Psychological Medicine (2017), **47**, 1816–1824. © Cambridge University Press 2017
doi:10.1017/S0033291717000277

ORIGINAL ARTICLE

Genetic and environmental influences on last-year major depression in adulthood: a highly heritable stable liability but strong environmental effects on 1-year prevalence

K. S. Kendler[1,2]* and C. O. Gardner[1]

[1] *Department of Psychiatry, Virginia Commonwealth University School of Medicine, Richmond, VA, USA*
[2] *Department of Human and Molecular Genetics, Virginia Commonwealth University School of Medicine, Richmond, VA, USA*

Background. This study seeks to clarify the contribution of temporally stable and occasion-specific genetic and environmental influences on risk for major depression (MD).

Method. Our sample was 2153 members of female–female twin pairs from the Virginia Twin Registry. We examined four personal interview waves conducted over an 8-year period with MD in the last year defined by DSM-IV criteria. We fitted a structural equation model to the data using classic Mx. The model included genetic and environmental risk factors for a latent, stable vulnerability to MD and for episodes in each of the four waves.

Results. The best-fit model was simple and included genetic and unique environmental influences on the latent liability to MD and unique wave-specific environmental effects. The path from latent liability to MD in the last year was constant over time, moderate in magnitude (+0.65) and weaker than the impact of occasion-specific environmental effects (+0.76). Heritability of the latent stable liability to MD was much higher (78%) than that estimated for last-year MD (32%). Of the total unique environmental influences on MD, 13% reflected enduring consequences of earlier environmental insults, 17% diagnostic error and 70% wave-specific short-lived environmental stressors.

Conclusions. Both genetic influences on MD and MD heritability are stable over middle adulthood. However, the largest influence on last-year MD is short-lived environmental effects. As predicted by genetic theory, the heritability of MD is increased substantially by measurement at multiple time points largely through the reduction of the effects of measurement error and short-term environmental risk factors.

Received 8 August 2016; Revised 11 November 2016; Accepted 23 January 2017; First published online 15 February 2017

Key words: Environment, heritability, longitudinal study, major depression.

Introduction

Lifetime major depression (MD), when assessed at one occasion during adulthood, has a heritability of approximately 40% (Sullivan *et al.* 2000; Kendler *et al.* 2006). When lifetime MD is evaluated twice several years apart (Kendler *et al.* 2001*a*), the latent heritability increases substantially. This is because unreliability of measurement or the impact of short-term environmental risks (which cannot easily be distinguished) set an upper limit on heritability. If such 'errors' are substantial, measuring MD more than once and modeling MD as the latent or stable liability assessed at multiple time points would be expected, by standard statistical theory (Falconer, 1989), to increase heritability estimates.

Indeed, questions have been raised about the validity of the lifetime assessment for MD (Coughlin, 1990; Andrews *et al.* 1999; Patten, 2003; Wells & Horwood, 2004; Kruijshaar *et al.* 2005). Many episodes are subject to forgetting (Moffitt *et al.* 2010). Recall is influenced by the level of current depressive symptoms (Kendler *et al.* 2001*b*). Finally, twin studies of lifetime MD (Sullivan *et al.* 2000; Kendler *et al.* 2006) assume that genetic risk factors for MD are stable over adulthood – that is that the same genetic factors would impact on risk for MD at ages 25, 30, 35 years, etc. We are aware of only a single test of this hypothesis conducted over a quite short period (~18 months) (Kendler *et al.* 1993*b*). This is a legitimate concern because genetic risk factors have been shown to change for depressive symptoms over childhood and adolescence (Silberg *et al.* 2001; Kendler *et al.* 2008; Nivard

* Address for correspondence: K. S. Kendler, M.D., Department of Psychiatry and Human and Molecular Genetics, Virginia Institute for Psychiatric and Behavioral Genetics of VCU, Box 980126, Richmond, VA 23298-0126, USA.
(Email: Kenneth.Kendler@vcuhealth.org)

et al. 2014; Waszczuk *et al.* 2016), and for major depression in late adult life (Kendler *et al.* 2009).

To address these concerns, we examined last-year MD assessed at personal interview in a population-based sample of female–female twin pairs in mid-adult life assessed at four waves spread out over an 8-year time period. We fit a series of structural equation models that address the following questions:

(1) Are the genetic influences on last-year MD stable over middle adulthood?
(2) What proportion of the environmental contributions to MD are (i) occasion-specific – that is, reflect short-term environmental stressors and/or measurement error *v*. (ii) temporally stable?
(3) Does the heritability of MD change over time?
(4) How much does the latent heritability for MD measured over these four interview waves increase over that obtained from each individual assessment? In particular, is it possible that the liability to MD is highly heritable but heritability assessments are substantially attenuated because of the large impact of diagnostic error and/or short-term environmental stressors?

