Expert and self-assessment of lifetime symptoms and diagnosis of major depressive disorder in large-scale genetic studies in the general population: comparison of a clinical interview and a self-administered checklist

Jessica Martin[a,*], Fabian Streit[a,*], Jens Treutlein[a], Maren Lang[a], Josef Frank[a], Andreas J. Forstner[b,c], Franziska Degenhardt[b,c], Stephanie H. Witt[a], Thomas G. Schulze[d,e], Sven Cichon[b,c,f,g], Markus M. Nöthen[b,c], Marcella Rietschel[a] and Jana Strohmaier[a]

Major depression disorder (MDD) is a complex neuropsychiatric disorder and an increasing number of genetic risk variants are being identified. Investigation of their influence in the general population requires accurate and efficient assessment of depressive symptoms. Here, clinical interviews conducted by clinicians are the gold standard. We investigated whether valid and reliable clinical phenotypes can be obtained efficiently using self-administered instruments. Lifetime depressive symptoms and lifetime MDD diagnosis were assessed in 464 population-based individuals using a clinical interview and a structured, self-administered checklist. Analyses were carried out of the following: (i) intraclass correlations (ICC) between checklist and interview; (ii) sensitivity/specificity of the checklist; and (iii) the association of interview and checklist with a positive family history of MDD (FH-MDD+). The correspondence of the self-administered checklist with the clinical interview was good for most depressive symptoms (ICC = 0.60–0.80) and moderate for MDD diagnosis (ICC = 0.45). With the consecutive inclusion of MDD diagnostic criteria, sensitivity decreased from 0.67 to 0.46, whereas specificity remained high (0.95). For checklist and interview, strong associations were found between FH-MDD+ and most depressive symptoms and MDD diagnosis (all odds ratio ≥ 1.83). The self-administered checklist showed high reliability for both the assessment of lifetime depressive symptoms and screening for individuals with no lifetime diagnosis of MDD. However, attention is warranted when the aim is to identify MDD cases. The positive association between depressive symptomatology and FH-MDD+ indicates the usefulness of both instruments to assess patients in genetic studies. Our data suggest that the more time-efficient and cost-efficient self-administered instruments also allow for the assessment of depressive symptoms accurate enough to investigate the influence of MDD genetic risk variants in the general population. *Psychiatr Genet* **27:187–196 Copyright © 2017 The Author(s). Published by Wolters Kluwer Health, Inc.**

Psychiatric Genetics 2017, **27:**187–196

Keywords: general population, genome-wide association study, major depression, self-administered assessment

[a]Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Mannheim, [b]Institute of Human Genetics, [c]Department of Genomics, Life and Brain Center, University of Bonn, Bonn, [d]Institute of Psychiatric Phenomics and Genomics, Ludwig-Maximilians-University, Munich, [e]Department of Psychiatry and Psychotherapy, University of Göttingen, Göttingen, [f]Institute of Neuroscience and Medicine (INM-1), Research Centre Jülich, Jülich, Germany and [g]Department of Biomedicine, Division of Medical Genetics, University of Basel, Basel, Switzerland

Correspondence to Jana Strohmaier, PhD, Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, J5, 68159 Mannheim, Germany
Tel: +49 621 1703 6067; fax: +49 621 1703 6065;
e-mail: jana.strohmaier@zi-mannheim.de

*Jessica Martin and Fabian Streit contributed equally to the writing of this article.

Received 16 September 2016 **Revised** 28 March 2017 **Accepted** 29 June 2017

Introduction

Major depressive disorder (MDD) is one of the most common psychiatric disorders worldwide and has a lifetime prevalence of ~15% (Kessler *et al.*, 2003; Bromet *et al.*, 2011; Wittchen *et al.*, 2011). The prevalence of MDD is twice as high in women as in men (Kessler *et al.* 1993; Weissman *et al.*, 1993; Bebbington, 1998). MDD is associated with increased morbidity (Kessler *et al.*, 2003; Hasin *et al.*, 2005; Wittchen *et al.*, 2011) and mortality (Angst *et al.*, 2002; Cuijpers and Smit, 2002), and results in considerable individual and societal costs (Luppa *et al.*, 2007; Insel, 2008). World Health Organization (2008) projections indicate that by 2030, MDD will be the leading cause of the global burden of disease.

The estimated heritability of MDD is ~40% (Sullivan *et al.*, 2000) and the identification of genetic risk factors

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (*www.psychgenetics.com*).

therefore represents a promising approach to the elucidation of MDD etiology. For psychiatric disorders such as schizophrenia and bipolar affective disorder with heritability estimates of ~80% (Cardno *et al.*, 1999; Sullivan *et al.*, 2003; Craddock and Sklar, 2009), recent genome-wide association studies (GWAS) have identified numerous genetic risk variants (Sklar *et al.*, 2011; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). For MDD, initial large GWAS and meta-analyses (Sullivan *et al.*, 2009; Muglia *et al.*, 2010; Rietschel *et al.*, 2010; Kohli *et al.*, 2011; Shi *et al.*, 2011; Shyn *et al.*, 2011; Wray *et al.*, 2012; Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium *et al.*, 2013) have failed to detect replicable genome-wide significant risk variants. Only recently, with an increase in sample sizes, this is now changing. An increasing number of contributing genetic risk variants have been identified in a considerably larger study (>120 000 cases and > 330 000 controls) analyzing data collected through 23andMe (Hyde *et al.*, 2016). Further, contributing genetic risk variants were identified for a severe subtype of MDD (CONVERGE consortium, 2015; Cai *et al.*, 2017).

To investigate the influence of identified genetic risk variants, genes, and pathways for MDD in the general population, the valid, reliable, and efficient assessment of depressive symptomatology is crucial. The gold standard for both the assessment of depressive symptoms and the assignment of an MDD diagnosis is the performance of a clinical interview, such as the internationally recognized Structured Clinical Interview for *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed. (DSM-IV), Axis I Disorders (SCID-I; First *et al.*, 1996; Wittchen *et al.*, 1997). However, these interviews are time-intensive and cost-intensive, and the interviewers require extensive training. For the purposes of large-scale GWAS, self-ratings may represent a more efficient method of assessing depressive symptomatology in the general population.

Research has shown that in the primary care setting, self-rating questionnaires are reliable tools for both the assessment of depressive symptoms and MDD screening (Mulrow *et al.*, 1995; Williams *et al.*, 2002). The newly developed Cross Cutting Symptom Measure of the DSM-V (American Psychiatric Association, 2013) allows the participant to self-rate selected depressive symptoms (i.e. A1 and A2, Table 1) and has yielded high test–retest reliabilities (Narrow *et al.*, 2013). These established screening instruments are designed primarily to detect the presence of depressive symptoms within a defined time period (e.g. within the preceding 2 weeks). However, in the context of genetic analyses, assessment of their lifetime occurrence is required.

