0.021$). Importantly, no age differences emerged between ME and FE ($p = 0.918$) ruling out the possibility that ME participants developed more episodes than FE participants because of their older age.

3.1.2. Outcome variables at baseline

CD vs. ME: For the total sample (Tables 2-4), the intercepts did not differ between the CD and the ME group in depression severity (PHQ-9 and QIDS), the SF-12 mental component, and the FEP-2 scales. Solely for the SF-12 physical component ($p = 0.019$), a significant difference towards stronger impairment in CD was observed at baseline. However,

when the analysis was conducted for the CAU and Deprexis condition separately (Supplemental Tables 1–3), the intercepts of all dependent variables did not differ between CD and ME.

CD vs. FE: The intercepts of all investigated variables differed, with a stronger impairment in CD ($p < 0.001$) for the total sample (Tables 2-4). This was also the case when running the analysis separately for the Deprexis and the CAU condition (Supplemental Tables 1–3).

3.2. Comparisons between the three depression groups in trajectories of outcome variables over the 3-year observation period

The time courses of the primary outcome, as well as the secondary outcome variables, are reflected by the estimates for the fixed effect time (slope) summarized in Tables 2-4 for the total sample. For all analyzed dependent variables, no differences were observed between CD and ME as well as between CD and FE.

Separate statistical analyses of the two treatment conditions (Supplemental Table 1–3) were in accordance with the analysis of the total sample, except for differences in the control group sample between CD and FE in the FEP-2 scales interpersonal problems ($p = 0.036$), incongruence ($p = 0.028$) and the total score ($p = 0.044$). The slopes were more positive in FE compared to CD in these variables and this means less improvement over time for FE compared to CD.

3.3. Time to remission

The three depression groups differed in the estimated time to onset of remission from depressive symptoms (PHQ-9) within the three years observation period after randomization (log rank $\chi^2 = 22.84$, $p < 0.001$; Breslow $\chi^2 = 24.97$, $p < 0.001$, Taron-Ware $\chi^2 = 24.60$, $p < 0.001$). The CD group (estimated $M = 723$ days (± 29.12), lower bound: 666 days, upper bound: 780 days) and the ME group (estimated $M = 767$ days (± 46.84), lower bound: 675 days, upper bound: 859 days) showed an approximately similar time to onset of remission (log rank $\chi^2 = 0.324$, $p = 0.570$; Breslow $\chi^2 = 0.328$, $p < 0.567$, Taron-Ware $\chi^2 = 0.332$, $p < 0.564$), whereas the estimated mean symptom duration in group FE was significantly shorter than in CD ($M = 555$ days (± 27.79), lower bound: 500 days, upper bound: 609 days; log rank $\chi^2 = 10.350$, $p < 0.001$; Breslow $\chi^2 = 11.068$, $p < 0.001$, Taron-Ware $\chi^2 = 11.006$, $p < 0.001$; Fig. 2).

Stratifying the total sample according to condition, differences regarding the duration of the depressive symptoms among the three depression groups remained significant in both conditions (Deprexis: log rank $\chi^2 = 6.21$, $p = 0.045$; Breslow $\chi^2 = 5.83$, $p = 0.054$, Taron-Ware $\chi^2 = 6.21$, $p = 0.045$; CAU: log rank $\chi^2 = 18.46$, $p < 0.001$; Breslow $\chi^2 = 23.46$, $p < 0.001$, Taron-Ware $\chi^2 = 21.46$, $p < 0.001$). Thus, results for the depression groups in the two subsamples concurred with those observed for the total sample (Supplemental Fig. 1–2).

4. Discussion

The current study was conducted to examine whether participants with CD resemble those with ME but differ from those with FE.

At baseline, significant differences emerged between CD and FE in all investigated outcome variables, but only one significant difference was observed between CD and ME (physical quality of life) in the total sample. Regarding the trajectories over three years, CD and ME did not differ in any of the parameters, whereas CD and FE differed in several outcome variables in the CAU condition. Furthermore, time to remission from depressive symptoms was similar in CD and ME but differed between CD and FE.