Method

Sample and measures

Participants were the Caucasian female–female subsample from the Virginia Adult Twin Study of Psychiatric and Substance Use Disorders (VATSPSUD; Kendler & Prescott, 2006). All subjects were ascertained from the population-based Virginia Twin Registry formed from a systematic review of birth certificates in the Commonwealth of Virginia. These pairs were born during 1934–1974 and were eligible if both members responded to a mailed questionnaire. As detailed in Table 1, this cohort was interviewed four times in person or by telephone over a decade with a total attrition of only 20.7%. At the first wave, the mean (S.D.) of years of education of this cohort was 14.1 (2.3). The % cooperation (of eligible twins) across the four waves equaled: 92%, 93%, 88%, and 85% (Kendler & Prescott, 2006).

Interviews were separated by a minimum of 1 year and were conducted by trained mental health professionals blind to knowledge about the co-twin. This research was approved by the VCU IRB. Informed consent was obtained prior to personal interviews and assent prior to all phone interviews. Zygosity was determined by discriminate function analyses using standard twin questions validated against DNA genotyping in 496 pairs (Kendler & Prescott, 1999).

In all interviews, twins were asked about the occurrence in the last year of 15 symptoms reflecting all DSM-IV A criteria for MD (APA, 1987) which are nearly identical to those in DSM-5 (APA, 2013). As detailed elsewhere (Kendler & Prescott, 2006), questions for this section were adapted from the SCID interview (Spitzer & Williams, 1985). With the assistance of the interviewer, the twin subjects then aggregated these symptoms in time, reported total number of episodes and dated, to the month, the onset and offset of each episode. In earlier parts of the interview, stressful events in the last year were also reported by month, so the respondent was repeatedly helped by the interviewer to structure in time their last year experiences. Inter-rater reliability for last-year MD was good: κ (Cohen, 1960) = +0.74 (S.E. = 0.08), tetrachoric correlation $r = 0.96 \pm 0.03$. In 375 twins interviewed twice by different interviewers with a mean (S.D.) inter-interview interval of 30 (9) days, the test–retest reliability of the diagnosis of MD in the last year was: $\kappa = +0.66$ [95% confidence interval (CI) 0.58–0.74], and tetrachoric correlation $r = 0.88$ (95% CI 0.82–0.93) (Kendler *et al.* 2005).

Data analysis

We fitted three different models to the four interview waves on the presence/absence of one or more MD episodes meeting DSM-III-R criteria in the last year: independent pathway, common pathway and Cholesky decomposition (CD). The common pathway model is a subtype of the independent pathway model so their relative fits can be directly compared. This is not true for the Cholesky model but we nonetheless relied on a comparison of fit indices as a rough guide, a commonly used approach in structural modeling.

As with traditional twin modeling, we divided the individual differences in liability to MD into three categories: additive genetic effects (A), shared or family environment (C), and unique environment (E) (Guerrini *et al.* 2005). Shared environment reflects family and community experiences that increase similarity in twins raised together. Unique environment included environmental experiences not shared by siblings that impact on the phenotype under study and diagnostic error. Both independent and common pathway models further divide the sources of individual differences in liability to last-year MD into those that are in common across all four waves of assessment (indicated by subscript c) and those that are occasion-specific in their effect (subscript s).

The independent pathway model allows the common A, C and E effects to directly impact on last-year MD. The common pathway model, by contrast, constrains the A, C and E effects to impact on a latent liability to MD which in turn influences last-year MD. The full Cholesky model assumes four A, C and E factors the first of which impacts on all four times

https://doi.org/10.1017/S0033291717000277 Published online by Cambridge University Press

of assessment. The second and third factors impact, respectively, on the last three and the last two waves. The final factors influence only the fourth wave.

Model fitting was conducted using the classic Mx program (Neale *et al.* 2003). The goal of our model fitting is to achieve the best possible balance of explanatory power and simplicity. This goal is operationalized by Akaike's Information Criterion (AIC; Akaike, 1987), which equals χ^2 – 2df, where df is degrees of freedom. We seek to minimize the AIC value.

Errors of measurement should contribute to occasion-specific individual environmental effects (E_s). Because we have an estimate of the test–retest reliability of MD in our sample (+0.88), we can subdivide E_s into two parts reflecting measurement error and wave-specific unique environmental effects. We assume that measurement error contributes 1 minus 0.88 or 12% of the variance in liability. If the total $e_s^2 = x$, then x – 0.12 equals the wave-specific unique environmental effects.

