Chapter 2

Key risk factors for onset and recurrence of Major Depression: results from Lifelines, a large representative population cohort

Lian Beijers, Klaas J. Wardenaar, Trynke R. de Jong, Robert A. Schoevers, Hanna M. van Loo

Submitted.

Abstract

Background

Major Depression (MD) is a multifactorial disorder with a substantial disease burden, which often has recurrent episodes. Prevention requires intricate knowledge about the key risk factors of MD. This study aims to determine which of the many previously identified risk factors are most important for predicting first onset and recurrence of MD by investigating multivariable models in a longitudinal population study with sufficient sample size.

Methods

The Dutch Lifelines study is a large longitudinal representative population cohort. We selected 21 risk factors for MD, such as socio-demographic variables, neuroticism, family history, stressful life events, childhood trauma, health behaviors, general health status, and metabolic and inflammatory markers. MD onset and recurrence were measured in two follow-up waves (n = 42,724). Relative importance analysis was used to identify key risk factors for MD onset and recurrence.

Results

A family history of anxiety and depression, childhood trauma, higher neuroticism, female sex, younger age, chronic stress, lower physical quality of life and current anxiety disorders were all key risk factors for MD onset. Most key risk factors for MD onset also predicted MD recurrence. Comorbid anxiety and female sex predicted first onset only, whereas lower education levels specifically predicted recurrence.

Conclusion

We identified several key risk factors relevant for onset and recurrence of MD, which could guide primary as well as secondary prevention programs. Our findings suggest that educational inequality plays a role in the course of the disorder, and emphasizes the importance of screening for MD among family members of depressed individuals.

Introduction

Major depression (MD) is among the most prevalent mental disorders worldwide, and is associated with a substantial burden.[1–3] This burden is highest in patients who have a course with chronic or recurrent episodes.[4–9] Since the 1970s, increasing numbers of people in Western countries are receiving psychotherapy or pharmacotherapy for the disorder, yet epidemiological data do not indicate a drop in MD prevalence.[10] The effectiveness of current therapies relative to placebo is modest, and other approaches are necessary to address the public health burden of MD.[11–13] Preventive interventions for both first onsets and recurrent episodes of MD seem like a promising avenue.[14,15] Selective prevention, which targets individuals or subgroups that are at high risk of MD, is thought to be more effective compared to universal interventions, which target the whole population, regardless of risk status.[16,17] Risk factors that cannot be changed with interventions (e.g., gender and age) can still be used to determine which people are at highest risk - interventions to increase resilience in these people specifically may potentially reduce the prevalence of MD.[15] Identifying key risk factors for MD will therefore help us determine which interventions are most likely to succeed in preventing MD episodes.[14,15,17]

A plethora of risk factors for MD have been identified. For example, although the exact mechanisms are unclear, it is well known that demographic factors such as younger age and female sex are risk factors for MD.[18–20] We also know that depression runs in families.[20–22] Based on twin studies, the heritability of MD is estimated to be about 37%[23–26], but rearing experiences are estimated to contribute just as much to trans-generational transmission of MD risk as genetic risk.[27–30] A number of physiological problems such as dysregulations of neuroendocrine[31,32], metabolic[33,34], and inflammatory[34,35] systems and the presence of somatic disorders[36–39] have also been related to higher MD risk, as have several aspects of personality, especially neuroticism.[20,21] Finally, environmental risk factors such as traumatic life experiences[40,41], socioeconomic status[20,42,43] and lifestyle factors like the consumption of alcohol[44,45] and tobacco[46] or the amount of physical movement[47,48] also contribute to MD risk.

Most of this knowledge comes from studies investigating single risk factors, or risk factor domains, and the differences in sample characteristics (e.g., different distributions of sex, age, ethnicity, or socio-economic status) and methodology (e.g., self-report vs. clinician-rated, different time intervals) make it difficult to compare the effects of different risk factors.[49,50] Furthermore, measuring effects of individual risk factors in different samples increases the risks associated with unidentified confounding or mediation, because many of the aforementioned risk factors interact.[51] For example, lower education

levels might predict the onset of MD directly, but this effect could be explained by lower income. Measuring the risks associated with both variables in independent samples might lead to the conclusion that there is a similar effect when in fact it is the same variance that is being explained by both risk factors. Furthermore, to investigate the directionality of the relationship between MD and these factors, longitudinal data are needed from a large number of participants. In summary, in order to identify key risk factors for MD, it is crucial that multiple risk factors are investigated in concert, using a multivariable model in a longitudinal population study with sufficient sample size.

One method to identify key risk factors for MD is relative importance analysis, which calculates the proportion of explained variance of each risk factor, by comparing the statistical fit of possible models including that risk factor to that of all possible models.[52,53] However, this type of analysis has never been applied to risk factors of MD before, since population studies that include sufficient numbers of risk factors are rare, and the computational power required for this type of analysis increases exponentially with each additional risk factor. Most previous studies reported a number of models investigating individual risk factors instead, or opted to specify a single multivariable model including all risk factors that were significant in univariable analyses.[54–56]

In order to investigate the key risk factors of MD onset and recurrence, we performed Relative Importance Analysis using the Lifelines cohort, a large longitudinal population study.[57] From this rich dataset, we selected known risk factors for MD, such as socio-demographic variables, neuroticism, family history, stressful life events, childhood trauma, health behaviors, general health status, and metabolic and inflammatory markers.[21,49,58] Our first aim was to investigate which of these risk factors were most important for predicting onset of the first MD episode. Our second aim was to investigate whether similar or different key risk factors predict recurrence of MD.

Methods

This study was preregistered on the Open Science Framework in February 2020 (https://osf.io/7bptq/).

The Lifelines Cohort Study

The Lifelines Cohort Study is a large population-based cohort study and biobank that is used for research on complex interactions between environmental, phenotypic and genomic factors in the development of chronic diseases and healthy ageing.[59] Between 2006 and

2013, inhabitants of the northern part of the Netherlands were invited to participate through their general practitioners, in a three-generation design. At wave 1, data were collected for 167,729 participants, aged 6 months to 93 years. Participants visited one of the Lifelines research sites for a physical examination, including lung function, ECG and cognition tests, and completed extensive questionnaires. Fasting blood and 24-h urine samples were processed on the day of collection and stored at -80 ˚C in a fully automated storage facility. The baseline questionnaire consisted of two parts containing questions on, among other topics, demographics, health status, lifestyle, and psychosocial aspects. We made use of the baseline measurement (wave 1, 2007-2013) the first follow-up wave (wave 2, 2014-2017), and the Lifetime Depression Assessment Self-report[60], an add-on online questionnaire administered in 2018 for the Biobanks Netherlands Internet Collaboration project. We included all subjects who participated at wave 1 and the 2018 add-on survey (n = 42,724). The mean intervals between waves 1-2 and waves 2 and the 2018 add-on survey were 3.88 (SD = 1.19) and 2.83 (SD = 1.06) years, respectively.