In summary, we found more differences between CD and FE than between CD and ME, although more than half of CD-participants (53%) had a history of up to five depressive episodes as had the FE group. Yet, we also observed a difference regarding physical aspects of health-related quality of life at baseline between CD and ME. These results

Table 1

Sociodemographic characteristics of the depression groups at baseline. Frequency and percentage.

	CD	ME	FE	Statistics	p
N (%)	376 (37.1%)	215 (21.2%)	422 (41.7%)		
Age, mean (SD)	44.0 (10.7)	42.7 (10.4)	41.9 (11.5)	F (2, 1012) = 3.65	0.026
Gender				χ^2 = 8.18	0.017
Male	111 (29.5%)	55 (25.6%)	152 (36.0%)		
Female	265 (70.5%)	160 (74.4%)	270 (64.0%)		
Marital status				χ^2 = 16.81	0.079
Married	154 (41.0%)	92 (42.8%)	179 (42.4%)		
Married, but living separated	14 (3.7%)	3 (1.4%)	11 (2.6%)		
Single	85 (22.6%)	49 (22.8%)	113 (26.8%)		
In relationship	63 (16.8%)	50 (23.3%)	76 (18.0%)		
Divorced	54 (14.4%)	21 (9.8%)	40 (9.5%)		
Widowed	6 (1.6%)	0 (0%)	3 (0.7%)		
Educational level				χ^2 = 8.64	0.734
No graduation	1 (0.3%)	1 (0.5%)	1 (0.2%)		
Lower secondary	22 (5.9%)	13 (6.0%)	18 (4.3%)		
Middle secondary	95 (25.3%)	52 (24.2%)	96 (22.7%)		
Higher secondary	67 (17.8%)	37 (17.2%)	68 (16.1%)		
Highest secondary	182 (48.4%)	110 (51.2%)	228 (54.0%)		
Other	9 (2.4%)	2 (0.9%)	11 (2.6%)		

Abbreviations: CD: chronic forms of depression. **ME:** more than five depressive episodes. **FE:** up to five depressive episodes. **N:** sample size. **SD:** standard deviation. χ^2: chi square.

Table 2

Estimates of the multilevel model with PHQ-9 and QIDS as dependent variable.

Parameter	Estimate	SE	df	t	p
PHQ-9					
Intercept (CD)	9.44	0.17	1261.87	54.57	< 0.001
ME	−0.24	0.29	1275.35	−0.83	0.408
FE	−1.35	0.24	1276.53	−5.66	< 0.001
Time (CD)	−0.00169	0.00020	7658.35	−8.28	< 0.001
Time x ME	0.00050	0.00035	7677.36	1.44	0.149
Time x FE	−0.00023	0.00028	7674.98	−0.82	0.413
QIDS					
Intercept (CD)	8.94	0.20	1529.51	44.06	< 0.001
ME	−0.22	0.34	1527.49	−0.65	0.518
FE	−1.28	0.28	1538.13	−4.56	< 0.001
Time (CD)	−0.00495	0.00078	1496.45	−6.38	< 0.001
Time x ME	−0.00146	0.00129	1500.00	−1.13	0.258
Time x FE	−0.00194	0.00107	1504.24	−1.81	0.071

Abbreviations: QIDS: Quick Inventory of Depressive Symptomatology. **PHQ-9:** Patient Health Questionnaire-9. **SE:** standard error. **CD:** chronic forms of depression. **ME:** more than five depressive episodes. **FE:** up to five depressive episodes.

Table 3

Estimates of the multilevel model with SF-12 as dependent variable.