Ethical standards

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

Results

At the first wave of interviews, the 2153 cooperative twins reflected 596 complete monozygotic (MZ) pairs, 433 complete dizygotic (DZ) pairs, and 95 unmatched twins who were included in the analysis and provided information about threshold placements. The prevalence of last-year MD was relatively stable over the four interview waves ranging from 9.3% to 10.8% (Table 1). Among individuals who reported an episode of MD in the last year at the first interview, the median and mean (S.D.) age at onset, number of lifetime episodes and longest duration of an episode (in days) were, respectively: 20 and 22.1 (9.7), 3 and 9.0 (17.9), and 90 and 214.0 (568.6).

The tetrachoric correlation matrix for last-year MD (Table 1) shows no clear attenuation over time. Table 1 also contains the mean inter-wave intervals which ranged from 17 to 44 months. There is no evident relationship between inter-wave intervals and correlations in MD occurrence. The mean interval between waves 1 and 4 was 7.7 years.

Table 2 presents model-fitting results. For each of our three models – independent pathway, common pathway and CD – we began with a full model and simplified until we could no longer improve the model fit as indexed by AIC. For the common pathway model, we could constrain the loadings from the common factor to the four waves of last-year MD to equality (model II), eliminate all C effects (model III), constrain the specific A and E effects on the last-year assessments to equality (model IV) and drop the specific genetic effects all with continued improvements of AIC (model V). Model V was the best-fit common pathway model.

For the independent pathway model, dropping the C effects (model II) improved the AIC value but it deteriorated when we tried to constrain the common A and E loadings to equality (model III). Dropping the occasion-specific A effects (model IV) further improved the fit index but constraining the specific E loadings to equality across the four waves did not (model V). Model IV was therefore the best-fit independent pathway model.

For the CD, we could, with an improvement in AIC, drop all C effects (model II), drop the fourth genetic factors (model III), drop the third genetic factor (model IV) and drop the second genetic factor (model V). We tried to further simplify the model by dropping all the off-diagonal E effects (so that the four E factors only impacted respectively, waves 1, 2, 3, and 4 with no cross-wave loadings) (model VI). Here, however, AIC deteriorated so model V was the best-fit CD.

As seen from Table 2, the best-fit common pathway model (CP-V) had a considerably lower AIC than the best-fit independent pathway (IP-IV) or Cholesky decomposition (CD-V) models. So the CP-V was the overall best-fit model. All three modeling approaches agreed on two key findings. First, shared environmental effects did not substantially contribute to individual differences in MD liability (hence the improved fits seen with models CP-III, IP-II and CD-II). Second, given one common genetic factor impacting on all four assessment waves, there was no need for other genetic influences (improved fits with models CP-V, IP-IV and CD-V). Furthermore, the failure of model CD-VI indicated the existence of important unique environmental effects that were shared across waves.

The best-fitting overall model (CP-V) was very simple (Fig. 1). The latent liability to last-year MD as assessed four times over an 8-year period resulted from only genetic and individual specific environmental factors. The heritability of this latent liability was high (78%) with the remaining 22% resulting from environmental factors that were both unique to individual twins and temporally stable in their effect.

By far the strongest contribution to risk for last-year MD, however, was our wave-specific E effects accounting for 58% of the overall liability. As detailed in the methods section, using our test-retest reliability results

https://doi.org/10.1017/S0033291717000277 Published online by Cambridge University Press

Table 1. *Descriptive statistics for the twin sample*

Wave	Age at interview (S.D.)	Sample size	Prevalence last-year MD	Tetrachoric correlations of MD across waves			Mean between interview interval in months (S.D.)	
				2	3	4	Waves	Interval
1	30.1 (7.6)	2153	10.1	+0.41	+0.35	+0.43	Waves 1–2	17.0 (3.7)
2	31.5 (7.5)	1995	9.9	–	+0.37	+0.49	Waves 2–3	44.3 (4.0)
3	35.1 (7.5)	1886	9.3		–	+0.49	Waves 3–4	31.1 (6.7)
4	37.7 (7.5)	1706	10.8			–	Waves 1–4	92.2 (6.6)

MD, Major depression.