Baseline predictors

A total of 21 putative risk factors were included in the analyses (see Online Supplement). These risk factors were classified into nine major risk domains: (1) Sex and age, (2) Current social and economic environment (education and income level, unfavourable work status, number of social contacts), (3) Health behaviors (physical movement, current smoking status, drinking alcohol), (4) Somatic health (physical quality of life (QoL), cardiovascular problems, cancer, inflammatory disorders, low-grade inflammation, metabolic syndrome) (5) Anxiety disorders (number of current diagnoses), (6) Family history of anxiety and depression, (7) Personality (neuroticism), (8) Early adverse life events (childhood trauma), (9) Acute and chronic stress. Most of these predictors were assessed at baseline using self-report instruments. The number of anxiety disorders was determined by trained research assistants using the Mini-International Neuropsychiatric Interview (MINI). Low-grade inflammation was measured through serum levels of high-sensitivity C-reactive protein (CRP). The metabolic syndrome diagnosis included measurements of waist circumference and blood pressure, as well as serum levels of glucose, high-density lipoprotein (HDL) cholesterol, and triglycerides. Childhood trauma was not measured at baseline, but in a separate questionnaire that took place an average of 5.6 years after the baseline measurement (SD = 1.3). Family history of anxiety and depression was assessed at the 2018 survey.

Outcomes

For the first two waves the MINI was used to measure MD in the past two weeks.[61] The 2018 survey measured lifetime MD status, age of onset, and the presence of an episode in the past year using the LIfetime Depression Assessment Self-report.[60] Both questionnaires are validated instruments assessing MD according to DSM-IV-TR criteria.[62]

The incidence rate was calculated as the number of new cases per 100 person years.[63] We divided the number of new onsets between wave 1 and the 2018 survey by the cumulative number of years at risk during this period. Among incident cases, we counted the time at risk as the age of onset minus the age at wave 1. We assumed that the average point when a new case emerges lies halfway through the year, so we subtracted half a year from this number.[55,64,65]

To study predictors of the onset of depression, we used an outcome that contrasted all subjects with a first onset of MD between wave 1 and the 2018 survey (i.e., MD not present at wave 1 and age of onset after wave 1) with all subjects who did not qualify for a MD diagnosis at any wave, nor reported lifetime MD. To study predictors of the recurrence of depression we selected all individuals at risk of recurrence, i.e., with at least one episode of MD before wave 1, but not at wave 1. We contrasted subjects with a new episode at wave 2 and/or 3 ('recurrence') with all subjects without episodes at waves 2 or 3 ('non-recurrence').

Statistical analysis

All analyses were performed in *R_3.5.2*.[66]

Missing data handling

Multiple Imputation by Chained Equations was performed on the complete dataset using R-package *mice_3.8.0*.[67] Ten imputed datasets were used and all estimates were pooled across the datasets.

Multicollinearity

We investigated the correlations between the risk factors (Supplementary Figure 1). Correlations higher than 0.2 were observed in 26 out of 253 possible combinations of risk factors, although none were higher than 0.8, so we did not exclude any predictors for reasons of redundancy or multicollinearity.

Relative importance analysis

To gain more insight into the contributions of individual risk factors to the outcomes of interest, we performed relative importance analyses using the R-package *MuMIn_1.43.17*.[68] Each potential risk factor was investigated as an independent variable in a univariable logistic regression analysis with either onset or recurrence as the dependent variable. In the first step of relative importance analysis, we ran multivariable logistic regression models with all possible combinations of significant risk factors from the univariable analyses. In the second step, the importance value for a particular risk factor, which can be interpreted as the probability that a risk factor will be included in the best model, was calculated by summing the Akaike weights for the models in which the risk factor appears, and dividing this number by the sum of the Akaike weights of all models (see Supplementary Methods).[52] The importance value was calculated in every imputed dataset, and the final model combined all risk factors for which the average importance value was over 50%. In order to facilitate future meta-analyses, we also ran a multivariate model with all risk factors that were significant in the univariable analyses.

Results

At baseline, n = 34,694 subjects had never experienced MD. Of this group, 6.9% (n = 2,046) developed at least one episode of MD before the 2018 survey (see Figure 1). This corresponds with 10.5 new cases per 100 person-years. At baseline, there were 6826 individuals at risk of recurrence, 34.1% (n = 2,326) of which developed a new episode at subsequent measurement points. See Table 1 for the baseline characteristics of both samples.

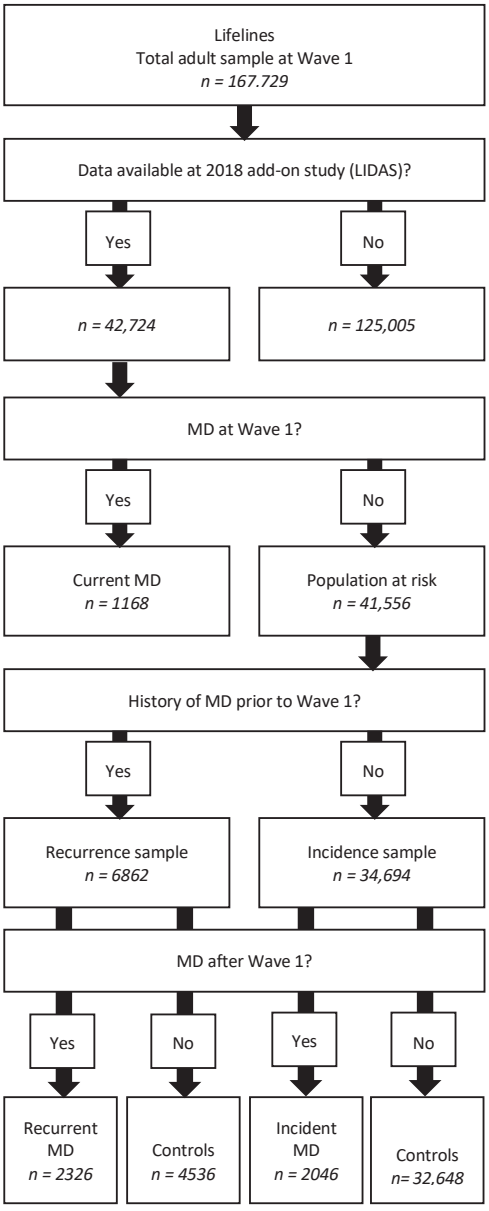


Figure 1. Sample selection

Flowchart of data selection for the incidence and recurrence samples.
LIDAS, LIfetime Depression Assessment Self-report; MD, Major Depression

