High Levels of Neuroticism as a Predictor for the First Onset of Depression in Youth: A Systematic Review

Remo Blauw

s4750861

Department of Psychology, University of Groningen

PSB3E-BT15: Bachelor Thesis

Group number: 2324_1a_22

Supervisor: Dr. M. Poppelaars

Second evaluator: Dr. C. Vrijen

In collaboration with: R. Dijkstra, S. Fietjes, L. W. Hopman, V. van den Bosch, A.L. Voogd

February 18th, 2024

A thesis is an aptitude test for students. The approval of the thesis is proof that the student has sufficient research and reporting skills to graduate, but does not guarantee the quality of the research and the results of the research as such, and the thesis is therefore not necessarily suitable to be used as an academic source to refer to. If you would like to know more about the research discussed in this thesis and any publications based on it, to which you could refer, please contact the supervisor mentioned

Abstract

Aim: Identifying neuroticism and negative affectivity as vulnerability factor for the first onset depressive disorders. We collectively conducted a systematic review for the predictors of depressive- and anxiety disorder(s) onset in childhood or adolescence up to 17 years of age. **Method**: We systematically searched the Web of Science database for eligible reports between January 1st, 2000 to November 7th, 2023. Subjects were limited to first onset of depression- or anxiety diagnosis without comorbidities. Studies were eligible if they were longitudinal, non-experimental, and included mentions of vulnerability or risk factors for the development of first onset depression. **Results**: Six reports were included in the systematic review. Samples were predominantly caucasian (~80%), female (65%), and only from high income countries. Two reports found a nonsignificant effect, three reports found high neuroticism scores to be a vulnerability factor for first onset depression. One report found an significant effect through interaction of low fear and stressors. **Discussion**: Results indicate mixed support for neuroticism as vulnerability for first onset of depression. The majority of results support a overall small effect size of vulnerability through neuroticism for youth aged between 11 and 17. Insufficient reports were found to make any nuanced claims for ages below 10. Future recommendations are made to further the understanding of this subject.

Keywords: First onset, Mood disorders, Temperament, Personality, Systematic review.

High Levels of Neuroticism as a Vulnerability for the First Onset of Depression in Youth: A Systematic Review

The negative effects of depressive disorders (DD) on individuals worldwide steers the scientific community to finding improved methods of prevention (e.g. El-Monshed et al., 2021), diagnosis (e.g. Wu et al., 2023), and intervention (e.g. LoPilato et al., 2023; Dawson et al., 2022). These steps are of importance as depression is still one of the major psychopathologies worldwide and places profound cognitive, social, emotional, and financial burdens (König et al., 2020) on individuals and societies alike. Recent decades have seen a rise in research on youth and adolescent depressive disorders (e.g. Bernaras et al., 2019; Lewis et al.,2020) as differences between onset age seem to play a role in depression outcomes.

The onset of depression can occur in different developmental stages in life (Haddad & Boyce, 2023). The developmental stage in which a depressive disorder takes place can affect things such as symptoms. Notably, childhood- or adolescent onset depression (hereby called *early onset depression*; ≤17 y.o.) differs in a few aspects when compared to adult- or late onset depression. Some of these differences include a more chronic/recurrent nature (Gollan et al., 2005; Hammen et al., 2008), higher levels of suicide (Zisook et al., 2007), increased prevalence of weight change, insomnia, loss of energy (Rice et al., 2019) and higher chances of developing into a mania disorder (Todd & Botteron, 2002). These aspects highlight different symptoms that might be present in an early onset depression when compared to adult depression. The differences in symptoms might indicate that different intervention strategies are required to effectively combat depression in youth. Additionally early onset depression has been shown to be less responsive to pharmaceutical drug interventions compared to depression onset at later stages of life (Cipriani et al., 2016). This makes DD in youth more

difficult to treat effectively with medication. The combination of symptoms and biological aspects show that the developmental stage in which the onset takes place might affect the trajectory of the DD. Affecting DD symptoms, duration, and the potential for intervention. Therefore focussing on early onset depression can prove beneficial for prevention endeavours.

One potential reason why early onset depression can be more detrimental on the long run is due to the build up of symptoms over time. The *scarring hypothesis* (Lewinsohn et al., 1981) is a concept that embodies the buildup of symptoms over time after a disorder onset. The hypothesis states that a disorder can create permanent psychological damage, or 'scar'. Some aspects that might be altered include cognitions or personality traits (Kendler et al., 1993; Klein et al., 2011). These negative alterations can last even after episodes subside, leaving a permanent or long lasting vulnerability (scar) to following episodes.

Although promising, the scaring hypothesis is just that: a hypothesis. It is not fully supported in the literature as there is a lack of consensus or conclusive findings (Beever et al., 2007; Jylhä et al., 2009; Prince et al., 2021; Rohde et al., 1990; Steiger et al., 2015). Supportive evidence seems to be boasting small increases in neuroticism (Prince et al., 2021), while non-supportive evidence seems to find no effects when controlling for other relevant variables (Rohde et al., 1990), even despite boasting adequate statistical power (Beevers et al., 2007). Sadly, as of the writing of this paper, there is no meta analysis related to the scarring hypothesis. Therefore there is no nuanced overarching view on the current state of the scarring hypothesis in literature. Despite the inconclusive nature of the scaring hypothesis it does provide a reason for measuring the first onset of depression. This reason is that it ensures that no scarring effects can be measured, as the first onset provides an unaltered (or unscarred) measure since no scarring has taken place. Thereby completely circumventing the necessity of a meta-analysis on the scaring hypothesis.