Table 2. *Model fitting results*

Model	Common	Occasion specific	Other features	−2 log likelihood	Degrees of freedom	AIC
CP-I	ACE	ACE		4810.19	8690	−12 568.8
CP-II	ACE	ACE	Common loadings	4815.04	8693	−12 570.7
CP-III	AE	AE	constrained to equality	4815.57	8698	−12 580.4
CP-IV	AE	AE	Specific loadings also	4817.00	8701	−12 585.0
CP-V	AE	E	constrained to equality	4817.86	8702	−12 586.1[a,b]
IP-I	ACE	ACE		4802.96	8684	−12 565.1
IP-II	AE	AE		4809.34	8692	−12 574.7
IP-III	AE	AE	Common loadings	4817.08	8695	−12 573.0
IP-IV	AE	E	constrained to equality	4809.90	8696	−12 581.1[a]
IP-V	AE	E	Specific loadings also constrained to equality	4817.73	8699	−12 580.7
CD-I	A1–A4, C1–C4, E1–E4	–		4802.80	8678	−12 553.2
CD-II	A1–A4, E1–E4	–		4804.50	8688	−12 571.5
CD-III	A1–A3, E1–E4	–		4804.62	8689	−12 573.4
CD-IV	A1–A2, E1–E4	–		4804.97	8691	−12 577.0
CD-V	A1, E1–E4	–		4809.44	8694	−12 578.6[a]
CD-VI	A1, E1–E4	–	Eliminate all off-diagonal E estimates	4823.37	8700	−12 576.6

AIC, Akaike's Information Criterion; A, additive genetic effects; C, shared environmental effects; E, individual specific environmental effects.

[a] Best-fit version within model: CP (common pathway), IP (independent pathway) or CD (Cholesky decomposition).

[b] Best overall fit.

of last-year MD (estimated to equal +0.88) allows us to divide the influences on last-year MD into four components: stable genetic effects 33%, stable unique environmental effects 9%, unreliability of assessment 12%, and wave-specific unique environmental effects 46%.

Discussion

We examined last-year DSM-III-R MD assessed at four waves of personal interviews completed over an 8-year period years in a population-based sample of female twins in middle adult life. We fitted three classes of structural equation models and found one that clearly fitted best. From that model, we sought to answer the four questions we articulated above, the results of which we review in turn.

First, genetic influences on last-year MD were stable over the four waves of assessment. While our independent and common pathway models included wave-specific genetic influences and our Cholesky model allowed for 'genetic innovation' over time (that is new genetic effects at later waves), in all

https://doi.org/10.1017/S0033291717000277 Published online by Cambridge University Press

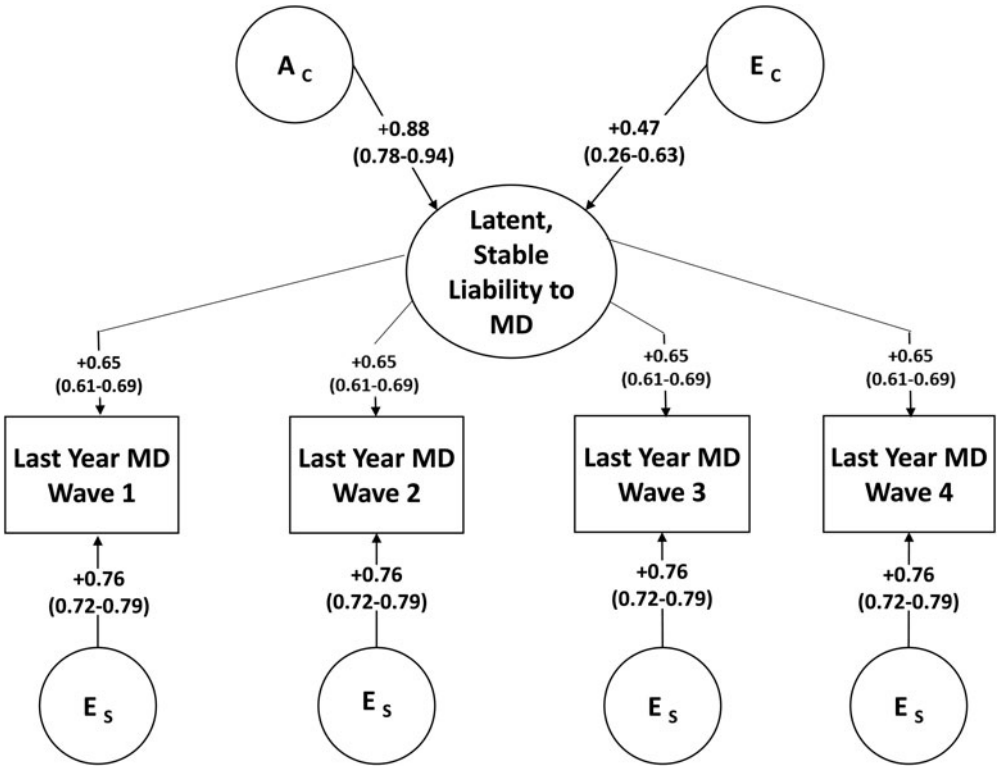

Fig. 1. The best fit structural equation model for last-year major depression assessed four times over 8 years in female–female twin pairs. The path estimates, presented along with their 95% confidence intervals, represent standardized partial regression coefficients and therefore must be squared to equal the proportion of variance accounted for. The following abbreviations are used: A, Additive genetic effects; E, individual specific environmental effects; subscript 'c' for common or stable; subscript 's' for occasion-specific.