The aim of the present study was to investigate whether clinical interviews can be replaced by self-administered instruments for the purposes of genetic studies in the general population. To test this, the reliability and validity of a self-administered checklist for MDD were investigated by comparison with a SCID-I interview conducted by clinicians. The self-administered checklist assesses lifetime depressive symptoms and lifetime diagnosis of MDD according to DSM-IV criteria (American Psychiatric Association, 2000). First, we investigated whether the symptoms and diagnosis assessed with the self-administered checklist were in accordance with the results of the SCID-I interview. Analyses were carried out of intraclass correlations (ICC) between checklist and interview data and the sensitivity and specificity of the checklist under the assumption that the interview elicited true lifetime depressive symptoms and MDD diagnosis. Second, formal genetic (Sullivan *et al.*, 2000; Flint and Kendler, 2014) and epidemiological (Kendler *et al.*, 1997; Lieb *et al.*, 2002) studies have shown that first-degree relatives of MDD patients have an increased MDD risk. Therefore, we investigated the usefulness of both instruments to assess patients in genetic studies by analyzing their association with the patients' family history of MDD.

Table 1 Description of single depressive symptoms A1–A9 and major depressive episode criteria A–E in accordance with DSM-IV

Items	Description
A1	Depressed mood
A2	Decreased interest or pleasure
A3	Significant weight change or change in appetite
A4	Change in sleep
A5	Change in activity
A6	Fatigue or loss of energy
A7	Feelings of worthlessness or guilt
A8	Problems with concentration or decision making
A9	Suicidal ideation
Criterion A	Depressive symptomatology: ≥ 1 symptom of A1 or A2, and ≥ 5 symptoms of A1 to A9
Criterion B	Exclusion of bipolar disorder
Criterion C	Clinically significant distress/impairment
Criterion D	Exclusion of substance or a general medical condition as a cause
Criterion E	Exclusion of bereavement as a cause

Fulfillment of MDE criteria A–E corresponds to a lifetime diagnosis of MDD.
DSM-IV, *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed.; MDD, major depressive disorder; MDE, major depressive episode.

Participants and methods

Study sample

Participants for the present study were drawn from the population-based cohort of the German National Genome Research Network (NGFN; Hoefgen *et al.*, 2005). The NGFN cohort comprises 1199 participants from the North Rhine-Westphalia region of Germany, who were identified from the national register and recruited between 2002 and 2003. A total of 464 of these participants (220 women, 244 men, $M = 47.38$ years, $SD = 14.91$; sample characteristics presented in Table 2) underwent face-to-face clinical interviews and were thus included in the present analyses. Written informed consent was obtained from all participants before inclusion. The study was carried out in accordance with the

Table 2 Descriptive characteristics of the sample used for the comparison of checklist and interview data

	Checklist and interview
N	464
Sex (%)	
Women	47.41
Men	52.59
Age [mean (SD)]	
Complete sample	47.38 (14.91)
Men	48.35 (14.71)
Women	46.30 (15.09)
Ethnicity[a] (%)	
German	95.04
Other European	2.37
Non-European	1.29
Unknown	1.29

[a]Ethnicity was determined according to the ethnicity of the grandparents. Participants were classified as German or European if at least three grandparents were reported to be of the respective origin.

Declaration of Helsinki and was approved by the local ethics commission.

Measures

All 464 participants completed a structured self-administered checklist. This is an adaptation of the major depressive episode (MDE) section of the International Diagnostic Checklist for DSM-IV (IDCL DSM-IV; Hiller *et al.*, 1997; Supplementary Appendix, Supplemental digital content 1, *http://links.lww.com/PG/A193*) implemented by our group. The main adaption included (i) lifetime assessment; (ii) phrasing of the criterion description as a question; (iii) additional questions for certain criteria (e.g. to A4 'Sleeplessness or increased sleep' questions asking about initial insomnia, middle insomnia, early morning wakening, and excessive sleep were added); and (iv) deletion of items assessing the differential diagnoses bipolar disorder and other psychotic disorders. In the Supplementary Appendix (Supplemental digital content 1, *http://links.lww.com/PG/A193*) the adapted checklist and an example criterion of the original IDCL DSM-IV are presented for comparison. Completion of this paper–pencil checklist required ~10 min. Further, all participants participated in the SCID-I interview (German version; First *et al.*, 1996; Wittchen *et al.*, 1997) administered by two trained clinicians. The SCID-I interview was administered within 1 week after the completion of the checklist. Interviewers were blinded to the checklist results.

Both the checklist and the interview assessed lifetime depressive symptomatology (DSM-IV criterion A) during at least one lifetime depressive episode (lasting ≥ 2 weeks). Furthermore, both instruments assessed MDE criteria C, D, and E, and the interview additionally assessed MDE criteria B. The nine single depressive symptoms A1–A9 and MDE criteria A–E are presented in Table 1.

In addition, all participants completed a self-rating questionnaire assessing demographic characteristics (including sex, age, years of education, and highest degree of graduation) and personal and family history of neurological and psychiatric disorder. Further details are provided in the Supplementary Text (Supplemental digital content 1, *http://links.lww.com/PG/A193*). Increased familial risk of MDD was defined as having a positive family history of MDD in first-degree or second-degree relatives (FH-MDD+).

Statistical analyses

All data analyses were carried out using IBM SPSS Statistics 22.0 (SPSS Inc., Chicago, Illinois, USA, 2013, *http://www.spss.com*). For the checklist and the interview, depressive symptom count was calculated as the sum score of the nine single depressive symptoms A1–A9. Participants for whom data on one or more depressive symptoms were missing were excluded from further analysis (checklist: 18/464; interview: 16/464). The data distribution was positively skewed, with the majority of participants reporting no depressive symptoms (checklist: 287/446; interview: 310/448). Therefore, for the depressive symptom count, separate comparisons were performed for the complete sample and for the subgroup of participants who reported at least one depressive symptom in both the checklist and the interview ($n = 119$).

For the checklist and the interview, a lifetime diagnosis of MDD was assigned on the basis of DSM-IV MDE criteria (Table 1). Participants for whom data on more than two depressive symptoms were missing or for whom fulfillment of criterion A could not be established were excluded from the analyses (checklist: 8/464; interview: 5/464). To determine whether a lifetime diagnosis of MDD could be assigned, fulfillment of MDE criteria A, C, D, and E (ACDE) was evaluated. Participants assigned a diagnosis of bipolar disorder with the SCID-I interview (DSM-IV criterion B) were excluded from all analyses. The fulfillment of MDE criterion A as well as the fulfillment of MDE criteria A and C (AC) and A, C, and D (ACD) were ascertained separately. In cases where data on one of the A–E criteria were missing, all consecutive criteria were rated as missing. For the depressive symptom count, a comparison between checklist and interview was performed using (i) the dependent-sample Wilcoxon singed-rank test in the complete sample and (ii) the dependent-sample *t*-test in the subgroup of participants with at least one depressive symptom. For lifetime diagnosis of MDD, a comparison between checklist and interview was performed using the χ^2-test. The χ^2-test was also used to compare participants reporting no depressive symptoms with participants reporting at least one depressive symptom. To analyze correlations between checklist and interview for single depressive symptoms A1–A9 and MDE criteria A, AC, ACD, and ACDE, ICC (two-way mixed model of absolute agreement) were used. For dichotomous variables, the ICC corresponds to Cohen's κ-coefficients (Wirtz and Caspar, 2002). The κ-coefficients were interpreted according to Fleiss and Cohen (1973), that is 'very good': ICC above 0.75; 'good': ICC between 0.60 and 0.75; and 'moderate': ICC between 0.40 and 0.60.