First onset of depression has become a focus of research in more recent years. Aligned with the scarring effect there are multiple reasons why the first onset is an important focus for research. As previous episodes seem to be a major predictor for recurring episodes (Patton et al., 2003), preventing the first episode from taking place is the most natural goal for prevention programmes. In addition, acquired symptoms tend to interact with a person their context and develop into more negative experiences, symptoms and alike. The symptoms build on each other as they are not static constructs, but function more like adaptive systems (see: Lewis, 1997). This active nature gives symptoms the potential to develop additional maladaptive behaviours, symptoms or entrenching existing ones. For example, a depressed person with anhedonia becomes less active, thereby reducing positive reinforcement and becoming more socially secluded. The distinction between the first onset and recurrent episodes is further shown in the works of Pettit et al.. (2006). They found that symptoms present before a depressive episode differ between first- and recurrent episodes. The predictive value of previous episodes, potential scarring effect, and dynamic aspects of symptoms are some examples of influences that previous episodes of depression might have on the long run. Therefore, the first episode has the possibility of cascading into multiple different vulnerabilities and can be a prime target for prevention.

Prevention tactics for DD use risk factors and vulnerabilities to predict and avoid depression onset. For clarity a distinction is made between risk and vulnerability (Ramskogler & Bruckner, 2014), because this allows an improved methodical approach to predict and prevent DD onset, thereby improving efficiency. Risk factors are outside forces that exert the same level of stress on any individual. Vulnerabilities on the other hand are individual differences that decrease the resistance against risk factors, thus increasing their chances of having an effect on an individual. For example, childhood abuse can affect anyone’s risk to

develop a distress disorder, but reduced effectiveness of serotonin transporters in the brain due to genetic dispositions increases the risk for certain individuals to develop a distress disorder (for other examples see also: Vikander & Strand, 2023). This distinction is made to differentiate between the contextual risk factors and stable (inherent) vulnerability factors.

Personality traits or temperament have long been vulnerability factors for DD onset (e.g. Katz & McGuffin, 1987). Several personality traits have been linked to the onset of depression in meta analyses (Hakulinen et al., 2015; Uliaszek et al., 2010). These include conscientiousness, extraversion and neuroticism or their respective interactions (Vasey et al., 2014). Neuroticism, or similar temperamental constructs such as negative emotionality or negative affectivity (Grist & McCord, 2010; Ramos-Grille et al., 2022; Watson & Clark, 1992), are related to fear, sadness, or negative feelings (Lahey, 2009). These aspects are usually central to various mental disorders (McCrae & Costa, 2013) as disorders tend to invoke negative feelings which align with neuroticism. The central aspect of neuroticism for disorders can also be related to the general psychopathology factor (*p-factor*). The p-factor is an overarching common aspect between different psychopathologies, as it tends to correlate highly with neuroticism scores (Caspi et al., 2014).

Neuroticism is highly associated as a vulnerability factor for the onset of DD (Zinbarg et al., 2010; Chen et al., 2021; Lehto et al., 2018, Tachi et al., 2019) and subthreshold depression (Toenders et al., 2022). For example: chronic high neuroticism increases the chance of a first onset Major Depressive Disorder (MDD; Prince et al., 2021; Vinkers et al., 2014) in adults. Within depressive disorders neuroticism influences symptom severity (Tonarely et al., 2020; van Eeden, 2019) and is related to specific depressive symptoms (Fried et al., 2014). Besides the relation between neuroticism and DD, high neuroticism can be a heritable trait (Atherton & Schofield, 2021; Mackin, 2022). This makes neuroticism a

predictable factor for prevention selection procedures, as parental neuroticism scores can reasonably act as a predictor for high scores in offspring. However there is a catch to using neuroticism as a predictive factor, as neuroticism measures in studies tend to be a mix of both first onset and recurrent onset depressions. Therefore, there is a potential that the scarring effect has altered the neuroticism scores of some of the participants in these studies. Since scarring has the potential of increasing neuroticism scores, this included the potential of skewed results and invalid results relating to first onset depression.

The predictive value of neuroticism for DD, along with the possibility of scarring effects and cascading effects after the first episode, make it a potential beneficial subject for prevention. This systematic review aims to map the current available literature related to the predictive value of high neuroticism for first onset depression in youth. Current literature would suggest that high neuroticism predicts the onset of depression in youth. However, there is a possibility of scarring effects altering data. Thereby potentially skewing the predictive value of high neuroticism. Focussing on youth ensures the highest possibility of measuring the first onset, as adolescence is the most common developmental stage for depression onset. The hypothesis for this review states that high neuroticism acts as a vulnerability for the first onset of depression in youth. This study uses the stress generation model (Hammen, 1991), which states that neuroticism is seen as a vulnerability factor that proceeds the onset of a first depression onset through naturally occurring stressors. Suggestive results can indicate that high neuroticism is a vulnerability for first onset depression, which in turn could help identify vulnerable individuals.

Methods

Search strategy

The review is setup according to the PRISMA guidelines (Page et al., 2021) as is recommended for systematic reviews. This research is part of a collective effort by multiple researchers, all aimed at the first onset of mood disorders. A systematic search of peer reviewed reports was performed in *Web of Science.*

Web of Science was searched on November 7th, 2023. The search terms were set up to select reports mentioning (1) depression or anxiety in the title or keypoints-plus, (2) risk- and vulnerability factors, (3) study design, and (4) sample ages in the title or abstract. The final part of the search string was used to (5) exclude general unrelated subjects or comorbidities in titles to limit search results. Search terms were approved if all predetermined fully eligible reports (n = 12; Bayer et al., 2021; Cohen et al., 2019; Hanson et al., 2015; Hudson et al., 2019; Kendall et al., 2015; Kopala-Sibley et al., 2017; Michelini et al., 2021; Muris et al., 2011; Pan et al., 2017; Platt et al., 2023; Sund & Wichstrøm, 2002; Tanaka et al., 2023) were present in the search results. These predetermined reports were related to all the study subjects. In addition, university experts were consulted for fine tuning of the search terms. The search string that was used in the advanced search can be found in Appendix A.