Parameter	Estimate	SE	df	t	P
Physical component					
Intercept (CD)	45.92	0.43	1236.98	105.95	< 0.001
ME	1.69	0.72	1235.27	2.35	0.019
FE	3.06	0.60	1237.51	5.15	< 0.001
Time (CD)	−0.00027	0.00046	3875.94	−0.58	0.562
Time x ME	−0.00119	0.00079	3886.23	−1.51	0.130
Time x FE	−0.00010	0.00064	3873.44	−0.16	0.874
Mental component					
Intercept (CD)	33.01	0.45	1499.27	73.40	< 0.001
ME	1.19	0.74	1498.10	1.59	0.111
FE	3.29	0.62	1497.85	5.33	< 0.001
Time (CD)	0.00702	0.00061	4106.88	11.50	< 0.001
Time x ME	−0.00128	0.00104	4126.60	−1.24	0.216
Time x FE	0.00082	0.00084	4103.22	0.97	0.333

Abbreviations: SF-12: Short-Form Health Survey. **SE:** standard error. **CD:** chronic forms of depression. **ME:** more than five depressive episodes. **FE:** up to five depressive episodes.

Table 4

Estimates of the multilevel model with FEP-2 as dependent variable.

Parameter	Estimate	SE	df	t	p
Well-being					
Intercept (CD)	3.43	0.03	1461.96	102.82	< 0.001
ME	−0.09	0.06	1462.63	−1.56	0.119
FE	−0.32	0.05	1465.64	−6.89	< 0.001
Time (CD)	−0.00038	0.00004	4138.32	−8.62	< 0.001
Time x ME	0.00011	0.00007	4155.17	1.47	0.142
Time x FE	−0.00003	0.00006	4141.49	−0.51	0.611
Complaints					
Intercept (CD)	2.80	0.03	1395.67	91.90	< 0.001
ME	−0.08	0.05	1394.49	−1.53	0.129
FE	−0.25	0.04	1399.23	−6.01	< 0.001
Time (CD)	−0.00036	0.00004	4088.43	−9.35	<0.001
Time x ME	0.00009	0.00007	4102.03	1.33	0.184
Time x FE	0.000001	0.00005	4091.46	0.01	0.990
Interpersonal problems					
Intercept (CD)	2.57	0.03	1223.82	87.59	< 0.001
ME	−0.09	0.05	1222.64	−1.82	0.069
FE	−0.20	0.04	1226.74	−4.99	< 0.001
Time (CD)	−0.00021	0.00003	3922.78	−7.06	< 0.001
Time x ME	0.00002	0.00005	3930.18	0.34	0.735
Time x FE	0.00003	0.00004	3925.18	0.73	0.463
Incongruence					
Intercept (CD)	3.03	0.03	1345.33	99.20	< 0.001
ME	−0.05	0.05	1345.67	−1.04	0.301
FE	−0.27	0.04	1348.76	−6.34	< 0.001
Time (CD)	−0.00032	0.00004	4044.45	−8.77	< 0.001
Time x ME	0.00003	0.00006	4057.52	0.53	0.593
Time x FE	0.00001	0.00005	4047.33	0.26	0.797
Total					
Intercept (CD)	2.90	0.03	1349.09	107.40	< 0.001
ME	−0.08	0.04	1349.40	−1.68	0.093
FE	−0.25	0.04	1352.51	−6.77	< 0.001
Time (CD)	−0.00031	0.00003	4043.13	−9.60	< 0.001
Time x ME	0.00006	0.00005	4055.76	1.03	0.301
Time x FE	0.00001	0.00004	4045.96	0.14	0.889

Abbreviations: FEP-2: Questionnaire for the Evaluation of Psychotherapeutic Progress-2. **SE:** standard error. **CD:** chronic forms of depression. **ME:** more than five depressive episodes. **FE:** up to five depressive episodes.

provide partial support for our hypothesis.