All of the above analyses were carried out for the complete sample, and then separately for men and women.

Under the assumption that the interview elicited the true depressive symptomatology and lifetime diagnosis of MDD, the sensitivity and specificity of the checklist were determined in relation to the interview results. In accordance with other studies comparing questionnaire data with clinical interviews, sensitivity and specificity estimates higher than 0.80 were interpreted as 'high' and values ~0.50 as 'moderate'/'low' (Gräfe *et al.*, 2004; Kroenke *et al.*, 2010). The χ^2-test was used to investigate the association between FH-MDD+ and single depressive symptoms A1–A9, depressive symptom count, and MDE criteria A, AC, ACD, and ACDE from checklist and interview. For this purpose, the depressive symptom count was dichotomized at the following five thresholds: at least one, at least two, at least three, at least four, and at least five depressive symptoms.

Furthermore, to explore whether more heritable single depressive symptoms – as observed in a twin study by Kendler *et al.* (2013) – show a stronger association with FH-MDD+, the strength of association between single depressive symptoms and FH-MDD+ for checklist and interview was compared with their respective heritability estimates (h^2). For this purpose, the log-transformed odds ratios (ORs) were correlated with the respective h^2.

The effects of the properties of the checklist on the necessary sample sizes to detect specific genetic effects were explored using the Genetic Power Calculator (*http://zzz.bwh.harvard.edu/gpc/cc2.html*; Purcell *et al.*, 2003). Power analyses were carried out for criteria ACDE, that is lifetime diagnosis of MDD, and criterion A. Calculations were carried out on the basis of the assumption of a multiplicative model and a risk allele frequency of 40%. For the interview, the respective observed MDD frequency was used as prevalence and a genotype relative risk of 1.15 was assumed. For the checklist, the observed sensitivity and specificity were used to calculate the corresponding allele distribution and respective genotype relative risk.

Results

Descriptive statistics for depressive symptoms and major depressive episode criteria

Depression characteristics are shown in Table 3.

Supplementary Fig. 1a and b (Supplemental digital content 1, *http://links.lww.com/PG/A193*) shows the relative frequencies of (i) participants who reported a lifetime history of single depressive symptoms A1–A9; (ii) participants who reported a lifetime history of at least one depressive symptom; and (iii) participants who reported a lifetime history of fulfillment of MDE criteria A, AC, ACD, and ACDE. These data are shown for the complete sample and separately for men and women. Further details of these analyses are provided in the Supplementary Text (Supplemental digital content 1, *http://links.lww.com/PG/A193*).

Table 3 Depression characteristics according to checklist and interview

	Checklist	Interview
Lifetime diagnosis of MDD [*n* (%)][a]		
Complete sample	11.11	14.44
Men	7.14	8.40
Women	15.57	21.23
Depressive symptom count [mean (SD)]		
Complete sample	1.90 (2.87)	1.45 (2.54)
Men	1.53 (2.64)	1.11 (2.26)
Women	2.34 (3.07)	1.85 (2.78)
Depressive symptom count ≥ 1 [mean (SD)]		
Complete sample	5.70 (1.93)	5.13 (2.09)
Men	5.26 (2.05)	4.72 (2.28)
Women	6.06 (1.76)	5.46 (1.87)

MDD, major depressive disorder; *n*, number of participants.
[a]*n* differs between the diagnosis of MDD and symptom severity variables, depending on the respective inclusion criteria.

Intraclass correlations between checklist and interview

ICC coefficients for single depressive symptoms A1–A9, depressive symptom count, depressive symptom count of participants reporting at least one depressive symptom, and MDE criteria A, AC, ACD, and ACDE are shown in (i) Fig. 1a, for the complete sample and (ii) Fig. 1b for men and women separately. Further details of the analyses are provided in Supplementary Table 1 (Supplemental digital content 1, *http://links.lww.com/PG/A193*).

For all the assessed items and criteria, highly significant ICCs were observed between checklist and interview (all $P < 0.001$). For single depressive symptoms A1–A9, the mean ICC in the complete sample was good (0.68). ICCs were very good for A1 and A9 (ICC > 0.75); good for A2, A3, A4, A6, A7, and A8 (ICC: 0.60–0.75); and moderate for A5 (ICC: 0.40–0.60). ICCs were the highest for A1 and the lowest for A5. This was found in the complete sample (0.80 vs. 0.52), and in women (0.79 vs. 0.56) and men (0.80 vs. 0.44).

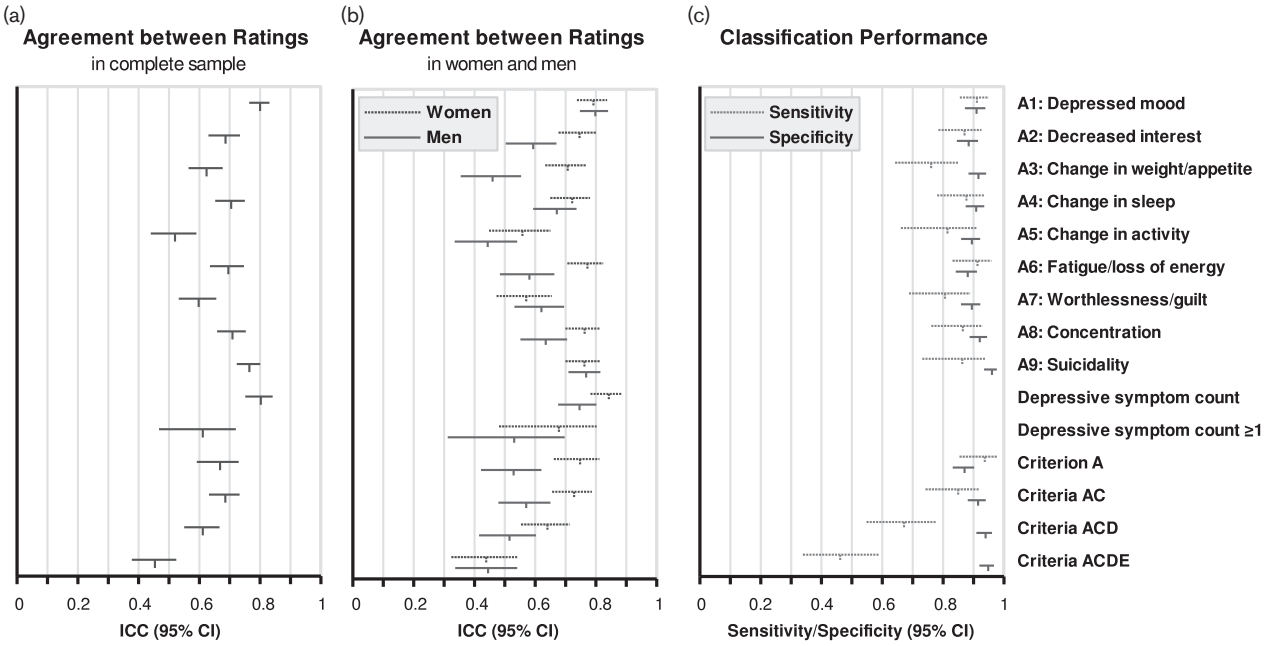
For the depressive symptom count, the ICC was very good in both the complete sample (0.80) and in women (0.84), whereas in men it was good (0.75). In participants reporting at least one depressive symptom, the ICC was good in both the complete sample (0.61) and in women (0.68), but only moderate in men (0.53).