Table 1. Baseline characteristics

	Onset		Recurrence	
	n	Value	*n*	Value
Sex, % female (SE)	24910	57.74 (0.31)	6659	72.95 (0.54)
Age, mean (SD)	24910	46.77 (12.19)	6659	45.69 (10.92)
Social and economic environment				
Education[1], % (SE)	24431		6533	
Low		25.86 (0.28)		21.54 (0.51)
Intermediate		39.44 (0.31)		41.37 (0.61)
High		34.71 (0.3)		37.09 (0.6)
Unfavourable job status[2], % (SE)	24894	5.87 (0.15)	6653	13.06 (0.41)
Income[3], % (SE)	21648		5937	
Low		16.28 (0.25)		20.53 (0.52)
Intermediate		55.80 (0.34)		54.24 (0.65)
High		27.92 (0.30)		25.23 (0.56)
N contacts past two weeks, mean (SD)	24659	19.22 (18.5)	6578	16.72 (15.96)
Health behaviors				
Physically active[4], % (SE)	23266	41.8 (0.32)	6275	47.41 (0.63)
Smoking status, % (SE)	24252		6504	
Non-smoker		46.61 (0.32)		39.48 (0.61)
Former smoker		37.17 (0.31)		39.45 (0.61)
Current smoker		16.21 (0.24)		21.06 (0.51)
Alcohol consumption, % (SE)				
Heavy drinker (≥ 6 drinks per drinking day)	23450	7.73 (0.17)	6366	7.37 (0.33)
Binge drinker (≥ 2 drinks per day on average)	23453	4.49 (0.14)	6366	3.82 (0.24)
Somatic health				
Physical quality of life (RAND-36), mean (SD)	24268	49.71 (7.57)	6459	49.87 (9.00)
Cancer[5], % (SE)	24881	1.01 (0.06)	6649	1.26 (0.14)
Cardiovascular problems[6], % (SE)	24894	2.02 (0.09)	6653	2.24 (0.18)
Metabolic syndrome (NCEP ATPIII criteria), % (SE)	24586	31.65 (0.30)	6515	33.84 (0.59)
Inflammatory disorders[7], % (SE)	24604	9.91 (0.19)	6606	13.02 (0.41)
Low-grade inflammation (CRP), mean (SD)	10348	2.43 (4.29)	2477	2.6 (4.76)
MD characteristics				
MD at wave 1 (past 2 weeks), % (SE)	24910	0 (N.A.)	6659	0 (N.A.)
MD at wave 2 (past 2 weeks), % (SE)	24355	0.39 (0.04)	4818	7.35 (0.38)
MD at the 2018 survey (past year), % (SE)	24910	4.13 (0.13)	6659	30.86 (0.57)
Lifetime MD (measured at the 2018 survey), % (SE)	24910	7.95 (0.17)	6659	100 (N.A.)
MD age of onset, mean (SD)	2908	41.44 (13.2)	6659	28.46 (11.67)
N anxiety disorders, mean (SD)	24910	0.05 (0.23)	6659	0.18 (0.47)
Family history of depression/anxiety, % (SE)	24795	48.47 (0.32)	6618	79.84 (0.49)
Neuroticism (NEO PI-R), mean (SD)	6550	-0.14 (0.94)	2201	0.62 (1.03)
Childhood trauma (CTQ), % (SE)	19475	21.97 (0.3)	4815	41.43 (0.71)
Acute and chronic stress				
Threatening events (LTE), mean (SD)	24397	1.04 (1.24)	6421	1.47 (1.53)
Chronic stress (LDI), mean (SD)	24395	2.04 (2.06)	6483	3.67 (2.71)

Sample characteristics at wave 1, based on complete data.

[1]Highest completed education: Low = junior general secondary education (mavo/vmbo-t) or lower, or no education; Intermediate = secondary vocational education (mbo), senior general secondary education (havo, vwo, hbs, mms); High = higher vocational education (hbo) or university.

[2]Unfavorable working conditions: being unemployed/looking for work, disabled for work, or on welfare.

[3]Net household income: low (< 1100), intermediate (1100-1899) and high (≥ 1900).

[4]The Dutch Movement Norm classifies physical activity as sufficient when participants report being active for at least half an hour on at least five days per week.

[5]Self-reported life time heart attack, aneurysm in aorta, heart failure, or stroke

[6]Self-reported cancer of any type, current

[7]Self-reported lifetime asthma, ulcerative colitis, rheumatoid arthritis, Crohn's disease, or celiac disease

CRP, C-reactive protein; CTQ, Childhood Trauma Questionnaire; LDI, Long-term Difficulties Index; LTE, List of Threatening Events; MD, Major Depression; NEO PI-R, Revised NEO Personality Inventory; NCEP ATPIII, National Cholesterol Education Programs Adults Treatment Panel III; RAND-36, Research and Development-36 (Dutch version of Medical Outcomes Study 36-Item Short Form Health Survey)

Key risk factors for new onset of MD

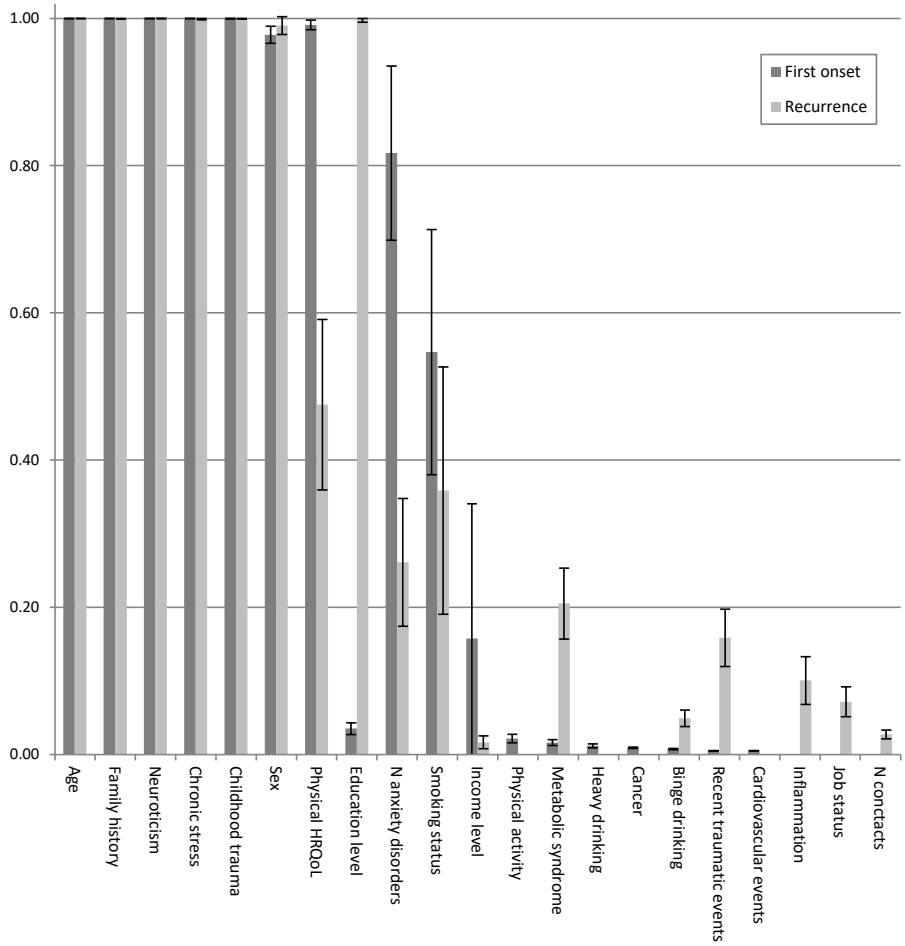
The relative importance of age, sex, family history of depression, neuroticism, childhood trauma, chronic stress, and physical QoL approached 100% (Figure 2). The relative importance of the number of anxiety disorders at wave 1 was also high (82%), whereas the relative importance of smoking behavior was 55%. The relative importance of education level, income, physical activity, the metabolic syndrome, drinking behavior, cancer, cardiovascular events, recent traumatic events, job status, and number of social contacts was low (1-16%). Similar risk factors remained significant in the multivariable model including all the significant risk factors from the univariable analyses (Supplementary Table 2). Family history was one of the strongest dichotomous predictors of incidence (odds ratio (OR) = 1.90) (Table 2). Similar risk (OR) was associated with having two additional anxiety diagnoses at wave 1, having problems with three additional domains of chronic stress (e.g., finances, health situation, or relationships), a 52-point difference on the Neuroticism dimension of the Revised Neuroticism-Extraversion-Openness Personality Inventory[69] (range 48-240), a 14-year age difference (n.b., younger people are at higher risk), or having a more than four standard deviations higher score on physical QoL. Childhood abuse, current smoking, and sex were weaker predictors (OR ~ 1.30).