Conventionally, several factors are examined when trying to validate a diagnostic category: antecedent validators (e.g. risk factors), concurrent validators (e.g. symptom similarity) and predictive validators (e.g. clinical course and treatment response). Our results provide
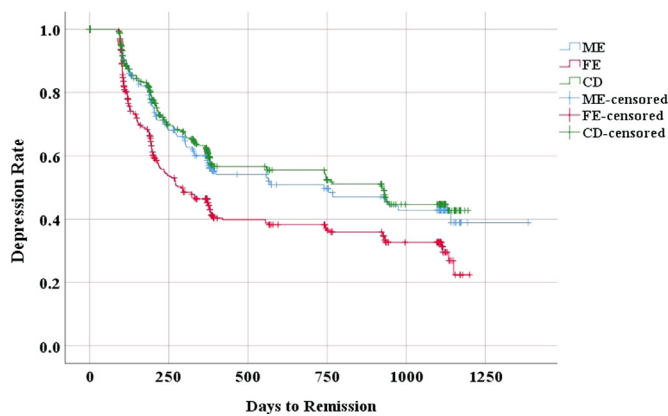
Fig. 2. Kaplan-Meier estimates of the duration of depressive symptoms until first remission by depression group. **CD**: chronic forms of depression. **ME**: more than five depressive episodes. **FE**: up to five depressive episodes.

concurrent (similar but not identical baseline symptoms) and predictive validators (similar three-year course) suggesting that patients with ME and CD belong to the same category. The studies by Bockting et al. (2005) and Ma and Teasdale (2004) provide antecedent evidence (childhood adversity and age of onset), suggesting that the ME and FE categories might be separate. Unfortunately, we were unable to compare participants with CD and ME with respect to these or other antecedent validators. Therefore, our results must be replicated and extended before firm conclusions can be drawn regarding the diagnostic validity and utility of a "frequent episode" category.

Altogether, the current findings emphasize the need for further refinement of the definition of CD, to improve treatments for those who have developed a more protracted course of the disorder but are currently not considered as patients with CD. According to our study an extension of the definition of chronic forms of depression including those patients with a high number of previous depressive episodes, should be considered and evaluated in future studies.

4.1. Strengths and limitations

Several strengths and limitations should be noted. The first strength of this study is the relatively long observation period of three-years, which allowed us to study whether the outcome trajectories differed among the three depression groups. A second strength is the large sample size, which allowed us to create relatively large subsamples with different depression forms. A third strength concerns the fact that participants were recruited from a broad array of clinical and non-clinical settings, which enhances external validity. A fourth strength is the broad generalizability of our findings as the trial's inclusion criteria were broad, allowing any form of concomitant treatment. A fifth strength is that all statistical analyses were not only conducted for the total sample, but also for the intervention and control group separately. These additional analyses largely confirmed that the results for the total sample are replicable in both conditions. A final strength is that the diagnosis of dysthymia and the diagnosis of current depression (MINI) as well as depression severity (QIDS) were ascertained by trained clinical raters.

There are also limitations to consider when interpreting the results. The first and major limitation of the study is that CD was assessed with an interview, whereas the differentiation between ME and FE was based on self-reports. Recall bias regarding the illness course could have led to false classifications, at least in some participants (Andrews et al., 1999). Using the same method to assess CD, ME, and FE would have been ideal. A second limitation is that the differentiation between ME and FE was based on lifetime history of more or up to five episodes. In contrast, other studies used different cut-offs, such as three (Ma and

Teasdale, 2004), four (Bockting et al., 2009), or five (Bockting et al., 2005) previous episodes. To avoid the problem of selecting a cut-off score to dichotomize the sample into those with multiple vs. few prior episodes, continuous scoring of the episodes should be used in future research, which should also apply validated clinical interviews to assess the number of previous depressive episodes. These future studies should also include a detailed assessment of the subtypes of CD, as the chronic forms investigated here included dysthymia and double depression only (the MINI interview evaluated current major depression as well as dysthymia but no other depressive disorders). Therefore, we do not know whether and how many participants met the criteria of persistent major depressive episode, intermittent major depressive episode with current episode, or intermittent major depressive episode without current episode as defined in DSM-5. It might be that the chronic forms of depression evaluated here (dysthymia and double depression) more closely resemble ME than FE, but other forms of CD might not show the same pattern. Another shortcoming is related to the lack of biological samples due to the online study design, e. g., cortisol levels (also in response to stimulus) might help to differentiate the groups. Moreover, there were age and gender differences between the groups. Including gender and age as moderators in further studies would be interesting to see if results are comparable for women and men and across age groups. A further limitation of the study is that generalizability is limited by the mild to moderate symptom severity range in our sample; thus, results might not generalize to patients with severe depression. Moreover, participants had higher levels of education compared to a representative population-based sample with the same range of depressive symptom severity (Späth et al., 2017). Therefore these results might not generalize to all persons with mild to moderate depressive symptoms. A final limitation is that participants were allocated to two conditions, with the timing of access to the intervention representing a confounder (immediate access in Deprexis vs. delayed access in CAU). However, to analyze whether the observed differences for the three depression groups are replicable among both conditions, separate analysis for both groups were performed, which confirmed the findings for the total sample.