For MDE criteria, the ICC was good for criterion A (0.67), criteria AC (0.69), and criteria ACD (0.61). For criteria ACDE, that is lifetime diagnosis of MDD, the ICC was moderate (0.45). The ICC for criteria A, AC, and ACD were higher in women than in men (0.75, 0.73, 0.64 vs. 0.53, 0.57, 0.52, respectively). For criteria ACDE, ICCs were moderate in both women (0.44) and men (0.45).

Sensitivity and specificity

Under the assumption that the interview elicited the true depressive symptomatology and lifetime diagnosis of MDD, the sensitivity and specificity of the checklist were high for single depressive symptoms A1–A9. Here,

Fig. 1



Agreement between ratings [intraclass correlation coefficients (ICC)]: (a) in the complete sample and (b) for men and women. (c) Graph shows the sensitivity and specificity for DSM-IV single depressive symptoms A1–A9, depressive symptom count, depressive symptom count of participants reporting at least one depressive symptom, and major depressive episode criteria A, AC, ACD, and ACDE, that is a lifetime diagnosis of major depressive disorder. Confidence intervals (CI) of 95% are shown. DSM-IV, *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed.

sensitivity ranged between 0.76 (A3) and 0.91 (A6), and specificity ranged between 0.88 (A6) and 0.96 (A9). For MDE criterion A, both sensitivity (0.94) and specificity (0.87) were high. Although sensitivity decreased with the consecutive inclusion of criteria AC, ACD, and ACDE (0.85, 0.67, 0.46), specificity remained high (0.92, 0.94, 0.95). The sensitivity and specificity of the checklist and interview are presented in Fig. 1c. Further details of the analysis are provided in Supplementary Table 2 (Supplemental digital content 1, *http://links.lww.com/PG/A193*).

Additional analyses of the agreement between checklist and interview for the lifetime diagnosis of MDD (i.e. concordance rate of checklist and interview; relative frequency of true/false-positive/negative findings) are provided in the Supplementary Text and Supplementary Fig. 2 (Supplemental digital content 1, *http://links.lww.com/PG/A193*).

Association with the familial risk of MDD: FH-MDD+

Among participants assigned a lifetime diagnosis of MDD by the checklist, 42% had a FH-MDD+, compared with 15% of participants with no lifetime diagnosis of MDD. For the interview, the percentages were 38 and 14%, respectively.

For the checklist, the association with FH-MDD+ was very strong for A1, A2, and A6–A9 (ORs: 2.24–3.01); strong for A4 and A5 (ORs = 1.94–2.11; all $P < 0.05$); and nonsignificant for A3 (OR = 1.37; $P = 0.30$). For the interview, all single depressive symptoms A1 to A9 yielded very strong associations with FH-MDD+ (ORs: 2.24–4.35; all $P < 0.05$).

For all five thresholds, the dichotomized depressive symptom count was significantly associated with FH-MDD+ for both the checklist (ORs: 1.83–2.10) and the interview (ORs: 2.51–2.97; all $P < 0.05$).

Descriptively, the single depressive symptoms A1–A9 previously found to be more heritable by Kendler *et al.* (2013) were more strongly associated with FH-MDD+ (higher ORs) for both the checklist and the interview. The correlation between the strength of association with FH-MDD+, that is the ORs, and h^2 of the single depressive symptoms A1–A9, reached significance for the checklist ($r = 0.72$, $P = 0.029$; interview: $r = 0.44$, $P = 0.24$). More details can be found in Supplementary Fig. 3a–c (Supplemental digital content 1, *http://links.lww.com/PG/A193*).

From criterion A to criteria AC, ACD, and ACDE, that is lifetime diagnosis of MDD, the association between checklist and interview and FH-MDD+ consecutively increased. The increase was more pronounced for the checklist (2.18, 2.43, 3.51, 4.13) than for the interview (2.77, 2.99, 3.22, 3.69; all $P < 0.05$). Figure 2 shows the ORs for checklist and interview. Additional information is provided in Supplementary Table 3 (Supplemental digital content 1, *http://links.lww.com/PG/A193*).

FH-MDD+ was more prevalent in participants assigned a lifetime diagnosis of MDD by both instruments (52%; OR = 6.78) than in participants assigned a lifetime diagnosis of MDD only by one instrument (checklist: 26%; OR = 2.28; interview: 27%; OR = 2.26). The lowest prevalence of FH-MDD+ was found in participants for whom no lifetime diagnosis of MDD was assigned by either instrument (FH-MDD+ = 14%).

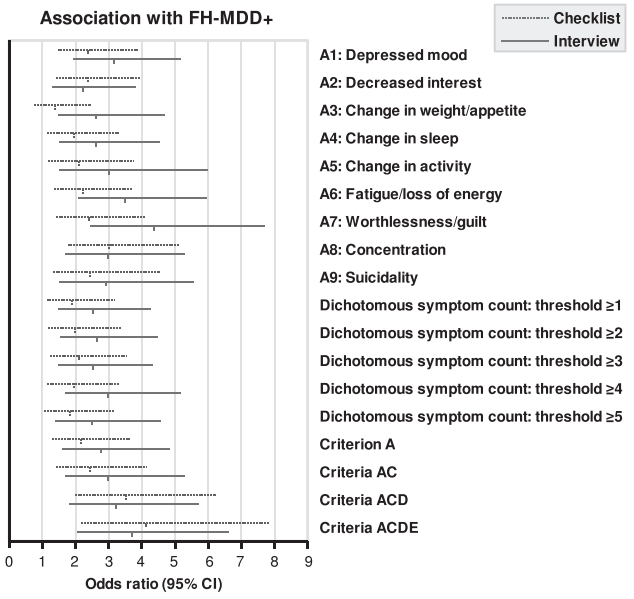
Effects of sensitivity and specificity on required sample size for association studies

The power analysis indicated that to achieve a comparable power as with the interview to identify genetic risk variants, a study using the checklist would require a sample more than four times larger for criteria ACDE and ~2 times larger for criterion A. The results of the power calculation can be found in Supplementary Tables 4 and 5 (Supplemental digital content 1, *http://links.lww.com/PG/A193*).