Key risk factors for recurrent MD

Relative importance analyses showed that risk factors were equally likely to be included in the multivariable models for onset and recurrence, except for the following differences. Education level had a high importance value for MD recurrence, but not for onset (99.8% vs. 3.5%). Physical QoL and the number of anxiety disorders at wave 1 were important risk factors for MD onset, but their importance values for recurrence were only 48% and 26%, respectively. Family history was still one of the strongest dichotomous predictors

(OR = 1.40) in the final multivariable model (see Table 2), but lower education level, which was not included for onset, was the strongest predictor for recurrence of MD (OR = 1.52). Finally, male sex rather than female sex was a significant risk factor for recurrence.

Figure 2. Relative importance of each variable averaged over the imputed datasets



Relative weights for all risk factors that were significant in in univariable analyses for onset and recurrence of MD, averaged over ten imputed datasets, including error bars indicating the standard deviation.

Table 2. Final multivariable binomial regression models for onset and recurrence of MD

	Onset			Recurrence		
Intercept	0.24 (0.20-0.30)	-13.61	< 0.001*	0.72 (0.54-0.96)	-2.22	0.03*
Risk factor	**OR (95% CI)**	**F**	**P**	**OR (95% CI)**	**F**	**P**
Sex (female)	1.25 (1.13-1.39)	4.40	< 0.001*	0.74 (0.66-0.84)	-4.93	< 0.001*
Age (z-transformed)	0.57 (0.54-0.60)	-21.01	< 0.001*	0.76 (0.71-0.81)	-8.64	< 0.001*
Social and economic environment						
Education[1]						
Low	-	-	-	1.56 (1.35-1.80)	6.07	< 0.001*
Intermediate	-	-	-	1.29 (1.15-1.46)	4.20	< 0.001*
Health behaviors						
Smoking status						
Former smoker	1.04 (0.92-1.16)	0.60	0.55	-	-	-
Current smoker	1.31 (1.17-1.48)	4.48	< 0.001*	-	-	-
Somatic health						
Physical quality of life (RAND-36)	1.16 (1.09-1.24)	4.43	< 0.001*	-	-	-
N comorbid anxiety disorders (count 0-4)	1.33 (1.14-1.55)	3.65	< 0.001*	-	-	-
Family history of depression/anxiety	1.90 (1.72-2.09)	12.66	< 0.001*	1.40 (1.22-1.61)	4.76	< 0.001*
Neuroticism (NEO PI-R, z-transformed)	1.29 (1.21-1.37)	8.53	< 0.001*	1.28 (1.20-1.36)	7.71	< 0.001*
Childhood trauma (CTQ)	1.36 (1.21-1.54)	4.99	< 0.001*	1.31 (1.17-1.47)	4.63	< 0.001*
Chronic stress (LDI, z-transformed)	1.26 (1.20-1.33)	9.01	< 0.001*	1.16 (1.11-1.22)	5.78	< 0.001*

Multivariable logistic regression for onset of MD pooled over ten imputed datasets, using all risk factors for which the average importance value was over 50%. Risk factors are dichotomous absent vs. present, unless otherwise specified (i.e., count data, z-transformed continuous risk factors).

*p < 0.05

[1]Highest completed education: Low = junior general secondary education (mavo/vmbo-t) or lower, or no education; Intermediate = secondary vocational education (mbo), senior general secondary education (havo, vwo, hbs, mms); High = higher vocational education (hbo) or university.

CTQ, Childhood Trauma Questionnaire; LDI, Long-term Difficulties Index; NEO PI-R, Revised NEO Personality Inventory; RAND-36, Research and Development-36 (Dutch version of Medical Outcomes Study 36-Item Short Form Health Survey)

Discussion

This study aimed to identify key risk factors for new onsets of MD and compare these with risk factors for recurrence of MD. To this end, we performed relative importance analysis of a comprehensive collection of 21 potential risk factors that are known to be related to MD in a large representative general population sample. We found that age, sex, family history of depression/anxiety, neuroticism, childhood trauma, chronic stress, physical QoL, the number of anxiety disorders at wave 1, and smoking status were key predictors of onset. This means that in the context of these risk factors, other risk factors such as low amounts of physical movement, binge drinking or heavy drinking, and specific issues related to somatic health are of lesser importance. The key risk factors of MD recurrence were similar to those of onset, minus the number of anxiety disorders at wave 1, physical QoL, and smoking. Additionally, they included lower education levels. This risk factor was a stronger predictor for recurrence than family history, which was the strongest predictor for onset. Finally, women had a higher risk of onset, but this was not the case for recurrence. In summary, we have identified a number of possible targets for preventive interventions, including potentially important differences between those for first onset and recurrence of MD.

It is difficult to compare these findings with previous literature directly, because no previous studies have used relative importance analysis. Relative importance analysis is less likely to designate a risk factor as relevant compared to, for example, studies using standard multivariable approaches that present a final model including all risk factors that are significant in univariable analyses. Still, since there are no previous studies that have used the same approach, we will compare our findings with previous general population studies that investigated risk factors for onset and recurrence of MD.