cases we could constrain to zero these novel sources of genetic effects with an improvement in model fit. We looked hard but could find no evidence from our results that new genetic influences on MD were activated over the 8 years of our study period. These results provide further support for the hypothesis that contrary to adolescence (Silberg *et al.* 2001; Kendler *et al.* 2008) and late adult life (Kendler *et al.* 2009), genetic influences on depression are stable though middle adulthood.

Second, our model allowed us to clarify the nature and timing of the environmental influences on last-year MD. The large majority of these effects are wave-specific – that is, were important only over the short time period of the last year. Excluding the impact of measurement error (which is typically confounded in most), 54% of the liability to last-year MD resulted from environmental influences. Of this aggregate effect, 17% reflected the enduring impact of environmental stressors experienced prior to the first assessment wave that impacted on MD risk over the entire period of the study. However, 83% of this effect resulted from environmental exposures over the last year whose influence on MD risk was short-term and had entirely attenuated by the next interview wave. Prior studies of this twin cohort provide some insight into the nature of these experiences. Consistent with other studies (Bebbington *et al.* 1981; Surtees *et al.* 1986; Brown & Harris, 1978), we found that typical stressful life events (such as assault, job loss, marital problems, and major illnesses) produce robust elevations in risk for MD that typically last only 1–2 months (Kendler *et al.* 1998). What might be responsible for the enduring environmental effects? Tentative answers come from a qualitative study of MZ twins discordant for lifetime MD conducted in this sample (Kendler & Halberstadt, 2013). In these pairs, a joint life-history interview revealed four classes of explanations for the discordance in the experiences of the affected *v.* well twin: (i) severe traumatic events, (ii) chronic relationship difficulties beginning with a traumatic romantic break-up, (iii) repeated, severe occupational difficulties, and (iv) life-changing injuries (Kendler & Halberstadt, 2013). Each of these groups of experiences likely contributed to the enduring environmental. We also found that, as in other many other samples (Fergusson & Mullen, 1999; Nelson *et al.* 2002), childhood sexual abuse could produce a life-long enduring vulnerability to MD (Kendler *et al.* 2000).

Third, the heritability of MD was stable over the 8 years of this study. This result, interesting in its own

https://doi.org/10.1017/S0033291717000277 Published online by Cambridge University Press

right, also provides information about the time course of the most common environmental exposures that predispose to MD in adult life: stressful life events (Brown & Harris, 1978; Kendler *et al.* 1998). Important for our point is consistent evidence that exposure to most classes of events are only modestly correlated in twin pairs (Kendler *et al.* 1993*a*; Middeldorp *et al.* 2005; Boardman *et al.* 2011). If these stressors produced long-lasting increases in the liability to MD, this should lead to declining twin resemblance for MD over time in both MZ and DZ twin pairs. This, in turn, would decreases in heritability. That we see no such effect is consistent with studies showing that the impact of typical stressful live events on risk for MD are short-lived (Brown & Harris, 1978; Bebbington *et al.* 1981; Surtees *et al.* 1986; Kendler *et al.* 1998). To be clear, we cannot rule out some events in middle adulthood which produce enduring effects on depression liability. While they likely occur, they are probably too rare for us to detect in our analyses given the available sample size.

Fourth, we showed a large increase in heritability of MD moving from that observed for a 1-year period (33%) compared to the latent stable estimate acting across all four waves (78%). To be clear, these two estimates are not directly comparable as they apply to two different constructs for 'MD'. In the first, we examine only the presence or absence of a depressive episode in a 1-year time period. That is, we analyze 'observed MD'. In the second, we examine the latent liability to MD as measured over four different time periods. That is, we analyze the 'liability to MD'. The increased heritability of the latter measure results directly from this model's ability to reduce the impact of short-term environment and errors of measurement thereby increasing the proportion of liability to MD due to genetic factors.