Discussion

The aim of the present study was to investigate whether, for the purposes of genetic studies in the general population, valid and reliable information on lifetime MDD symptoms and diagnosis can be obtained using self-administered instruments. The results suggest that self-administered instruments can be applied in the general population to (i) assess lifetime depressive symptoms and (ii) identify participants with no lifetime diagnosis of MDD. Participants identified as having a lifetime history of depressive symptoms showed an increased familial risk for MDD. However, the self-administered checklist was not reliable in terms of identifying participants with a lifetime diagnosis of MDD.

Fig. 2



Odds ratios for the association of a positive family history of major depressive disorder (FH-MDD+) with DSM-IV single depressive symptoms A1–A9, dichotomous depressive symptom count variables (threshold ≥ 1 and ≥ 5 depressive symptoms), and criterion A, criteria AC, criteria ACD, and criteria ACDE, that is lifetime diagnosis of MDD. Confidence intervals (CI) of 95% are shown. DSM-IV, *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed.

To our knowledge, the present study is the first to compare self-report and clinical interview assessment of lifetime depressive symptoms and lifetime DSM-IV criteria for MDD. Although two different methods of assessment were compared, our reliability findings for lifetime depressive symptoms are consistent with previously reported retest and inter-rater reliabilities. In a study comparing interview and videotape ratings, good to very good inter-rater reliabilities were obtained for most current depressive symptoms (Hilsenroth *et al.*, 2004). A study investigating the inter-rater reliability of the IDCL DSM-IV reported good to very good κ values for two thirds of the depressive symptoms of a lifetime MDE (Hiller *et al.*, 1990). For symptoms A1 and A2, the present study achieved test–retest reliabilities comparable to those of the Cross Cutting Symptom Measure of the DSM-V, which assesses A1 and A2 over the preceding 2 weeks (Narrow *et al.*, 2013).

For MDE criteria A, AC, and ACD, the reliability of the checklist was good. For a lifetime diagnosis of MDD, that is ACDE, reliability was only moderate. However, previous authors have considered moderate reliability estimates ($\kappa = 0.40$–0.60) to be a realistic goal for the assignment of psychiatric diagnoses compared with other standard medical diagnostic procedures (Kraemer *et al.*, 2012). Good and very good inter-rater reliabilities have been reported for (i) the SCID-I, for the current diagnosis of MDD ($\kappa = 0.61$ and 0.80; Zanarini *et al.*, 2000; Zanarini and Frankenburg, 2001; Lobbestael *et al.*, 2011) and (ii) the IDCL DSM-IV, for a lifetime diagnosis of MDD ($\kappa = 0.73$; Hiller *et al.*, 1990). However, in the pooled field trials for DSM-V, the inter-rater reliability was very low. The authors attributed this to the presence of high comorbidity with other psychiatric disorders in their sample (ICC = 0.28; Regier *et al.*, 2013). In contrast to the present study, which investigated self-report and clinical interview, these investigations did not compare different assessment methods and their cohorts were mainly comprised of psychiatric patients. Compared with population-based samples, the investigation of patient cohorts is associated with decreased external validity as reliability estimates depend on the base-rate of the respective diagnosis in the investigated population (Polsky *et al.*, 2006; Regier *et al.*, 2013). The present sample was population based, thereby implying high generalizability.

Although sensitivity showed a consecutive decrease for MDE criteria AC, ACD, and ACDE, specificity remained consistently high. Although the respective study did not assess lifetime symptoms, the high sensitivity and specificity

for MDE criteria A and AC correspond to the findings for the German version of the Patient Health Questionnaire-9 (Gräfe *et al.*, 2004), and indicates that the checklist is reliable in terms of identifying participants who fulfill and participants who do not fulfill MDE criteria A and AC.

Although the present checklist can reliably identify individuals with no lifetime diagnosis of MDD, the rather low sensitivity for MDE criteria ACD and lifetime diagnosis of MDD indicates that caution is warranted when considering its use for the identification of individuals with a lifetime history of MDD.

The rather low sensitivity for MDE criteria ACD and lifetime diagnosis of MDD was primarily because of large differences between the checklist and interviewer rating for MDE criteria D (exclusion of substance or general medical condition as cause) and E (exclusion of bereavement as a cause). These criteria accounted for markedly more MDD diagnosis exclusions in the checklist than in the interview. Therefore, although participants affirmed single depressive symptoms more frequently compared with interviewers, they were more likely to attribute their symptoms to substance use, a general medical condition, or bereavement. This attribution prevented them from receiving an MDD diagnosis and resulted in lower rates of MDD diagnosis in the checklist than in the interview. More detailed information on inconsistencies between checklist and interview for criteria D and E, suggestions for the optimization of the checklist, and the rationale for excluding criterion E from the DSM-V are provided in the Supplementary Text (Supplemental digital content 1, *http://links.lww.com/PG/A193*).

In summary, the checklist provides a reliable assessment of lifetime depressive symptoms in the general population and in previously diagnosed cases, and identifies individuals with no lifetime diagnosis of MDD in the general population. Although healthy controls may also be identified using short established screening questionnaires such as the WHO-5, the specificity values of the present checklist are superior (Primack, 2003). If the aim is to identify lifetime MDD cases, the checklist could serve as a screening instrument: in a two-step assessment, individuals who fulfill MDE criteria A and C in the checklist could be interviewed by an expert.

In both the checklist and the interview, the single depressive symptoms (A1–A9) showed a strong association with FH-MDD+, with the exception of A3 (change in weight/appetite) in the checklist. This discrepancy between checklist and interview for A3 could be because of the fact that to be rated as present, the change in weight/appetite cannot be attributable to medication. Although this is clear to clinicians during the interview, the checklist does not inform the participant of this exclusion rule. This shortcoming requires amendment in future investigations. In the interview, A6 (fatigue/loss of energy) and A7 (worthlessness/guilt) yielded higher associations with FH-MDD+ than a lifetime diagnosis of MDD, and equally high associations as a checklist lifetime diagnosis of MDD. This finding suggests that these two depressive symptoms may represent a particularly familial, and therefore genetic, component of MDD. This hypothesis warrants further investigation in larger sets of genetic data. If this hypothesis were confirmed, individuals from the general population with a high familial risk of MDD could be identified by interview assessment of these two depressive symptoms. Interestingly, the explorative analysis showed that the depressive symptoms for which higher heritability estimates (h^2) were observed in the twin study by Kendler *et al.* (2013) were also more strongly associated with FH-MDD+ in both the interview and the checklist. For both the checklist and the interview, the association with FH-MDD+ increased with the consecutive inclusion of MDE criteria A, AC, ACD, and ACDE. This finding supports the external validity of both the checklist and the interview in terms of the assessment of a lifetime diagnosis of MDD. FH-MDD+ was especially prevalent in participants diagnosed with MDD by the interview and/or the checklist. These observations are in line with the assumption that FH-MDD+ could serve as a marker (proxy) for familial (partially genetic) risk for MDD.