Our findings regarding family history align with the results of the Baltimore Epidemiologic Catchment Area study, in which subjects with a first onset were twice as likely to have reported a parent with depression.[54] Most other general population studies investigating onset and recurrence of MD either excluded family history or used a different definition, precluding direct comparison.[55,65,70–75] Our results also correspond with earlier observations from clinical cohort studies[76–78], including a recent study showing that family history is a much stronger predictor of MD onset than polygenic risk of MD[79], and underline the potential of this risk factor as a target for screening and preventive interventions.[80–83] Since rearing experiences are thought to contribute just as much to trans-generational transmission of MD risk as genetic risk, targeted interventions to improve parenting skills in families with a history of MD are of paramount importance.[15,27–30]

Our findings also confirmed that the well-known gender gap in the prevalence of MD primarily relates to higher incidence in women, as women had a significantly

higher risk of MD onset (7.1% vs 4.4%), but a lower risk of MD recurrence (33.2% vs 35.9%).[54,55,65,75,84–87] This sex difference in MD onset arises in puberty, and is likely due to a combination of factors ranging from genetic and hormonal differences to heightened exposure to severe adversity and structural inequity.[87–89] However, whereas smaller studies often showed no effect of sex on MD recurrence[90–96], in our study the risk of MD recurrence was somewhat higher in men than in women, which calls for awareness of the importance of recurrence prevention in *both* men and women.

In contrast to most other longitudinal population studies, education was not a key risk factor for MD onset in our study, possibly because we included other, potentially confounding, risk factors such as age, income, and health behaviors.[54,55,65,74,75] However, we did find that education level was a strong risk factor for recurrence of MD, which is in line with a meta-analysis showing that education levels were a stronger risk factor for chronic course than for incidence.[97] Thus, strategies for tackling inequality in depression are needed, especially in relation to the course of the disorder.[98]

We also did not confirm a dose-response effect of number of anxiety disorders on MD recurrence[90–92,96], but our finding that this risk factor predicts MD onset is in line with earlier studies.[65,73,74,84] Still, the presence of any anxiety disorder after the MD index episode appears to be a relevant indicator for recurrence[92,96,99], and as such should still be monitored for.

Strengths and limitations

Strengths of the current study included the large general population sample, the longitudinal study design, the inclusion of both men and women and a wide age range, the high number of available risk factors, and the presence of thorough assessments with validated structured questionnaires that enabled us to investigate numerous risk factors for first onsets and recurrence. Furthermore, our use of relative importance analysis enabled us to investigate which risk factors were the most important for predicting onset and recurrence of MD, which is not possible with conventional model selection methods.[54–56,100] However, the results should also be interpreted in the context of several limitations.

First, although the number of included risk factors is large, some potential key risk factors have been excluded. For example, in order to increase the comparability with first onsets of MD, we did not include any risk factors related to the initial MD episodes, such as age of onset or duration, in the recurrence analyses.[96,99] Other potential key risk factors include variables related to the pathophysiology of MD (e.g., monoamine dysregulation[31], increased stress response[32], altered neurocircuitry[101,102]), but at the moment, the Lifelines sample only includes data related to inflammatory and metabolic dysregulations.

Second, there has been considerable discussion in the literature about how to define terms relating to the recurrence of MD.[103,104] Here, we assigned subjects suffering from MD prior to but not at wave 1 to the recurrent group when there was an episode at any of our post-baseline measurements. Unfortunately, due to missing information, we cannot be sure that the subjects assigned to the non-recurrent group were in remission/recovery in the whole period between the assessment waves.

Third, there might have been selective attrition in the sample due to MD or other factors such as higher age or lower education levels.[105–110] This might have led to an underestimation of the incidence and recurrence rates. It is difficult to ascertain the effects of selective attrition on our analyses because we cannot be sure which subjects dropped out due to developing MD after wave 1, which is a common limitation of longitudinal cohort studies.[107–109,111]

Finally, the onset and recurrence analyses were performed in different subsets of the Lifelines sample. Ideally, future waves of the Lifelines study could be used to investigate key risk factors for onset and recurrence in the same subjects to replicate the differences identified here. However, the sample for MD onset was already about three times larger than the recurrence sample, and this difference will only increase when onset and recurrence of MD are studied in the same subjects.[54] Differences in sample size produce differences in statistical power, meaning it is easier for weaker candidate risk factors to reach significance in univariable models for onset. However, this is especially problematic when this is the only criterion for inclusion in the final model, which was not the case here.

Conclusion

We identified a number of key risk factors relevant for population screening to identify subjects at risk of onset or recurrence of MD. For example, the risk of MD recurrence was higher in men than in women in our study, which calls for awareness of the importance of recurrence prevention in *both* men and women. Furthermore, the importance of lower education levels as a predictor for recurrence of MD suggests that strategies for tackling educational inequality in MD are needed, especially in relation to the course of the disorder. Finally, screening for MD among family members of depressed individuals may lead to more timely interventions. Future studies using relative importance analysis in similarly large samples are needed to confirm these results, as well as expand them to include other potential key risk factors.