Date of publishing was set to start from January 1st, 2000 and ends at November 7th, 2023. The publishing limitation was applied to reduce the amount of reports due to time constraints. Simultaneously, the publishing range focusses on the first onset of disorder, as this subject is scarcely found before the year 2000. Only English peer reviewed reports are taken into consideration for this review, due to it being the most commonly used language in the scientific community. Additionally, peer reviewed reports are the bulk of the scientific literature available.

Eligibility criteria

Eligibility criteria were specified according to the Population, Exposure, Comparison, Outcome (PIECO) criteria and in line with PRISMA guidelines (see Table 1). Longitudinal-, cohort-, twin-, prospective-, observational-, and follow-up study designs published between January 1st, 2000 and November 7th, 2023, were included.

Criteria for inclusion were: (a) the main outcome was a diagnosis of the first onset of a depressive disorder to avoid scarring effects; (b) comorbidity of depression and anxiety disorders was allowed due to the prevalent comorbidity of these disorders (Saha et al., 2021). Other comorbidities were excluded because they could influence measurements and skew results. (c) The initial measurement took place before the age of 17. This was done to maximise the likelihood of finding first onset depression cases (Hasin et al., 2005). If participants were measured before age ten then a depression measure would be preferable but not required. These studies would be approved on the assumption that childhood first onset depression is not as prevalent as in later stages of development (Institute of Health Metrics and Evaluation, 2020). (d) A minimum of 20 participants per group was used to ensure ample statistical power was available and thereby reduce Type I & II error likelihood. (e) A minimum of 1 month between measures was used to ensure that diagnoses were possible as symptoms require a minimal duration of two weeks (American Psychiatric Association, 2013a, 2013b). Full inclusion and exclusion criteria can be found in Table 1. If the same sample was used for multiple reports, this would be a reason for exclusion unless unique variables or participants were used when compared to the initial report that used the sample. A case-by-case decision was made for which report to include based on review subject relatedness (see: Appendix B).

Study selection

All database results were collected and uploaded to open service platform *Rayyan: Legacy* (software; Ouzzani et al., 2016). After merging all the results, any duplicates were removed. To ensure similarity in eligibility judgment from all investigators a blind inter-rater reliability (IRR) was trained This was done through collective paper reviewing until a blind IRR ≥90% (see also: McHugh, 2012) was reached. After collectively reviewing 80 reports an IRR of 94% was achieved.

Due to time constraints, the initial report selection round was performed by the integrated Rayyan artificial intelligence (AI) software (Flowchart 1, 'Records marked as ineligible by automation tools'). This AI makes use of a Support Vector Machine Classificator, which is trained with the title and abstract characteristics from reviewed reports. The AI was trained through reviewed reports (n = 877) that came from the database search results. These reports included the predetermined fully eligible reports (n = 12) that were used for the search term fine tuning and the initial IRR training reports (n = 80). The AI training reports were approved by the investigators according to the eligibility of the title and abstract as stated in Table 1.

Based on the reviewed reports the AI was able to rate non-reviewed reports. The Rayyan AI uses a star-ranking to mark relevant reports. A maximum of five- and minimum of a half star can be given to a report. These ratings are given in half star increments. Due to time constraints the 6000 highest ranked reports (≥2.5 stars) were screened by title and abstract for eligibility. The ratings of the reviewed reports ranged from 2.5 to 4.5 stars.

During the second selection phase the highest rated reports (n = 6,000) were evenly divided and reviewed individually by six investigators (n = 1,000). The setup used is made available for improved replicability (see: Appendix C). There was a report overlap of 10% (n

= 100) between each of the investigators for an IRR check. Post-hoc testing found a IRR of 92%. The second selection was made by the investigators based on title and abstract in accordance with the selection criteria (see: Table 1). In cases where eligibility was uncertain (e.g. due to insufficient information or unspecific language) the report was approved for full text review.

Table 1

Inclusion- and Exclusion Criteria for Reviewed Reports.

	Inclusion	Exclusion
Population	I. Sample aged ≤17 years old at first measurement II.Sample aged ≤10 years old without measure of current depression or anxiety disorder(s) III. Comorbidity of Depression- and Anxiety disorders is allowed IV. Sample size of ≥20 participants	I. Sample consists of individuals with specific other psychological disorder(s) (e.g. personality disorder) II. Sample consists of individuals with other medical disorders (e.g. heart disorders, Parkinson's) III. Sample is selected solely on specific subgroups (e.g. African American, Menarche) IV. Sample is refilled after attrition with participants other than those present at the first measurement
Predictors	Vulnerabilities or risk factors: I. Neurology, neurobiological structures. II. Personality, Temperament III. Behavioural inhibition IV. Cognition V. Social environment VI. Parental psychopathology	I. Experimental stressful exposure(s). II. Other non-ecological stressful exposure(s).
Research design	I.Longitudinal study II.Cohort study III.Twin study IV.Prospective V.Observational VI.Follow-up VII.Study duration >1 month.	I.Prevention trials II.Intervention trails III.Single measure research design(s). IV.Related to specific events (e.g. earthquake, tsunami, COVID).
Outcome	I.Mention of a diagnosed Depressive- or Anxiety disorder.	