The power analysis indicated that compared with studies using clinical interviews, studies using the checklist to investigate genetic risk variants for MDD would require a sample size four and two times larger for the diagnosis of MDD and criterion A, respectively. However, as the power calculation was based on the assumption that the interview correctly assesses a lifetime diagnosis of MDD, the power of the checklist might be underestimated by this approach.

The present study had several strengths. First, to our knowledge, this is the first investigation to compare the reliability of the assessment of lifetime depressive symptoms and a lifetime diagnosis of MDD by a self-administered checklist and standard clinical interview, and to investigate the usefulness of both instruments to assess individuals in genetic studies. Second, it shows the ability of the checklist to assess lifetime depressive symptomatology in detail, facilitating a symptom-based analysis of genetic risk variants for depression. Symptom-based analyses allow, in comparison with diagnosis-based analyses, for the following: (i) investigation of genetic risk variants for the individual depressive symptoms as they have been found to differ in their heritability and underlying genetic factors (Kendler *et al.*, 2013); (ii) cross-diagnostic analysis of depressive symptoms in other psychiatric disorders (e.g. in anxiety disorders); and (iii) investigation of depressive symptoms in controls who do not fulfill the entire MDD criteria. Recent developments in psychiatric research such as the Research Domain Criteria approach acknowledge the necessity of addressing the heterogeneity and dimensionality of psychopathology and underlying neurobiological mechanisms

by dissecting global diagnoses (Cuthbert and Insel, 2013; Woody and Gibb, 2015). Symptom-based approaches have already been applied successfully in the formal genetic studies (Kendler *et al.*, 2013) and our group has applied this approach successfully to molecular genetic studies of bipolar disorder and depression (Breuer *et al.*, 2011; Meier *et al.*, 2012; Miro *et al.*, 2012; Treutlein *et al.*, 2017). In addition, the detailed assessment of single symptoms allows for the selection of specific MDD subgroups, a strategy that has been proven successful in GWAS investigating cases with severe depression or using stratification for age at onset (CONVERGE con sortium, 2015; Cai *et al.*, 2017; Power *et al.*, 2017). Third, in the context of genetic studies of depressive symptomatology in the general population, large samples and thus the highly efficient acquisition of reliable phenotype data are required. As clinical interviews by trained experts are expensive, recruitment on the basis of self-ratings could lead to significant savings and increase possible sample sizes (Abbasi, 2017), especially in the light of continuously decreasing prices for genome-wide genotyping (arrays reaching prices <$40/unit). Now and in the future, large studies investigating genetic risk factors for physical health and somatic diseases are being and will be carried out. The availability of efficient self-administered instruments for mental health could decide whether or not mental health will be included in the list of the assessed phenotypes. Therefore, self-administered instruments may contribute toward the availability of large-scale samples for the investigation of the genetic underpinnings of mental disorders in the future. The value of self-rating for genetic studies is strongly supported by the findings of a recent study that analyzed the self-report data of a total of more than 120 000 cases and more than 330 000 controls, which were collected through 23andMe (Hyde *et al.*, 2016). Finally, in addition to the conventional approach of using only the categorical diagnosis (binary classification) as the phenotype of interest, the properties of single depressive symptoms and partial diagnosis criteria (i.e. A, AC, ACD) were investigated to explore the source of agreement/disagreement between interview and self-administered checklist. In addition, symptom count measures were included in our analysis as they have been suggested to reflect the continuous distribution of depressive symptomatology in the general population (Melzer *et al.*, 2002) and to represent a proxy for depression severity of MDD (Musliner *et al.*, 2015; Ware *et al.*, 2015; Okbay *et al.*, 2016).

The present study had several limitations. First, the usefulness of both instruments to assess participants in genetic studies was investigated by analyzing their association with the participants' family history of MDD. However, it is important to note that a positive family history of MDD should not be equated with genetic risk for MDD because environmental factors may also add to the familial aggregation of MDD. Nevertheless, the evidence for a meaningful contribution of shared environmental factors to the familial aggregation for MDD is sparse (Sullivan *et al.*, 2000; Flint and Kendler, 2014). Future studies should include a genetic risk score (Purcell *et al.*, 2009; Wray *et al.*, 2014) as the genetic validity criterion. Second, the checklist does not assess differential or comorbid diagnoses, and therefore precludes the identification of individuals with a lifetime bipolar (mixed) episode. Complex termination rules are difficult to understand and implement in self-administered formats, where no additional information is provided by an interviewer, particularly in paper–pencil versions. The checklist may be implemented in a computerized format, with automatized termination rules being triggered on the basis of the participant's response. Fully structured computerized checklists for the assessment of psychiatric disorders are already available (e.g. Diagnostic Interview Schedule; Robins *et al.*, 2000). Third, the checklist only requires participants to continue with A3 if A1 and/or A2 have been answered in the affirmative. Participants who do not report depressive symptoms A1 and/or A2 during one or more lifetime depressive episodes (lasting ≥ 2 weeks) are instructed to terminate the checklist. For these participants, no information on depressive symptoms A3–A9 is available. Deletion of such termination rules may increase the value of self-administered instruments for the purposes of genetic studies in the general population. Finally, the main finding is that depressive symptoms can be assessed reliably with the self-rating checklist, whereas this is not the case for lifetime diagnosis of MDD. The checklist in its current form should therefore be applied with caution if the study aim is to screen for individuals with a lifetime diagnosis of MDD.

In conclusion, our results suggest that for the purposes of time-efficient and cost-efficient recruitment in genetic studies of MDD in the general population, self-administered instruments can replace the clinical interview to (i) assess lifetime depressive symptoms and (ii) identify individuals with no lifetime diagnosis of MDD.

Acknowledgements

This study was supported by the German Federal Ministry of Education and Research (BMBF) within the context of the following projects: (i) the Integrated Genome Research Network (IG) MooDS of the German National Genome Research Network plus (NGFNplus) (grant 01GS08144 to Sven Cichon and Markus M. Nöthen; grant 01GS08147 to Marcella Rietschel); and (ii) the Integrated Network IntegraMent (Integrated Understanding of Causes and Mechanisms in Mental Disorders; grant 01ZX1314A to Sven Cichon and Markus M. Nöthen; grant 01ZX1314G to Marcella Rietschel). The latter project is being conducted under the auspices of the BMBF e:Med Systems Medicine Programme. Markus M. Nöthen is a member of the DFG-funded Excellence-Cluster ImmunoSensation. Marcella Rietschel

and Stephanie H. Witt are members of the Collaborative Research Center SFB636 of the University of Heidelberg. This study was further supported by the German Research Foundation (DFG; grant FOR2107; grant RI908/11-1 to M.R.; grant NO246/10-1 to M.M.N.).