References

1. Kessler, R. C. The costs of depression. *Psychiatr. Clin. North Am.* **35**, 1–14 (2012).
2. Whiteford, H., Ferrari, A. & Degenhardt, L. Global Burden Of Disease Studies: Implications For Mental And Substance Use Disorders. *Health Aff. (Millwood).* **35**, 1114–20 (2016).
3. World Health Organization. Investing in treatment for depression and anxiety leads to fourfold return. http://www.who.int/mediacentre/news/releases/2016/depression-anxiety-treatment/en/ (2016).
4. Greden, J. F. The burden of disease for treatment-resistant depression. *J. Clin. Psychiatry* **62**, 26–31 (2001).
5. Crown, W. H. *et al.* The impact of treatment-resistant depression on health care utilization and costs. *J. Clin. Psychiatry* **63**, 963–971 (2002).
6. Mauskopf, J. A. *et al.* Nonresponse, partial response, and failure to achieve remission: humanistic and cost burden in major depressive disorder. *Depress. Anxiety* **26**, 83–97 (2009).
7. Ivanova, J. I. *et al.* Direct and indirect costs of employees with treatment-resistant and non-treatment-resistant major depressive disorder. *Curr. Med. Res. Opin.* **26**, 2475–2484 (2010).
8. Lepine, B. A., Moreno, R. A., Campos, R. N. & Couttolenc, B. F. Treatment-Resistant Depression Increases Health Costs and Resource Utilization. *Rev. Bras. Psiquiatr.* **34**, 379–388 (2012).
9. Mrazek, D. A., Hornberger, J. C., Altar, C. A. & Degtiar, I. A review of the clinical, economic, and societal burden of treatment-resistant depression: 1996-2013. *Psychiatr. Serv.* **65**, 977–87 (2014).
10. Jorm, A. F., Patten, S. B., Brugha, T. S. & Mojtabai, R. Has increased provision of treatment reduced the prevalence of common mental disorders? Review of the evidence from four countries. *World Psychiatry* **16**, 90–99 (2017).
11. Turner, E. H., Matthews, A. M., Linardatos, E., Tell, R. A. & Rosenthal, R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N. Engl. J. Med.* **358**, 252–260 (2008).
12. Coe, R. It's the effect size, stupid. What effect size is and why it is important. in *British Educational Research Association annual conference* 1–18 (2002).
13. Cuijpers, P., Andersson, G., Donker, T. & Van Straten, A. Psychological treatment of depression: Results of a series of meta-analyses. *Nord. J. Psychiatry* **65**, 354–364 (2011).
14. Beshai, S., Dobson, K. S., Bockting, C. L. H. & Quigley, L. Relapse and recurrence prevention in depression: Current research and future prospects. *Clin. Psychol. Rev.* **31**, 1349–1360 (2011).
15. Ormel, J., Cuijpers, P., Jorm, A. F. & Schoevers, R. Prevention of depression will only succeed when it is structurally embedded and targets big determinants. *World Psychiatry* **18**, 111 (2019).
16. Werner-Seidler, A., Perry, Y., Calear, A. L., Newby, J. M. & Christensen, H. School-based depression and anxiety prevention programs for young people: A systematic review and meta-analysis. *Clinical Psychology Review* vol. 51 30–47 (2017).
17. van Zoonen, K. *et al.* Preventing the onset of major depressive disorder: A meta-analytic review of psychological interventions. *Int. J. Epidemiol.* **43**, 318–329 (2014).
18. Ferrari, A. J. *et al.* Global variation in the prevalence and incidence of major depressive disorder: A systematic review of the epidemiological literature. *Psychol. Med.* **43**, 471–481 (2013).
19. Patten, S. B. Incidence of major depression in Canada. *CMAJ* (2000).

20. Hardeveld, F., Spijker, J., De Graaf, R., Nolen, W. A. & Beekman, A. T. F. Prevalence and predictors of recurrence of major depressive disorder in the adult population. *Acta Psychiatr. Scand.* **122**, 184–191 (2010).

21. Burcusa, S. L. & Iacono, W. G. Risk for recurrence in depression. *Clin. Psychol. Rev.* **27**, 959–985 (2007).

22. Weissman, M. M. *et al.* Offspring of depressed parents: 30 years later. *Am. J. Psychiatry* **173**, 1024–1032 (2016).

23. Sullivan, P. F., Neale, M. C. & Kendler, K. S. Genetic epidemiology of major depression: Review and meta-analysis. *American Journal of Psychiatry* vol. 157 1552–1562 (2000).

24. Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).

25. Hyde, C. L. *et al.* Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat. Genet.* **48**, 1031–1036 (2016).

26. Howard, D. M. *et al.* Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. *Nat. Commun.* **9**, 1470 (2018).

27. Kendler, K. S., Ohlsson, H., Sundquist, J. & Sundquist, K. The Rearing Environment and Risk for Major Depression: A Swedish National High-Risk Home-Reared and Adopted-Away Co-Sibling Control Study. *Am. J. Psychiatry* **177**, 447–453 (2020).

28. Kendler, K. S., Ohlsson, H., Lichtenstein, P., Sundquist, J. & Sundquist, K. The Nature of the Shared Environment. *Behav. Genet.* **49**, 1–10 (2019).

29. Kendler, K. S., Ohlsson, H., Sundquist, K. & Sundquist, J. Sources of parent-offspring resemblance for major depression in a national Swedish extended adoption study. *JAMA Psychiatry* **75**, 194–200 (2018).

30. Weissman, M. M. Is Depression Nature or Nurture? Yes. *American Journal of Psychiatry* vol. 177 376–377 (2020).

31. Albert, P. R., Benkelfat, C. & Descarries, L. The neurobiology of depression-revisiting the serotonin hypothesis. I. cellular and molecular mechanisms. *Philos. Trans. R. Soc. B Biol. Sci.* **367**, 2378–2381 (2012).

32. Pariante, C. M. & Lightman, S. L. The HPA axis in major depression: classical theories and new developments. *Trends Neurosci.* **31**, 464–468 (2008).

33. Pan, A. *et al.* Bidirectional association between depression and metabolic syndrome: A systematic review and meta-analysis of epidemiological studies. *Diabetes Care* **35**, 1171–1180 (2012).

34. Milaneschi, Y., Lamers, F., Berk, M. & Penninx, B. W. J. H. Depression Heterogeneity and Its Biological Underpinnings: Toward Immunometabolic Depression. *Biological Psychiatry* vol. 88 369–380 (2020).

35. Valkanova, V., Ebmeier, K. P. & Allan, C. L. CRP, IL-6 and depression: A systematic review and meta-analysis of longitudinal studies. *J. Affect. Disord.* **150**, 736–744 (2013).

36. Clarke, D. M. & Currie, K. C. Depression, anxiety and their relationship with chronic diseases: A review of the epidemiology, risk and treatment evidence. *Med. J. Aust.* **190**, S54-60 (2009).

37. Katon, W. J. Epidemiology and treatment of depression in patients with chronic medical illness. *Dialogues Clin. Neurosci.* (2011).

38. Gaynes, B. N., Burns, B. J., Tweed, D. L. & Erickson, P. Depression and health-related quality of life. *J. Nerv. Ment. Dis.* **190**, 799–806 (2002).

39. Penninx, B. W. J. H., Milaneschi, Y., Lamers, F. & Vogelzangs, N. Understanding the somatic consequences of depression: biological mechanisms and the role of depression symptom profile. *BMC Med.* **11**, 129 (2013).

40. Jaworska-Andryszewska, P. & Rybakowski, J. K. Childhood trauma in mood disorders: Neurobiological mechanisms and implications for treatment. *Pharmacol. Reports* **71**, 112–120 (2019).

41. Wiersma, J. E. *et al.* The importance of childhood trauma and childhood life events for chronicity of depression in adults. *J. Clin. Psychiatry* **70**, 983 (2009).

42. Wang, J. L., Schmitz, N. & Dewa, C. S. Socioeconomic status and the risk of major depression: The Canadian national population health survey. *J. Epidemiol. Community Health* **64**, 447–452 (2010).

43. Meng, X. & D'Arcy, C. The projected effect of risk factor reduction on major depression incidence: A 16-year longitudinal Canadian cohort of the National Population Health Survey. *J. Affect. Disord.* **185**, 56–61 (2014).

44. Hartka, E. *et al.* A meta-analysis of depressive symptomatology and alcohol consumption over time. *Br. J. Addict.* **86**, 1283–1298 (1991).