The third selection phase consisted of report selection based on subject relevancy. As due to time constraints it was not feasible to perform full-text reviews of the reports that were approved in the second phase. The fourth selection phase used the same report distribution method as phase two (see: Appendix D). In this phase the subject relevant reports were reviewed in the full text according to the selection criteria (see: Table 1). No training was performed beforehand, as this phase did not rely on interpretation but on careful reading. A post-hoc IRR of 96.3% was found. If report eligibility was unclear, a group discussion and majority vote was used to decide on eligibility.

The final selection phase was performed individually by each of the investigators. Within this review the selection of reports were aimed at the personality trait domain(s) neuroticism and temperament aspects negative affect/-emotionality. These traits were measured at baseline, before the first onset of a depression. In addition to previous reviewing steps, the remaining undecided reports (n = 3,790) of the first selection phase were reviewed for a mention of neuroticism, personality, or temperament using the Rayyan software. The resulting reports (n = 342) were reviewed by a single investigator according to the afore mentioned selection criteria. This did not provide additional eligible reports for reviewing.

Bias prevention

Despite there being no minimum of reports for systematic reviews (Charrois, 2015), a five reports minimum was set to ensure a mixed perspective of the subject at hand through multiple separate samples. Thereby allowing comparison through multiple results and a nuanced conclusion. Additionally, the financial sources of the selected reports were checked for potential biases or other notable influences of the results. These biases were checked by the investigator that reviewed the report for eligibility during the full text review. This check did not yield any notable results.

Results remain as stated in the original report. Eligible reports their conclusions would be weighed based on the significance of findings, statistical power, effect sizes and their confidence intervals, number of repeated measures, time between measures, and to conclude estimated bias. Through the combination of these measurements, the conclusions were weighed and integrated in the final conclusion.

Results

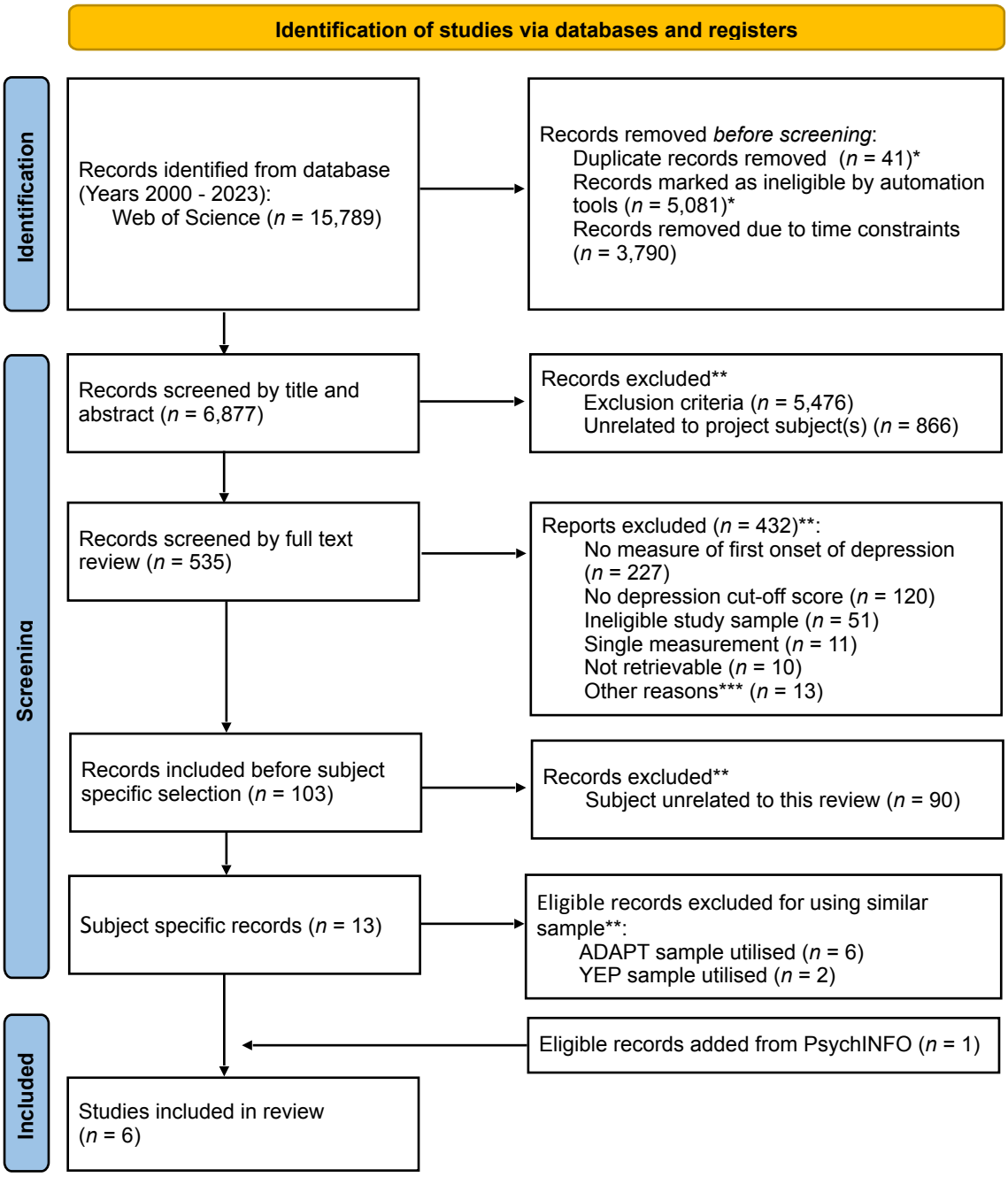
The search terms in Web of science resulted in 15,789 reports. A total of 6,877 reports were screened based on title and abstract (see Figure 1), of which a smaller proportion (n = 877) was used during AI- and IRR training. The 1,401 approved reports were screened on the relatedness to the varying subjects of this project. This resulted in a total of 535 reports eligible for full text screening. A total of six fully eligible reports were included in this systematic review. Reasons for exclusion in the full text screening varied, but most related to (a) no baseline measure of previous onset of depression, (b) the sample being too old for an assumption of no first onset, (c) the sample being too old, or (d) grouping ages above and below the set age of 17. All selection results can be found in the PRISMA diagram (Figure 1).