4.2. Conclusions

We aimed to examine whether participants with ME and CD resemble each other closely, which might justify their inclusion in the same diagnostic category. Our results provide concurrent (similar but not identical baseline symptoms) and predictive validators (similar three-year course) suggesting that patients with ME and CD belong to the same category. More research is needed to elucidate whether ME and CD should be included in a combined diagnostic category, as these findings tentatively suggest.

CRediT authorship contribution statement

Elke Humer: Formal analysis, Writing - original draft. **Krisztina Kocsis-Bogar:** Formal analysis. **Thomas Berger:** Investigation. **Johanna Schröder:** Investigation. **Christina Späth:** Investigation. **Björn Meyer:** Investigation. **Steffen Moritz:** Investigation. **Wolfgang Lutz:** Validation. **Thomas Probst:** Data curation, Formal analysis, Writing - original draft. **Jan Philipp Klein:** Investigation, Data curation, Writing - original draft.

Declaration of Competing Interests

Acknowledgements

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.psychres.2020.113235.

References

American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders, 5th ed. American Psychiatric Association, Washington, DC.

American Psychiatric Association, 1994. Diagnostic and Statistical Manual of Mental Disorders, 4th ed. American Psychiatric Association, Washington, DC.

Andrews, G., Anstey, K., Brodaty, H., Issakidis, C., Luscombe, G., 1999. Recall of depressive episode 25 years previously. Psychol. Med. 29, 787–791. https://doi.org/10.1017/S0033291799008648.

Bird, T., Tarsia, M., Schwannauer, M., 2018. Interpersonal styles in major and chronic depression: a systematic review and meta-analysis. J. Affect. Disord. 239, 93–101. https://doi.org/10.1016/j.jad.2018.05.057.

Bockting, C.L.H., Spinhoven, P., Wouters, L.F., Koeter, M.W.J., Schene, A.H., the DELTA Study Group, 2009. Long-term effects of preventive cognitive therapy in recurrent depression: a 5.5-year follow-pp study. J. Clin. Psychiatry 70, 1621–1628. https://doi.org/10.4088/JCP.08m04784blu.

Fuhr, K., Schröder, J., Berger, T., Steffen, M., Meyer, Lutz, W., Hohagen, F., Hautzinger, M., Klein, J.P., 2018. The association between adherence and outcome in an internet intervention for depression. J. Affect. Disord. 15 (229), 443–449. https://doi.org/10.1016/j.jad.2017.12.028.

Hoertel, N., 2017. A comprehensive model of predictors of persistence and recurrence in adults with major depression: results from a national 3-year prospective study. J. Psychiatr. Res. 95, 19–27. https://doi.org/10.1016/j.jpsychires.2017.07.022.

Ildirli, S., Sair, Y.B., Dereboy, F., 2015. Persistent depression as a novel diagnostic category: results from the menderes depression study. Noro Psikiyatri Arsivi 52, 359–366. https://doi.org/10.5152/npa.2015.7589.