45. Graham, K., Massak, A., Demers, A. & Rehm, J. Does the association between alcohol consumption and depression depend on how they are measured? *Alcohol. Clin. Exp. Res.* **31**, 78–88 (2007).

46. Fluharty, M., Taylor, A. E., Grabski, M. & Munafò, M. R. The Association of Cigarette Smoking With Depression and Anxiety: A Systematic Review. *Nicotine Tob. Res.* **1**, 3–13 (2017).

47. Mammen, G. & Faulkner, G. Physical activity and the prevention of depression: A systematic review of prospective studies. *Am. J. Prev. Med.* **45**, 649–657 (2013).

48. Hiles, S. A., Lamers, F., Milaneschi, Y. & Penninx, B. W. J. H. Sit, step, sweat: Longitudinal associations between physical activity patterns, anxiety and depression. *Psychol. Med.* **47**, 1466–1477 (2017).

49. Kendler, K. S. The dappled nature of causes of psychiatric illness: Replacing the organic-functional/hardware-software dichotomy with empirically based pluralism. *Mol. Psychiatry* **17**, 377–388 (2012).

50. Kendler, K. S. The structure of psychiatric science. *Am. J. Psychiatry* **171**, 931–938 (2014).

51. Babyak, M. A. Understanding confounding and mediation. *Evid. Based. Ment. Health* **12**, 68–71 (2009).

52. Burnham, K. P. & Anderson, D. R. Multimodel Inference: understanding AIC and BIC in Model Selection, Amsterdam Workshop on Model Selection. *Sociol. Methods Res.* **33**, 261–304 (2004).

53. Tonidandel, S. & LeBreton, J. M. Relative Importance Analysis: A Useful Supplement to Regression Analysis. *J. Bus. Psychol.* **26**, 1–9 (2011).

54. Eaton, W. W. *et al.* Population-based study of first onset and chronicity in major depressive disorder. *Arch. Gen. Psychiatry* **122**, 184–191 (2008).

55. De Graaf, R., Bijl, R. V., Ravelli, A., Smit, F. & Vollebergh, W. A. M. Predictors of first incidence of DSM-III-R psychiatric disorders in the general population: Findings from the Netherlands Mental Health Survey and Incidence Study. *Acta Psychiatr. Scand.* **106**, 303–313 (2002).

56. ten Have, M. *et al.* Duration of major and minor depressive episodes and associated risk indicators in a psychiatric epidemiological cohort study of the general population. *Acta Psychiatr. Scand.* **136**, 300–312 (2017).

57. Klijs, B. *et al.* Representativeness of the LifeLines Cohort Study. *PLoS One* **10**, e0137203 (2015).

58. Hölzel, L., Härter, M., Reese, C. & Kriston, L. Risk factors for chronic depression - A systematic review. *J. Affect. Disord.* **129**, 1–13 (2011).

59. Scholtens, S. *et al.* Cohort Profile: LifeLines, a three-generation cohort study and biobank. *Int. J. Epidemiol.* **44**, 1172–1180 (2015).

60. Bot, M. *et al.* Validity of LIDAS (LIfetime Depression Assessment Self-report): A self-report online assessment of lifetime major depressive disorder. *Psychol. Med.* **47**, 279–289 (2017).

61. Sheehan, D. V *et al.* The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J. Clin. Psychiatry* **59**, 22–57 (1998).

62. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR*. (American Psychiatric Publishing, 2000).

63. Rothman, K. J., Greenland, S. & Associate, T. L. L. Modern Epidemiology, 3rd Edition. *Hastings Cent. Rep.* (2014)

64. Newman, S. C. & Bland, R. C. Incidence of mental disorders in Edmonton: Estimates of rates and methodological issues. *J. Psychiatr. Res.* **32**, 273–282 (1998).

65. Grant, B. F. *et al.* Sociodemographic and psychopathologic predictors of first incidence of DSM-IV substance use, mood and anxiety disorders: Results from the Wave 2 National Epidemiologic Survey on Alcohol and Related Conditions. *Mol. Psychiatry* **14**, 1051–66 (2009).

66. R Development Core Team. R: A Language and Environment for Statistical Computing. (2013)

67. van Buuren, S. & Groothuis-Oudshoorn, K. mice: Multivariate imputation by chained equations in R. *J. Stat. Softw.* **45**, 1–67 (2011).

68. Bartón, K. MuMIn: Multi-model inference. (2020).

69. Costa, P. T. & McCrae, R. R. The revised NEO personality inventory (NEO-PI-R). in *The SAGE Handbook of Personality Theory and Assessment: Volume 2 - Personality Measurement and Testing* (2008).

70. Mattisson, C. *et al.* Risk factors for depressive disorders in the Lundby cohort - A 50 year prospective clinical follow-up. *J. Affect. Disord.* **113**, 203–215 (2009).

71. Murphy, J. M., Laird, N. M., Monson, R. R., Sobol, A. M. & Leighton, A. H. Incidence of depression in the Stirling County Study: Historical and comparative perspectives. *Psychol. Med.* **30**, 505–514 (2000).

72. Lukat, J., Becker, E. S., Lavallee, K. L., van der Veld, W. M. & Margraf, J. Predictors of Incidence, Remission and Relapse of Axis I Mental Disorders in Young Women: A Transdiagnostic Approach. *Clin. Psychol. Psychother.* **24**, 322–331 (2017).

73. Klein, D. N. *et al.* Predictors of first lifetime onset of major depressive disorder in young adulthood. *J. Abnorm. Psychol.* **122**, 1–6 (2013).

74. Bromberger, J. T. *et al.* Predictors of first lifetime episodes of major depression in midlife women. *Psychol. Med.* **39**, 55–64 (2009).

75. Beard, J. R., Heathcote, K., Brooks, R., Earnest, A. & Kelly, B. Predictors of mental disorders and their outcome in a community based cohort. *Soc. Psychiatry Psychiatr. Epidemiol.* **42**, 623–630 (2007).

76. Havinga, P. J. *et al.* Doomed for disorder? High incidence of mood and anxiety disorders in offspring of depressed and anxious patients: A prospective cohort study. *J. Clin. Psychiatry* **78**, e8–e17 (2017).

77. Leijdesdorff, S., Van Doesum, K., Popma, A., Klaassen, R. & Van Amelsvoort, T. Prevalence of psychopathology in children of parents with mental illness and/or addiction: An up to date narrative review. *Curr. Opin. Psychiatry* **30**, 312–317 (2017).

78. Van Sprang, E. D. *et al.* Familial risk for depressive and anxiety disorders: Associations with genetic, clinical, and psychosocial vulnerabilities. *Psychol. Med.* 1–11 (2020)

79. Agerbo, E. *et al.* Risk of Early-Onset Depression Associated With Polygenic Liability, Parental Psychiatric History, and Socioeconomic Status. *JAMA psychiatry* **75**, 387–397 (2021).