Some reports made use of the same sample as included reports. A total of eight reports were therefore excluded from the final review. This includes six studies related to the Adolescent Development of Emotions and Personality Traits (ADEPT) study (Kopala-Sibley et al., 2017; Michelini et al., 2021; Mu et al., 2023a; Mu et al., 2023b; Perlman et al., 2018; Silver et al., 2021) and two related to the Youth Emotion Project (YEP; Metts et al., 2021; Zinbarg et al., 2010). Reasoning for exclusion can be found in Appendix B.

Of the six reports included, the majority of studies were performed in the USA (Bress et al., 2013; Bufferd et al., 2014; Goldstein et al., 2018; Kendall et al., 2015; Wilson et al.,

Figure 1

PRISMA Systematic Review Flowchart for Record Selection



Notes. PRISMA flow diagram for eligible report selection. Showing the stepwise selection process for the systematic review. This includes (1) the identification of reports through databases and duplicates removed, (2) the screening process through which title, abstract and full texts were screened, and (3) reasons for excluding—and amounts of—ineligible reports. To conclude (4) the included (and fully eligible) reports within this review. *Total amount of reports excluded by Rayyan software (*n* = 5,122) rated <2.5 stars or duplicates. **Total amount of reports excluded by investigators (*n* = 6,872). ***Reports did not adhere to other eligibility criteria, but are too few to mention separately.

2014) and one in Australia (Davey et al., 2015; see Table 2). Notably, no reports from low- or middle income countries (e.g. China) were eligible for inclusion. Time between first depression measure and follow-up across the reports ranged from nine to thirty-six months, with a median of twelve months (see Table 3). Study participants (n = 3,142) in reviewed reports were predominantly female (65%; n = 2,133) and caucasian (83.8% - 84.8%; see Table 2). One study (Davey et al., 2015) did not mention the percentages of caucasian participants. Therefore the sample range of 0% to 100% caucasian is taken into account. Baseline sample sizes ranged from 68 to 1,512 participants. Mean age at baseline ranged from three to seventeen years old. Measuring a depression diagnosis—as conform to either the DSM-III, -IV, -IV-TR, or -5—was the main outcome for these studies. The measured vulnerability factor for first onset depression was neuroticism, negative emotionality and negative affectivity. All measurement tools utilised in the studies can be found in Table 3. Study result were presented as either odds ratio (OR; n = 3), hazard ratio (HR; n = 1), or standard T-test (n = 2; see Table 4).

Bress and colleagues' (2013) focussed on neurobiological mapping of depressive vulnerabilities through functional magnetic resonance imaging. They oversampled for adolescents with a parental history of depression and had substantial (19%) attrition at follow-up. Initial neuroticism measures did not show differences at follow-up between the Major Depressive Episode onset group (n = 16) compared to no onset group (n = 52; see Tables 3 and 4). Therefore, the initial neuroticism score did not act as a vulnerability for the first onset of a depression in this sample.

In another neurobiological study similar results were found. Davey and colleagues' (2015) neurobiological study selected their sample from a larger pool of

Table 2

Reference and Demographics of Studies (n = 6) Reviewed.

Reference	Sample	Female (%)	Ethinicity (Caucasian %)	Age *M*(sd) -or- range	Country (State)
Bress et al. (2013)	68	100.0	92.6	15 - 17	USA (NY)
Bufferd et al. (2014)	456	46.3	86.6***	3.60 (.30)	USA (NY)
Davey et al. (2015)	56	44.6	-	16.50 (.50)	Australia (Vic)
Goldstein et al. (2018)	479	100.0	80.5 ***	14.40 (.63)	USA (NY)
Kendall et al. (2015)	627	69.2	49.0***	16.90 (.43)	USA (NY, CA)
Wilson et al. (2014)	1512*	54.0**	98.0**	11.72 (.43)	USA (MN)

Notes. *The study sample consists of a young cohort (from age 11) and old cohort (from age 17). Only the young cohort is used. ** Percentage based on young and old cohort combined. ***Specified as non-Hispanic caucasian. Percentages are rounded up to one decimal.

participants. Negative affectivity and effortful control scores in the selected participants were widely sampled above and below the mean score. Temperament and depression were measured at both time points (see Table 3 and 4). At follow-up eight participants received a MDD diagnosis and forty-eight did not receive a MDD diagnosis. A T-test, comparing the MDD onset group to the no onset group, found no difference in average negative affectivity scores at initial measurement. Thus the negative affectivity score was not a vulnerability for the initial onset of a depression in their sample.