The use of multiple measures to increase heritability is an old technique in statistical genetics first developed by animal and plant breeders (Falconer, 1989). In discussing its application milk yield in cattle and litter size in mice, Falconer notes that this method works because of

> increasing the number of measurements reduces the amount of variance due to special [aka occasion-specific] environment … When the repeatability is low, multiple measurements may lead to a worthwhile gain in accuracy [of genetic prediction] (Falconer, 1989, pp. 139–144).

He presents formulas for the expected increases in estimated heritability as a function of the repeatability of the trait. For last-year MD (with a 'repeatability' of ~+0.40), this formula suggests substantial gains for up to five or six independent measurements.

Our results address another concern with heritability estimates of lifetime MD – the possible impact of problems with long-term recall. Critics can argue, given well-demonstrated problems in the accuracy and bias of long-term memory (Coughlin, 1990; Andrews *et al.* 1999; Kendler *et al.* 2001*b*; Wells & Horwood, 2004; Moffitt *et al.* 2010), that the heritability of lifetime MD might be substantially 'contaminated' by such recall effects. For example, a lifetime history of MD might assess, at least in part, a 'plaintive set' of cognitive biases rather than the true occurrence of MD episodes. Our results argue against this hypothesis both because our assessments were based on much shorter time periods of recall (1 year) and we employed interviews that aided the respondent in temporally structuring their last year experiences, an approach shown to improve the accuracy of human recall (Belli, 1998).

We have previously demonstrated that two independent reports of lifetime MD also increases heritability substantially (Kendler *et al.* 2001*a*). These and the present results have implications for molecular genetic studies of MD which suffer from problems of low power due to small effect-size genetic variants (CONVERGE Consortium, 2015; Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium, 2012). When dealing with low heritability phenotypes like MD with only moderate stability over time especially in general population samples, multiple independent measurements is one potential way to increase, potentially substantially, the 'genetic signal'.

Limitations

Our findings should be considered in the light of five potential methodological limitations. First, our subjects are all female, white twins born in Virginia. While our sample has been shown to be broadly representative of the US population for a range of relevant traits (Kendler & Prescott, 2006), the degree to which our present findings extrapolate to other samples and to males is uncertain. With respect to the rates of psychopathology, twins are probably representative of the general population (Kendler *et al.* 1996; Kendler & Prescott, 2006). Our 1-year prevalence rates for MD in females were mid-way between those reported in the National Comorbidity Survey (12.9%) (Kessler *et al.* 1994) and the National Epidemiologic Survey on Alcoholism and Related Conditions (6.9%) (Hasin *et al.* 2005).

Second, in examining a community-based sample, we obtain cases of MD that are representative of the general population. However, their average level of severity and patterns of recurrence and comorbidity likely differ from those seen in clinical settings, especially in-patient treatment. So our findings may not extrapolate to clinically ascertained cases of MD in particular because more severe forms of MD and those

https://doi.org/10.1017/S0033291717000277 Published online by Cambridge University Press

with treatment tend to be more reliability reported (Bromet *et al.* 1986; Foley *et al.* 1998).

Third, because the power of our sample size of twins was only moderate, we cannot rule out modest changes in genetic influences on MD over time or small shared environmental effects (Neale *et al.* 1994).

Fourth, there have been traditional concerns about biases in the twin method that might increase heritability estimates, especially the 'equal environmental assumption' – that MZ and DZ twins have approximately equally similar childhood environments. This assumption has been tested several times in this sample and shown for MD and other psychiatric disorders to be well justified (Kendler & Prescott, 2006).

Fifth, our estimates of the role of environmental risk factors (E) in our analyses were all inferred from the patterns of resemblance for MD across time and in MZ and DZ twin pairs. This has the important advantage of cleanly separating sources of variance in liability based on our twin model assumptions. The disadvantage is that we cannot specify the actions of specific risk factors. We have, however, in previous large-scale path models in this sample, examined the action of both a wide variety of genetic, temperamental and environmental risk factors predicting onset of MD in the last year (Kendler *et al.* 2002) and the interrelationship on a month by month basis of a latent depressive vulnerability and aggregate stressful life events (Kendler & Gardner, 2016). These analyses provide complementary perspectives on the etiologic processes underlying risk for MD.

Summary

The latent liability to MD as assessed over four waves of measurement in middle-adulthood is highly heritable and developmentally stable. The large majority of the environmental risk for MD (83%) is occasion-specific and produces only transient elevations in MD risk. However, a small proportion of the environmental risk for MD (17%) arises from environmental experiences earlier in life that have an enduring impact on liability. As predicted from genetic theory, given the moderate stability of depressive episodes over adult life, heritability of MD is increased substantially by multiple measurements.