Conflicts of interest

There are no conflicts of interest.

References

Abbasi J (2017). 23andMe, Big Data, and the Genetics of Depression. *JAMA* **317**:14–16.

American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed., Text Revision. Washington, DC: American Psychiatric Association.

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed. Arlington, VA: American Psychiatric Publishing.

Angst F, Stassen HH, Clayton PJ, Angst J (2002). Mortality of patients with mood disorders: follow-up over 34–38 years. *J Affect Disord* **68**:167–181.

Bebbington PE (1998). Sex and depression. *Psychol Med* **28**:1–8.

Breuer R, Hamshere ML, Strohmaier J, Mattheisen M, Degenhardt F, Meier S, *et al.* (2011). Independent evidence for the selective influence of GABA(A) receptors on one component of the bipolar disorder phenotype. *Mol Psychiatry* **16**:587–589.

Bromet E, Andrade LH, Hwang I, Sampson NA, Alonso J, de Girolamo G, *et al.* (2011). Cross-national epidemiology of DSM-IV major depressive episode. *BMC Med* **9**:90.

Cai N, Bigdeli TB, Kretzschmar WW, Li Y, Liang J, Hu J, *et al.* (2017). 11,670 whole-genome sequences representative of the Han Chinese population from the CONVERGE project. *Sci Data* **4**:170011.

Cardno AG, Marshall EJ, Coid B, Macdonald AM, Ribchester TR, Davies NJ, *et al.* (1999). Heritability estimates for psychotic disorders: the Maudsley twin psychosis series. *Arch Gen Psychiatry* **56**:162–168.

CONVERGE consortium (2015). Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature* **523**:588–591.

Craddock N, Sklar P (2009). Genetics of bipolar disorder: successful start to a long journey. *Trends Genet* **25**:99–105.

Cuijpers P, Smit F (2002). Excess mortality in depression: a meta-analysis of community studies. *J Affect Disord* **72**:227–236.

Cuthbert BN, Insel TR (2013). Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med* **11**:126.

First MB, Spitzer RL, Gibbon M, Williams JBW (1996). *Structured Clinical Interview for DSM-IV Axis I Disorders, Clinician Version (SCID-CV)*. Washington, DC: American Psychiatric Press.

Fleiss JL, Cohen J (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas* **33**:613–619.

Flint J, Kendler KS (2014). The genetics of major depression. *Neuron* **81**:484–503.

Gräfe K, Zipfel S, Herzog W, Löwe B (2004). Screening psychischer Störungen mit dem 'Gesundheitsfragebogen für Patienten (PHQ-D. *Diagnostica* **50**:171–181.

Hasin DS, Goodwin RD, Stinson FS, Grant BF (2005). Epidemiology of major depressive disorder: results from the National Epidemiologic Survey on Alcoholism and Related Conditions. *Arch Gen Psychiatry* **62**:1097–1106.

Hiller W, von Bose M, Dichtl G, Agerer D (1990). Reliability of checklist-guided diagnoses for DSM-IIIR affective and anxiety disorders. *J Affect Disord* **20**:235–247.

Hiller W, Zaudig M, Mombour W (1997). *IDCL – Internationale Diagnosen Checklisten für DSM-IV und ICD-10 (Manual, 31 Checklisten nach DSM-IV und Screening-Blatt)* [ICDL- International diagnostic checklist for DSM-IV and ICD 10 (Manual, 31 checklists for DSM-IV and checklist sheet)]. Göttingen: Hogrefe.

Hilsenroth MJ, Baity MR, Mooney MA, Meyer GJ (2004). DSM-IV Major Depressive Episode criteria: an evaluation of reliability and validity across three different rating methods. *Int J Psychiatry Clin Pract* **8**:3–10.

Hoefgen B, Schulze TG, Ohlraun S, von Widdern O, Hofels S, Gross M, *et al.* (2005). The power of sample size and homogenous sampling: association between the 5-HTTLPR serotonin transporter polymorphism and major depressive disorder. *Biol Psychiatry* **57**:247–251.

Hyde CL, Nagle MW, Tian C, Chen X, Paciga SA, Wendland JR, *et al.* (2016). Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat Genet* **48**:1031–1036.

Insel TR (2008). Assessing the economic costs of serious mental illness. *Am J Psychiatry* **165**:663–665.

Kendler KS, Davis CG, Kessler RC (1997). The familial aggregation of common psychiatric and substance use disorders in the National Comorbidity Survey: a family history study. *Br J Psychiatry* **170**:541–548.

Kendler KS, Aggen SH, Neale MC (2013). Evidence for multiple genetic factors underlying DSM-IV criteria for major depression. *JAMA Psychiatry* **70**:599–607.

Kessler RC, McGonagle KA, Swartz M, Blazer DG, Nelson CB (1993). Sex and depression in the National Comorbidity Survey. I: lifetime prevalence, chronicity and recurrence. *J Affect Disord* **29**:85–96.

Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, *et al.* (2003). The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA* **289**:3095–3105.

Kohli MA, Lucae S, Saemann PG, Schmidt MV, Demirkan A, Hek K, *et al.* (2011). The neuronal transporter gene SLC6A15 confers risk to major depression. *Neuron* **70**:252–265.

Kraemer HC, Kupfer DJ, Clarke DE, Narrow WE, Regier DA (2012). DSM-5: how reliable is reliable enough? *Am J Psychiatry* **169**:13–15.

Kroenke K, Spitzer RL, Williams JB, Lowe B (2010). The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. *Gen Hosp Psychiatry* **32**:345–359.

Lieb R, Isensee B, Höfler M, Pfister H, Wittchen H (2002). Parental major depression and the risk of depression and other mental disorders in offspring: a prospective-longitudinal community study. *Arch Gen Psychiatry* **59**:365–374.

Lobbestael J, Leurgans M, Arntz A (2011). Inter-rater reliability of the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID I) and Axis II Disorders (SCID II). *Clin Psychol Psychother* **18**:75–79.

Luppa M, Heinrich S, Angermeyer MC, Konig HH, Riedel-Heller SG (2007). Cost-of-illness studies of depression: a systematic review. *J Affect Disord* **98**:29–43.

Ripke S, Wray NR, Lewis CM, Hamilton SP, Weissman MM, *et al.*, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium (2013). A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry* **18**:pp. 497–511.

Meier S, Mattheisen M, Vassos E, Strohmaier J, Treutlein J, Josef F, *et al.* (2012). Genome-wide significant association between a 'negative mood delusions' dimension in bipolar disorder and genetic variation on chromosome 3q26.1. *Transl Psychiatry* **2**:e165.

Melzer D, Tom BD, Brugha TS, Fryers T, Meltzer H (2002). Common mental disorder symptom counts in populations: are there distinct case groups above epidemiological cut-offs? *Psychol Med* **32**:1195–1201.