Along the same vein, Bufferd et al. (2014) used the negative emotionality sub-scores dysphoria and fear/inhibition to predict initial onset (see: Table 3). The sample consisted of preschool children starting at age three, with follow-up at age six (see: Table 2). A comparison

Table 3

Reference and Measurement Information of Studies (n = 6) Reviewed

Reference	*Trait measured*	*Traits measurement tool(s)[1]*	Diagnostic standard utilised	*Type of depression measured[2]*	*Depression measurement tool(s)[3]*	*Mean months between measures*	*Total amount of follow-ups*
Bress et al. (2013)	Neuroticism	BFI	DSM-IV	MDD	IDAS; PHQ-9	21	1
Bufferd et al. (2014)	Negative Emotionality (Dysphoria, Fear/inhibition)	LABTAB	DSM-IV-TR	MDD*, Dysthymia, DNOS	CAPA; PAPA	30	1
Davey et al. (2015)	Negative Affectivity	EATQ-R	DSM-5	MDD	K-SADS-PL, CES-D	24	1
Goldstein et al. (2018)	Neuroticism	BFI	DSM-IV	DD	K-SADS-PL; IDAS; DEQ	9	2
Kendall et al. (2015)	Neuroticism	EPQ-R-N; IPIP-NEO-PI-R	DSM-IV	DD	SCID-I/NP; MASQ-AD	12	10
Wilson et al. (2014)	Negative Emotionality	MPR; TRF; MPQ	DSM-III-R	MDD	DICA-R; SCID	36	6

Notes.[1] Personality measurement tools abbreviations: Big Five Inventory (BFI); Early Adolescent Temperament Questionnaire (EATQ-R); Eysenck Personality Questionnaire—Revised (EPQ-R); International Personality Item Pool NEO Personality Index - Revised (IPIP-NEO-PI-R); Laboratory Temperament Assessment Battery (LABTAB); Multidimensional Personality Ratings (MPR); Multidimensional Personality Questionnaire (MPQ); Teacher Rating Form (TRF). [2] Depression diagnoses abbreviations: Depressive Disorder (DD); Major Depressive Disorder (MDD); Depression Not Otherwise Specified (DNOS).[3] Depression measurement tools abbreviations: Inventory of Depression and Anxiety Symptoms (IDAS); Patient Health Questionnaire - 9 (PHQ-9); Childhood and Adolescent Psychiatric Assessment (CAPA); Center for Epidemiologic Studies Depression (CES-D); Diagnostic Interview for Children and Adolescents—Revised (DICA-R); Kiddie Schedule for Affective Disorders and Schizophrenia for School-Age Children–Epidemiological version–Present and Lifetime (K-SADS-PL); Preschool-Age Psychiatric Assessment (PAPA); Structured Clinical Interview for DSM-III-R (SCID); Structured Clinical interview for DSM IV - non-patient edition (SCID-I/NP); Mood and Anxiety Symptom Questionnaire - Anhedonic Depression scale (MASQ-AD). *Modified criteria for MDD (Juby et al., 2002) were used for diagnosis.

Table 4

Reference, Measurement Information and Results of Studies (n = 6) Reviewed

Reference	*Type of Analysis*	*Result*
Bress et al. (2013)	T-test	Comparing the Major Depressive Episode onset group to the no onset group showed no difference in average neuroticism score ($t(66) = -.28$, $p > .10$) at initial measurement. Therefore neuroticism was not a vulnerability factor within this sample.
Bufferd et al. (2014)	Odds ratio	Comparing negative emotionality sub-scores between non-depressed and depressed participants showed a non-significant results for predictors dysphoria (*OR* 1.26, 95% CI [0.84, 1.89], $p > .05$) and fear/inhibition (*OR* 1.06, 95% CI [0.73, 1.53], $p > .05$). Interactions between low fear/inhibition and early stressors were a vulnerability factor for first onset of a depressive disorder (*OR* 1.42, 95% CI [1.21, 1.68], $p < .001$) in this sample with a small effect size.
Davey et al. (2015)	T-test	Comparing the *MDD onset group* to the *no MDD onset group* showed no difference in average negative affectivity score ($t(54) = -.7$, $p = .49$) at initial measurement. Therefore negative affectivity was not a vulnerability factor for first onset depression in this sample.
Goldstein et al. (2018)	Odds ratio	A comparison of initial neuroticism in participants with first onset depression compared to no diagnosis showed a significant small effect size (*OR* 1.99, 95% CI [1.46, 2.72], $p < .001$). However, after controlling for clinical risk factors a smaller significant effect size was found only for the sub-scale depressivity (*OR* 1.54, 95% CI [1.00, 2.43], $p < .05$). Therefore, sub-scale depressivity was a vulnerability factor for first onset depression.
Kendall et al. (2015)	Hazard ratio	The hazard ratio of neuroticism on the development of first onset of a depressive disorder (*HR* 2.08, 95% CI [1.40, 3.08], $p = .001$) was statistically significant with a small effect size. Removal of the anhedonia indicator from the model (*HR* = 2.09, 95% CI [1.43, 3.07] $p < .001$) had no major effect on the results. Anhedonia was removed because it symptomatically differentiates DD from other emotional disorders. The results indicate that high neuroticism was a vulnerability factor for first onset depression in their sample.
Wilson et al. (2014)	Odds ratio	An increase of one standard deviation in negative emotionality showed significant increased prevalence of developing a first depressive disorder (OR 1.27, 95% CI [1.11, 1.46], $p < .001$) up to the age of 29. Early onset depression was more likely to develop compared to late onset depression when using negative emotionality as predictor (*OR* 1.43, 95% CI [1.14, 1.79], $p < .01$). Finally an early depressive onset with recurrent episodes had a higher likelihood than any other MDD variant of developing when predicted by the increased negative emotionality score (*OR* 2.13, 95% CI [1.23, 3.68], $p < .001$). All results found are categorised as small effect sizes.