Acknowledgements

This project was supported by grant MH40828. The Virginia Twin Registry is supported by grant UL1RR031990. Dr Carol Prescott played a central role in the design and implementation of this twin study. Dr Michael Neale provided advice on the statistical modeling.

Declaration of Interest

None.

References

Akaike H (1987). Factor analysis and AIC. *Psychometrika* **52**, 317–332.

Andrews G, Anstey K, Brodaty H, Issakidis C, Luscombe G (1999). Recall of depressive episode 25 years previously. *Psychological Medicine* **29**, 787–791.

APA (1987). *Diagnostic and Statistical Manual of Mental Disorders, Revised Third Edition*. American Psychiatric Association: Washington, DC.

APA (2013). *Diagnostic and Statistical Manual of Mental Disorders: Fifth Edition, DSM-5*. American Psychiatric Association: Washington, DC.

Bebbington PE, Tennant C, Hurry J (1981). Adversity and the nature of psychiatric disorder in the community. *Journal of Affective Disorders* **3**, 345–366.

Belli RF (1998). The structure of autobiographical memory and the event history calendar: potential improvements in the quality of retrospective reports in surveys. *Memory* **6**, 383–406.

Boardman JD, Alexander KB, Stallings MC (2011). Stressful life events and depression among adolescent twin pairs. *Biodemography and Social Biology* **57**, 53–66.

Bromet EJ, Dunn LO, Connell MM, Dew MA, Schulberg HC (1986). Long-term reliability of diagnosing lifetime major depression in a community sample. *Archives of General Psychiatry* **43**, 435–440.

Brown GW, Harris TO (1978). *Social Origins of Depression: A Study of Psychiatric Disorder in Women London*. The Free Press (a division of Macmillan Publishing Company): NY, USA.

Cohen J (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* **XX**, 37–46.

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

Coughlin SS (1990). Recall bias in epidemiologic studies. *Journal of Clinical Epidemiology* **43**, 87–91.

Falconer DS (1989). *Introduction to Quantitative Genetics*, 3 edn. Wiley: New York.

Fergusson DM, Mullen PE (1999). *Childhood Sexual Abuse: An Evidence Based Perspective*. Sage Publications, Inc.: Thousand Oaks, CA.

Foley DL, Neale MC, Kendler KS (1998). Reliability of a lifetime history of major depression: implications for heritability and co-morbidity. *Psychological Medicine* **28**, 857–870.

Guerrini I, Cook CC, Kest W, Devitgh A, McQuillin A, Curtis D, Gurling HM (2005). Genetic linkage analysis supports the presence of two susceptibility loci for alcoholism and heavy drinking on chromosome 1p22.1-11.2 and 1q21.3-24.2. *BMC Genetics* **6**, 11.

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. *Archives of General Psychiatry* **62**, 1097–1106.

Kendler KS, Bulik CM, Silberg J, Hettema JM, Myers J, Prescott CA (2000). Childhood sexual abuse and adult psychiatric and substance use disorders in women: an epidemiological and cotwin control analysis. *Archives of General Psychiatry* **57**, 953–959.

Kendler KS, Fiske A, Gardner CO, Gatz M (2009). Delineation of two genetic pathways to major depression. *Biological Psychiatry* **65**, 808–811.

Kendler KS, Gardner CO (2016). Depressive vulnerability, stressful life events and episode onset of major depression: a longitudinal model. *Psychological Medicine* **46**, 1865–1874.

Kendler KS, Gardner CO, Lichtenstein P (2008). A developmental twin study of symptoms of anxiety and depression: evidence for genetic innovation and attenuation. *Psychological Medicine* **38**, 1567–1575.

Kendler KS, Gardner CO, Neale MC, Prescott CA (2001*a*). Genetic risk factors for major depression in men and women: similar or different heritabilities and same or partly distinct genes? *Psychological Medicine* **31**, 605–616.

Kendler KS, Gardner CO, Prescott CA (2001*b*). Are there sex differences in the reliability of a lifetime history of major depression and its predictors? *Psychological Medicine* **31**, 617–625.

Kendler KS, Gardner CO, Prescott CA (2002). Toward a comprehensive developmental model for major depression in women. *American Journal of Psychiatry* **159**, 1133–1145.

Kendler KS, Gatz M, Gardner C, Pedersen N (2006). A Swedish National Twin Study of lifetime major depression. *American Journal of Psychiatry* **163**, 109–114.

Kendler KS, Halberstadt LJ (2013). The road not taken: life experiences in monozygotic twin pairs discordant for major depression. *Molecular Psychiatry* **18**, 975–984.