Miro X, Meier S, Dreisow ML, Frank J, Strohmaier J, Breuer R, *et al.* (2012). Studies in humans and mice implicate neurocan in the etiology of mania. *Am J Psychiatry* **169**:982–990.

Muglia P, Tozzi F, Galwey NW, Francks C, Upmanyu R, Kong XQ, *et al.* (2010). Genome-wide association study of recurrent major depressive disorder in two European case-control cohorts. *Mol Psychiatry* **15**:589–601.

Mulrow CD, Williams JW Jr, Gerety MB, Ramirez G, Montiel OM, Kerber C (1995). Case-finding instruments for depression in primary care settings. *Ann Intern Med* **122**:913–921.

Musliner KL, Seifuddin F, Judy JA, Pirooznia M, Goes FS, Zandi PP (2015). Polygenic risk, stressful life events and depressive symptoms in older adults: a polygenic score analysis. *Psychol Med* **45**:1709–1720.

Narrow WE, Clarke DE, Kuramoto SJ, Kraemer HC, Kupfer DJ, Greiner L, *et al.* (2013). DSM-5 field trials in the United States and Canada, Part III: development and reliability testing of a cross-cutting symptom assessment for DSM-5. *Am J Psychiatry* **170**:71–82.

Okbay A, Baselmans BML, De Neve JE, Turley P, Nivard MG, Fontana MA, *et al.* (2016). Genetic variants associated with subjective well-being, depressive symptoms and neuroticism identified through genome-wide analyses. *Nat Genet* **48**:624–633.

Polsky D, Doshi JA, Bauer MS, Glick HA (2006). Clinical trial-based cost-effectiveness analyses of antipsychotic use. *Am J Psychiatry* **163**:2047–2056.

Power RA, Tansey KE, Buttenschon HN, Cohen-Woods S, Bigdeli T, Hall LS, *et al.* (2017). Genome-wide Association for Major Depression Through Age at Onset Stratification: Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. *Biol Psychiatry* **81**:325–335.

Primack BA (2003). The WHO-5 Wellbeing Index performed the best in screening for depression in primary care. *ACP J Club* **139**:48.

Purcell SM, Cherny SS, Sham PC (2003). Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* **19**:149–150.

Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, *et al.* (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**:748–752.

Regier DA, Narrow WE, Clarke DE, Kraemer HC, Kuramoto SJ, Kuhl EA, *et al.* (2013). DSM-5 field trials in the United States and Canada, Part II: test–retest reliability of selected categorical diagnoses. *Am J Psychiatry* **170**:59–70.

Rietschel M, Mattheisen M, Frank J, Treutlein J, Degenhardt F, Breuer R, *et al.* (2010). Genome-wide association-, replication-, and neuroimaging study implicates HOMER1 in the etiology of major depression. *Biol Psychiatry* **68**:578–585.

Robins LN, Cottler LB, Bucholz KK, Compton WM, North CS, Rourke KM (2000). *Diagnostic Interview Schedule for the DSM-IV (DIS-IV)*. St. Louis, MO: Washington University, School of Medicine.

Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**:421–427.

Shi J, Potash JB, Knowles JA, Weissman MM, Coryell W, Scheftner WA, *et al.* (2011). Genome-wide association study of recurrent early-onset major depressive disorder. *Mol Psychiatry* **16**:193–201.

Shyn SI, Shi J, Kraft JB, Potash JB, Knowles JA, Weissman MM, *et al.* (2011). Novel loci for major depression identified by genome-wide association study of Sequenced Treatment Alternatives to Relieve Depression and meta-analysis of three studies. *Mol Psychiatry* **16**:202–215.

Sklar P, Ripke S, Scott LJ, Andreassen OA, Cichon S, Craddock N, *et al.* (2011). Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet* **43**:977–983.

Sullivan PF, Neale MC, Kendler KS (2000). Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry* **157**:1552–1562.

Sullivan PF, Kendler KS, Neale MC (2003). Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry* **60**:1187–1192.

Sullivan PF, de Geus EJ, Willemsen G, James MR, Smit JH, Zandbelt T, *et al.* (2009). Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. *Mol Psychiatry* **14**:359–375.

Treutlein J, Strohmaier J, Frank J, Witt SH, Rietschel L, Forstner AJ, *et al.* (2017). Association between neuropeptide Y receptor Y2 promoter variant rs6857715 and major depressive disorder. *Psychiatr Genet* **27**:34–37.

Ware EB, Mukherjee B, Sun YV, Diez-Roux AV, Kardia SLR, Smith JA (2015). Comparative genome-wide association studies of a depressive symptom phenotype in a repeated measures setting by race/ethnicity in the multi-ethnic study of atherosclerosis. *BMC Genet* **16**:118.

Weissman MM, Bland R, Joyce PR, Newman S, Wells JE, Wittchen HU (1993). Sex differences in rates of depression: cross-national perspectives. *J Affect Disord* **29**:77–84.

Williams JW Jr, Noel PH, Cordes JA, Ramirez G, Pignone M (2002). Is this patient clinically depressed? *JAMA* **287**:1160–1170.

Wirtz M, Caspar F (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität [Inter-rater agreement and inter-rater reliability]*. Göttingen: Hogrefe.

Wittchen HU, Zaudig M, Fydrich T . (1997). *Strukturiertes Klinisches Interview für DSM-IV* [Structured clinical interview for DSM-IV]. Göttingen: Hogrefe.

Wittchen HU, Jacobi F, Rehm J, Gustavsson A, Svensson M, Jonsson B, *et al.* (2011). The size and burden of mental disorders and other disorders of the brain in Europe 2010. *Eur Neuropsychopharmacol* **21**:655–679.

Woody ML, Gibb BE (2015). Integrating NIMH Research Domain Criteria (RDoC) into Depression Research. *Curr Opin Psychol* **4**:6–12.

World Health Organization (2008). *The global burden of disease: 2004 update*. Geneva: World Health Organization.

Wray NR, Pergadia ML, Blackwood DH, Penninx BW, Gordon SD, Nyholt DR, *et al.* (2012). Genome-wide association study of major depressive disorder: new results, meta-analysis, and lessons learned. *Mol Psychiatry* **17**:36–48.

Wray NR, Lee SH, Mehta D, Vinkhuyzen AA, Dudbridge F, Middeldorp CM (2014). Research review: polygenic methods and their application to psychiatric traits. *J Child Psychol Psychiatry* **55**:1068–1087.

Zanarini MC, Frankenburg FR (2001). Attainment and maintenance of reliability of axis I and II disorders over the course of a longitudinal study. *Compr Psychiatry* **42**:369–374.

Zanarini MC, Skodol AE, Bender D, Dolan R, Sanislow C, Schaefer E, *et al.* (2000). The Collaborative Longitudinal Personality Disorders Study: reliability of axis I and II diagnoses. *J Pers Disord* **14**:291–299.