Abreviations: Odds Ratio (OR); Confidence Interval (CI); Major Depressive Disorder (MDD); Depressive Disorders (DD).

between non-depressed (n = 423) and depressed participants (n = 33) found nonsignificant results for predictors dysphoria and fear/inhibition. Other sub-scores of negative emotionality were pre-emptively removed during initial explorations due to non-significance. However, interactions between low fear/inhibition and early life stressors were a vulnerability factor for first onset depression in their sample (see Table 4). Their finding were the opposite to what the diathesis-stress model would predict, as this model states that high fear should interact with stressors to predict depression.

Other studies found positive results for the predictive value of neuroticism. Goldstein et al. (2018) studied the ADEPT cohort, consisting solely of adolescent females, to predict depression using personality traits and their subcomponents. Using a logistical regression of initial neuroticism scores they compared depressed (n = 49) vs. non depressed (n = 443) participants at follow-up. Results showed that high neuroticism was a vulnerability factor for first onset depression in their sample. More conservative measures that adjusted for other clinical risk factors (e.g. baseline subthreshold depression) reduced neuroticism its effects on depression onset. Only the neuroticism subcomponent depressivity had a unique effect on DD onset and was therefore a vulnerability factor, whereas neuroticism subcomponents anger and anxiousness did no longer show any significant effects.

Similar to longitudinal design of the ADEPT cohort, Kendall and colleagues (2015) followed the Youth Emotion Project cohort of adolescents annually for ten years. Their main focus was on building a predictive model for first onset depression. The cohort was oversampled for higher scores of neuroticism (59% of the sample had a top tertiary neuroticism scores). Final follow-up diagnoses comprised of 88 participants with first DD onset and 539 without onset. Calculations were performed both with and without anhedonia, as this is a differentiating factor from other emotional disorders. Results were displayed by

use of a proportional hazard analysis (See: Cox, 1972). The study found that neuroticism was a vulnerability factor for the onset of DD at a hazard ratio of 2.08 (see Table 4). The hazard ratio translates to slightly more than twice as many depression onsets within the same timespan for high neuroticism when compared to average neuroticism scores.

The final reviewed study was performed by Wilson and colleagues (2014), which made use of the Minnesota Twin Family Study cohort. The sample consisted of a younger and older segment (see Table 2). Only the younger segment was used for the personality/ temperamental measures as predictor for MDD. Depression onset(s) were specified as (a) MDD onset or no onset, (b) early- or late onset, and (c) single or recurrent episodes up to age 29. Results showed that an increase of one standard deviation in negative emotionality significantly increased the odds of developing a first depressive disorder (see Table 4). Additionally, early onset and recurrent depressive episodes was significantly more prevalent (*OR* 2.13, 95% CI [1.23, 3.68], $p < .001$) than other MDD onset types (e.g. single episode early onset, late onset) when preceded by higher than average negative emotionality levels.

Discussion

In this systematic review it was hypothesised that high neuroticism scores would act as a vulnerability factor for the first onset of depression in youth. A total of six eligible reports were reviewed and found mixed results in support of the hypothesis. However, due to variations in study designs, sample sizes and sample ages this review suggests that the hypothesis is fairly plausible for the adolescent (11-17 y.o.) population.

Three reports found results that are consistent to the hypothesis (Kendall et al., 2015; Wilson et al., 2014; Goldstein et al., 2018) with all of them boasting small effect sizes (Chen et al, 2010). These studies all utilised samples over 450 participants, which gave them ample statistical power. In contrast, two reports found nonsignificant results (Bress et al., 2013;

Davey et al., 2015). Their findings came from comparing depression onset- vs no onset groups. These studies made use of samples with fewer than 70 participants and were underpowered. To conclude, one study found significant results in a sample of children aged 3-6 years old, but to the contrary direction. Low fear was a vulnerability when interacting with stressors contrary to the expected high fear. The combined results of the reports are therefore only suggestive if the report by Bufferd et al. (2014) is excluded. Separation of these results is reasonable, due to their samples being in substantially different developmental stages and therefore contain different biological and cognitive processes (Andersen & Teicher, 2008; Paus et al, 2008). Due to insufficient reports related to ages below ten years old, it is not feasible to make any nuanced statements on neuroticisms effect during those ages.

A potential reason for nonsignificant findings in Bress et al. (2013) and Davey et al. (2015) comes from the strength of the vulnerability factor that is analysed. The hypothesis consistent results (Goldstein et al, 2018; Kendall et al., 2015; Wilson et al., 2014) all showed small effect sizes (Chen et al., 2010). If an overall effect is small it falls within reason that some smaller sampled studies find nonsignificant results. This is because underpowered studies tend to find a wider range of effect sizes that deviate from the overall effect when compared to high powered studies. In general, the findings in this systematic review are in line with findings from meta analyses related to high neuroticism scores and depression. This includes cross-sectional (Lyon et al., 2021) and longitudinal relations between high neuroticism and depression in adults (Hakulinen et al., 2015) and college students (Liu et al., 2019).

The work of Bufferd and colleagues (2014) is an exceptional study due to their result which found an interaction between low fear/inhibition and stressors. This finding is opposite to what would be expected based on the diathesis-stress model. The diathesis-stress model for

instance claims that vulnerabilities interact with high stressors causing the onset of a disorder (Klein et al., 2011). This exceptional result might indicate a different underlying process for the first depression onset when comparing adolescent participants to young children. Findings do fall in line with the *social push hypothesis* (see Olino et al., 2020). In contrary to the diathesis stress model, the social push hypothesis proposes that the effects of positive personality traits (e.g. low fear) might be most beneficial in contexts which lack stressors. When stressors are present the advantages of a positive personality traits are nullified, or could become a vulnerability. (Kushner, 2015). This aligns with Bufferd et al. (2014) their findings, as low fear interacting with stressors was found to be a greater predictor for depression onset than high fear.