Kendler KS, Karkowski LM, Prescott CA (1998). Stressful life events and major depression: risk period, long-term contextual threat, and diagnostic specificity. *Journal of Nervous and Mental Disease* **186**, 661–669.

Kendler KS, Kuhn JW, Vittum J, Prescott CA, Riley B (2005). The interaction of stressful life events and a serotonin transporter polymorphism in the prediction of episodes of major depression: a replication. *Archives of General Psychiatry* **62**, 529–535.

Kendler KS, Neale M, Kessler R, Heath A, Eaves L (1993*a*). A twin study of recent life events and difficulties. *Archives of General Psychiatry* **50**, 789–796.

Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ (1993*b*). A longitudinal twin study of 1-year prevalence of major depression in women. *Archives of General Psychiatry* **50**, 843–852.

Kendler KS, Pedersen NL, Farahmand BY, Persson PG (1996). The treated incidence of psychotic and affective illness in twins compared with population expectation: a study in the Swedish Twin and Psychiatric Registries. *Psychological Medicine* **26**, 1135–1144.

Kendler KS, Prescott CA (1999). A population-based twin study of lifetime major depression in men and women. *Archives of General Psychiatry* **56**, 39–44.

Kendler KS, Prescott CA (2006). *Genes, Environment, and Psychopathology: Understanding the Causes of Psychiatric and Substance Use Disorders*, 1st edn. Guilford Press: New York.

Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, Wittchen HU, Kendler KS (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. *Archives of General Psychiatry* **51**, 8–19.

Kruijshaar ME, Barendregt J, Vos T, de GR, Spijker J, Andrews G (2005). Lifetime prevalence estimates of major depression: an indirect estimation method and a quantification of recall bias. *European Journal of Epidemiology* **20**, 103–111.

Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium (2012). A mega-analysis of genome-wide association studies for major depressive disorder. *Molecular Psychiatry* **18**, 497–511.

Middeldorp CM, Cath DC, Vink JM, Boomsma DI (2005). Twin and genetic effects on life events. *Twin Research and Human Genetics* **8**, 224–231.

Moffitt TE, Caspi A, Taylor A, Kokaua J, Milne BJ, Polanczyk G, Poulton R (2010). How common are common mental disorders? Evidence that lifetime prevalence rates are doubled by prospective versus retrospective ascertainment. *Psychological Medicine* **40**, 899–909.

Neale MC, Boker SM, Xie G, Maes HH (2003). *Mx: Statistical Modeling*, 6th edn. Box 980126. Dept. of Psychiatry, Virginia Commonwealth University Medical School: Richmond, VA.

Neale MC, Eaves LJ, Kendler KS (1994). The power of the classical twin study to resolve variation in threshold traits. *Behavior Genetics* **24**, 239–258.

Nelson EC, Heath AC, Madden PA, Cooper ML, Dinwiddie SH, Bucholz KK, Glowinski A, McLaughlin T, Dunne MP, Statham DJ, Martin NG (2002). Association between self-reported childhood sexual abuse and adverse psychosocial outcomes: results from a twin study. *Archives of General Psychiatry* **59**, 139–145.

Nivard MG, Dolan CV, Kendler KS, Kan KJ, Willemsen G, van Beijsterveldt CE, Lindauer RJ, van Beek JH, Geels LM, Bartels M, Middeldorp CM, Boomsma DI (2014). Stability in symptoms of anxiety and depression as a function of genotype and environment: a longitudinal twin study from ages 3 to 63 years. *Psychological Medicine*, 1–11.

Patten SB (2003). Recall bias and major depression lifetime prevalence. *Social Psychiatry and Psychiatric Epidemiology* **38**, 290–296.

Silberg JL, Rutter M, Eaves L (2001). Genetic and environmental influences on the temporal association between earlier anxiety and later depression in girls. *Biological Psychiatry* **49**, 1040–1049.

Spitzer RL, Williams JBW (1985). *Structured Clinical Interview for DSM-III-R* (*SCID*). Biometrics Research Department, New York State Psychiatric Institute: New York.

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

Surtees PG, Miller PM, Ingham JG, Kreitman NB, Rennie D, Sashidharan SP (1986). Life events and the onset of

affective disorder. A longitudinal general population study. *Journal of Affective Disorders* **10**, 37–50.

Waszczuk MA, Zavos HM, Gregory AM, Eley TC (2016). The stability and change of etiological influences on depression, anxiety symptoms and their co-occurrence across adolescence and young adulthood. *Psychological Medicine* **46**, 161–175.

Wells JE, Horwood LJ (2004). How accurate is recall of key symptoms of depression? A comparison of recall and longitudinal reports. *Psychological Medicine* **34**, 1001–1011.