Jobst, A., Brakemeier, E.-.L., Buchheim, A., Caspar, F., Cuijpers, P., Ebmeier, K.P., Falkai, P., Jan van der Gaag, R., Gaebel, W., Herpertz, S., Kurimay, T., Sabaß, L., Schnell, K., Schramm, E., Torrent, C., Wasserman, D., Wiersma, J., Padberg, F., 2016. European Psychiatric Association Guidance on psychotherapy in chronic depression across Europe. Eur. Psychiatry 33, 18–36. https://doi.org/10.1016/j.eurpsy.2015.12.003.

Klein, J.P., Berger, T., Schröder, J., Späth, C., Meyer, B., Caspar, F., Lutz, W., Arndt, A., Greiner, W., Gräfe, V., Hautzinger, M., Fuhr, K., Rose, M., Nolte, S., Löwe, B., Andersson, G., Vettorazzi, E., Moritz, S., Hohagen, F., 2016. Effects of a psychological internet intervention in the treatment of mild to moderate depressive symptoms: results of the EVIDENT study, a randomized controlled trial. Psychother. Psychosom. 85, 218–228. https://doi.org/10.1159/000445355.

Klein, J.P., Berger, T., Schröder, J., Späth, C., Meyer, B., Caspar, F., Lutz, W., Greiner, W., Hautzinger, M., Rose, M., Gräfe, V., Hohagen, F., Andersson, G., Vettorazzi, E., Moritz, S., 2013. The EVIDENT-trial: protocol and rationale of a multicenter randomized controlled trial testing the effectiveness of an online-based psychological

intervention. BMC Psychiatry 13, 239. https://doi.org/10.1186/1471-244X-13-239.

Klein, J.P., Gamon, C., Späth, C., Berger, T., Meyer, B., Hohagen, F., Hautzinger, M., Lutz, W., Vettorazzi, E., Moritz, S., Schröder, J., 2017a. Does recruitment source moderate treatment effectiveness? a subgroup analysis from the EVIDENT study, a randomised controlled trial of an internet intervention for depressive symptoms. BMJ Open 7, e015391. https://doi.org/10.1136/bmjopen-2016-015391.

Klein, J.P., Roniger, A., Schweiger, U., Späth, C., Brodbeck, J., 2015. The association of childhood trauma and personality disorders with chronic depression: a cross-sectional study in depressed outpatients. J. Clin. Psychiatry 76, e794–e801. https://doi.org/10.4088/JCP.14m09158.

Klein, J.P., Späth, C., Schröder, J., Meyer, B., Greiner, W., Hautzinger, M., Lutz, W., Rose, M., Vettorazzi, E., Andersson, G., Hohagen, F., Moritz, S., Berger, T., 2017b. Time to remission from mild to moderate depressive symptoms: one year results from the EVIDENT-study, an RCT of an internet intervention for depression. Behav. Res. Ther. 97, 154–162. https://doi.org/10.1016/j.brat.2017.07.013.

Köhler, S., Chrysanthou, S., Guhn, A., Sterzer, P., 2019. Differences between chronic and nonchronic depression: systematic review and implications for treatment. Depress. Anxiety 36, 18–30. https://doi.org/10.1002/da.22835.

Kroenke, K., Spitzer, R.L., Williams, J.B.W., 2001. The PHQ-9: validity of a brief depression severity measure. J. Gen. Intern. Med. 16, 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x.

Löwe, B., Kroenke, K., Herzog, W., Gräfe, K., 2004. Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). J. Affect. Disord. 81, 61–66. https://doi.org/10.1016/S0165-0327(03)00198-8.

Lutz, W., Böhnke, J.R., 2008. Der „Fragebogen zur Evaluation von Psychotherapieverläufen" (FEP-2). Trierer Psychologische Berichte 35, 1–21.

Lutz, W., Schürch, E., Stulz, N., Böhnke, J.R., Schöttke, H., Rogner, J., Wiedl, K.H., 2009. Entwicklung und psychometrische Kennwerte des Fragebogens zur Evaluation von Psychotherapieverläufen (FEP). Diagnostica 55, 106–116. https://doi.org/10.1026/0012-1924.55.2.106.