80. National Research Council and Institute of Medicine. *Depression in parents, parenting, and children: Opportunities to improve identification, treatment, and prevention*. (The National Acedemies Press, 2009).

81. Reedtz, C. *et al.* Promotion of Wellbeing for Children of Parents With Mental Illness: A Model Protocol for Research and Intervention. *Front. Psychiatry* **10**, 1–10 (2019).

82. Maciejewski, D., Hillegers, M. & Penninx, B. Offspring of parents with mood disorders: Time for more transgenerational research, screening and preventive intervention for this high-risk population. *Curr. Opin. Psychiatry* **31**, 349–357 (2018).

83. Havinga, P. J. *et al.* Prevention programmes for children of parents with a mood/anxiety disorder: Systematic review of existing programmes and meta-analysis of their efficacy. *Br. J. Clin. Psychol.* **60**, 212–251 (2021).

84. De Graaf, R., Ten Have, M., Tuithof, M. & Van Dorsselaer, S. First-incidence of DSM-IV mood, anxiety and substance use disorders and its determinants: Results from the Netherlands Mental Health Survey and Incidence Study-2. *J. Affect. Disord.* **149**, 100–107 (2013).

85. Wittchen, H. U. *et al.* The size and burden of mental disorders and other disorders of the brain in Europe 2010. *Eur. Neuropsychopharmacol.* **21**, 655–679 (2011).

86. Baxter, A. J. *et al.* Challenging the myth of an 'epidemic' of common mental disorders: trends in the global prevalence of anxiety and depression between 1990 and 2010. *Depress. Anxiety* **31**, 506–516 (2014).

87. Kuehner, C. Why is depression more commen among women than among men? *The Lancet Psychiatry* **4**, 146–58 (2017).

88. Kessler, R. C. Epidemiology of women and depression. *J. Affect. Disord.* **74**, 5–13 (2003).

89. Altemus, M., Sarvaiya, N. & Neill Epperson, C. Sex differences in anxiety and depression clinical perspectives. *Front. Neuroendocrinol.* **35**, 320–330 (2014).

90. Spijker, J. *et al.* Determinants of persistence of major depressive episodes in the general population. Results from the Netherlands Mental Health Survey and Incidence Study (NEMESIS). *J. Affect. Disord.* **81**, 231–240 (2004).

91. Hardeveld, F. *et al.* Recurrence of major depressive disorder across different treatment settings: Results from the NESDA study. *J. Affect. Disord.* **147**, 225–231 (2013).

92. ten Have, M. *et al.* Recurrence and chronicity of major depressive disorder and their risk indicators in a population cohort. *Acta Psychiatr. Scand.* **137**, 503–515 (2018).

93. Mattisson, C., Bogren, M., Horstmann, V., Munk-Jörgensen, P. & Nettelbladt, P. The long-term course of depressive disorders in the Lundby Study. *Psychol. Med.* **37**, 883–891 (2007).

94. Nöbbelin, L., Bogren, M., Mattisson, C. & Brådvik, L. Risk factors for recurrence in depression in the Lundby population, 1947–1997. *J. Affect. Disord.* **228**, 125–131 (2018).

95. Murphy, J. M., Olivier, D. C., Sobol, A. M., Monson, R. R. & Leighton, A. H. Diagnosis and outcome: depression and anxiety in a general population. *Psychol. Med.* **16**, 117–126 (1986).

96. Hoertel, N. *et al.* A comprehensive model of predictors of persistence and recurrence in adults with major depression: Results from a national 3-year prospective study. *J. Psychiatr. Res.* **95**, 19–27 (2017).

97. Lorant, V. *et al.* Socioeconomic inequalities in depression: A meta-analysis. *Am. J. Epidemiol.* **157**, 98–112 (2003).

98. Patel, V. *et al.* Income inequality and depression: a systematic review and meta-analysis of the association and a scoping review of mechanisms. *World Psychiatry* **17**, 76–89 (2018).

99. Buckman, J. E. J. *et al.* Risk factors for relapse and recurrence of depression in adults and how they operate: A four-phase systematic review and meta-synthesis. *Clin. Psychol. Rev.* **64**, 13–38 (2018).

100. Harrell, F. E., Lee, K. L., Matchar, D. B. & Reichert, T. A. Regression models for prognostic prediction: Advantages, problems, and suggested solutions. *Cancer Treat. Rep.* **69**, 1071–1077 (1985).

101. Ferrari, F. & Villa, R. F. The Neurobiology of Depression: an Integrated Overview from Biological Theories to Clinical Evidence. *Mol. Neurobiol.* **54**, 4847–4865 (2017).

102. Nestler, E. J. *et al.* Neurobiology of depression. *Neuron* **34**, 13–25 (2002).

103. Frank, E. *et al.* Conceptualization and rationale for consensus definitions of terms in major depressive disorder: Remission, recovery, relapse, and recurrence. *Arch. Gen. Psychiatry* **48**, 851–855 (1991).

104. Rush, A. J. *et al.* Report by the ACNP Task Force on response and remission in major depressive disorder. *Neuropsychopharmacology* **31**, 1841–1853 (2006).

105. Marcellus, L. Are we missing anything? Pursuing research on attrition. *Can. J. Nurs. Res.* **36**, 82–98 (2004).

106. Lamers, F. *et al.* Sociodemographic and psychiatric determinants of attrition in the Netherlands Study of Depression and Anxiety (NESDA). *Compr. Psychiatry* **53**, 63–70 (2012).

107. Eaton, W. W., Anthony, J. C., Tepper, S. & Dryman, A. Psychopathology and attirition in the epidemiologic catchment area surveys. *Am. J. Epidemiol.* **135**, 1051–1059 (1992).

108. De Graaf, R., Bijl, R. V., Smit, F., Ravelli, A. & Vollebergh, W. A. M. Psychiatric and sociodemographic predictors of attrition in a longitudinal study: The Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Am. J. Epidemiol.* **152**, 1039–1047 (2000).

109. De Graaf, R., Van Dorsselaer, S., Tuithof, M. & Ten Have, M. Sociodemographic and psychiatric predictors of attrition in a prospective psychiatric epidemiological study among the general population. Result of the Netherlands Mental Health Survey and Incidence Study-2. *Compr. Psychiatry* **54**, 1131–1139 (2013).

110. Chatfield, M. D., Brayne, C. E. & Matthews, F. E. A systematic literature review of attrition between waves in longitudinal studies in the elderly shows a consistent pattern of dropout between differing studies. *J. Clin. Epidemiol.* **58**, 13–19 (2005).

111. Eerola, M., Huurre, T. & Aro, H. The problem of attrition in a Finnish longitudinal survey on depression. *Eur. Jouranl Epidemiol.* **20**, 113–120 (2005).