Ma, S.H., Teasdale, J.D., 2004. Mindfulness-based cognitive therapy for depression: replication and exploration of differential relapse prevention effects. J. Consult. Clin. Psychol. 72, 31–40. https://doi.org/10.1037/0022-006X.72.1.31.

Meyer, B., Berger, T., Caspar, F., Beevers, C.G., Andersson, G., Weiss, M., 2009. Effectiveness of a novel integrative online treatment for depression (Deprexis): randomized controlled trial. J. Med. Internet Res. 11, e15. https://doi.org/10.2196/jmir.1151.

Murphy, J.A., Byrne, G.J., 2012. Prevalence and correlates of the proposed DSM-5 diagnosis of Chronic Depressive Disorder. J. Affect. Disord. 139, 172–180. https://doi.org/10.1016/j.jad.2012.01.033.

Nelson, J., Klumparendt, A., Doebler, P., Ehring, T., 2017. Childhood maltreatment and characteristics of adult depression: meta-analysis. Br. J. Psychiatry 210, 96–104. https://doi.org/10.1192/bjp.bp.115.180752.

Nierenberg, A.A., DeCecco, L.M., 2001. Definitions of antidepressant treatment response, remission, nonresponse, partial response, and other relevant outcomes: a focus on treatment-resistant depression. J. Clin. Psychiatry 62 (Suppl. 16), 5–9.

Probst, T., Berger, T., Meyer, B., Späth, C., Schröder, J., Hohagen, F., Moritz, S., Klein, J.P., 2020. Social phobia moderates the outcome in the EVIDENT-study: a randomized controlled trial on internet-based psychological intervention for mild to moderate depressive symptoms. J. Consult. Clin. Psychol. 88, 82–89. https://doi.org/10.1037/ccp0000441.

Rush, A.J., Trivedi, M.H., Ibrahim, H.M., Carmody, T.J., Arnow, B., Klein, D.N., Markowitz, J.C., Ninan, P.T., Kornstein, S., Manber, R., Thase, M.E., Kocsis, J.H., Keller, M.B., 2003. The 16-Item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. Biol. Psychiatry 54, 573–583. https://doi.org/10.1016/S0006-3223(02)01866-8.

Sheehan, D.V., Lecrubier, Y., Sheehan, K.H., Amorim, P., Janavas, J., Weiller, E., Hergueta, T., Baker, R., Dunbar, G.C., 1998. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and 7 validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J. Clin. Psychiatry 59 (Suppl. 20), 34–57.

Späth, C., Hapke, U., Maske, U., Schröder, J., Moritz, S., Berger, T., Meyer, B., Rose, M., Nolte, S., Klein, J.P., 2017. Characteristics of participants in a randomized trial of an Internet intervention for depression (EVIDENT) in comparison to a national sample (DEGS1). Internet Interv. 9, 46–50. https://doi.org/10.1016/j.invent.2017.05.003.

ten Have, M., de Graaf, R., van Dorsselaer, S., Tuithof, M., Kleinjan, M., Penninx, B.W.J.H., 2018. Recurrence and chronicity of major depressive disorder and their risk indicators in a population cohort. Acta Psychiatr. Scand. 137, 503–515. https://doi.org/10.1111/acps.12874.

Ware, J.E., Kosinski, M., Keller, S.D., 1996. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med. Care 34, 220–233. https://doi.org/10.1097/00005650-199603000-00003.

Wittchen, H.U., Jacobi, F., Rehm, J., Gustavsson, A., Svensson, M., Jönsson, B., Olesen, J., Allgulander, C., Alonso, J., Faravelli, C., Fratiglioni, L., Jennum, P., Lieb, R., Maercker, A., van Os, J., Preisig, M., Salvador-Carulla, L., Simon, R., Steinhausen, H.-.C., 2011. The size and burden of mental disorders and other disorders of the brain in Europe 2010. Eur. Neuropsychopharmacol. 21, 655–679. https://doi.org/10.1016/j.euroneuro.2011.07.018.

World Health Organization, 2010. International Classification of Diseases, 10th ed. World Health Organization, Malta.