Strengths and limitations

Strengths of this review include its focus on the onset of clinical depression diagnosis. By using diagnosis as an outcome it ensures that people are affected to such a degree that intervention is preferable. Therefore results relate to a population that is of importance for intervention. The diagnosis is additionally connected to neuroticism or related scores, which are easily measurable and widely available in many languages (e.g. Laidra et al., 2017; Muris et al., 2005; Olivier & Herve, 2015;). Finally, as no depression has taken place before the first measure, there is no altered neuroticism measure through a scarring effect (Kendler et al., 1993) or interference from previous depressive symptoms. Results thereby show an unaltered personality/temperament score and its predictive value for first depression onset.

There are however limitations to this systematic review. Subthreshold depression is out of focus in most studies despite the similarity and potential for full depression onset (Wesselhoeft et al., 2016). Secondly is the variability of the diagnostic- and personality measurement tools used. This variation could have affected the findings due to differences in

specificity or sensitivity for disorders (Subica et al, 2014) and personality traits. Thirdly, findings are aimed at all potential depression diagnoses and multiple neuroticism measures. Focussing on specific depression disorders (e.g. Valerio et al., 2020) or neuroticism traits might result in alterations in the findings. Fourthly, comorbidity was filtered out to avoid interference from other disorders. However, depression is a highly comorbid with other disorders (Li et al., 2020; Saha et al., 2022; Sandstrom et al., 2021). Findings related to comorbidity are therefore of clinical relevance but not included in this review. Fifthly, eligible reports might have been missed in part to the AI, skipped reports due to time constraints, or the database search terms. Any missed reports might affect the conclusion, especially given the small amount of reports that were eligible. To conclude, this systematic review is the first performed by the writer for academic development purposes. This brings with it pitfalls related to the learning process including—but not limited to—writing oversights, time limitations, interpretation errors, and divided attention between subject and course requirements. Therefore, despite best intentions the reliability of the conclusion might be affected.

Future recommendations

Relatively few of the full text reports reviewed (~19%) were eligible for use in the final review. However, the majority of excluded reports were not eligible due to either (a) no measures of initial onset of depression, or (b) no use of a depression diagnosis or mention thereof. Therefore, the first recommendation is to include measures with a clinical cut-off score for depression diagnosis and participants their history of disorders. Focussing on a specific disorders within the depression category could be a beneficial avenue of research as well. Within this review the category might be too broad to find any strong effects from neuroticism. Including the recommendations increases available data for following reviews.

Additionally, full text reviews did illuminate multiple cohort longitudinal samples (e.g. NESDA, Preschool Depression Study) which do have sufficient information to perform a post-hoc data analysis of first onset. However, these samples are not yet used for this purpose as of the writing of this report, to the knowledge of the writer. Using existing samples could further bolster the body of knowledge for first onset depressions. To conclude, future systematic reviews could run reference searches and contact authors of reviewed reports for additional eligible reports. This would add to the findings from this review. Utilising the future recommendations would further our understanding of neuroticism as a vulnerability factor for first onset depression in youth.

Conclusion

In this review it was hypothesised that high neuroticism would act as a vulnerability factor for first onset depression. The findings in this review support that this is the case for the adolescent population. For adolescents the overall effect size found was small. However, this effect received mixed support from study results. Only underpowered reports that used an adolescent sample found nonsignificant effects. The single study with a sample below six years old found the opposite effect as what was expected. Contrary to the adolescent population, no conclusion could be made for the age group younger than ten years old. There were insufficient reports available to give a nuanced conclusion. As literature related to the subject of this review is scarce, additional supports are needed for more substantial conclusions.

References

References with an asterisk indicate reports included in the review

American Psychiatric Association. (2013a). Depressive disorders. In Diagnostic and statistical manual of mental disorders (5th ed.). https://doi.org/10.1176/appi.books.9780890425596.dsm05

American Psychiatric Association. (2013b). Anxiety disorders. In Diagnostic and statistical manual of mental disorders (5th ed.). https://doi.org/10.1176/appi.books.9780890425596.dsm05

Andersen, S. L., & Teicher, M. H. (2008). Stress, sensitive periods and maturational events in adolescent depression. Trends in Neurosciences, *31*(4), 183-191. https://doi.org/10.1016/j.tins.2008.01.004

Atherton, O. E., & Schofield, T. J. (2021). Personality and parenting. In O. P. John & R. W. Robins (Eds.), Handbook of personality: Theory and research., 4th ed. (pp. 352-366). The Guilford Press. http://search.ebscohost.com.proxy-ub.rug.nl/login.aspx?direct=true&db=psyh&AN=2021-06060-016&site=ehost-live&scope=site

Barrocas, A. L., & Hankin, B. L. (2011). Developmental pathways to depressive symptoms in adolescence: A multi-wave prospective study of negative emotionality, stressors, and anxiety. Journal of Abnormal Child Psychology, *39*(4), 489-500. https://doi.org/10.1007/s10802-010-9482-2

Beevers, C. G., Rohde, P., Stice, E., & Nolen-Hoeksema, S. (2007). Recovery from major depressive disorder among female adolescents: A prospective test of the scar hypothesis. *Journal of Consulting and Clinical Psychology*, *75*(6), 888-900. https://doi.org/10.1037/0022-006X.75.6.888

Bernaras, E., Jaureguizar, J., & Garaigordobil, M. (2019). Child and Adolescent Depression: A Review of Theories, Evaluation Instruments, Prevention Programs, and Treatments [Review]. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00543

Brandes, C. M., Kushner, S. C., Herzhoff, K., & Tackett, J. L. (2021). Facet-level personality development in the transition to adolescence: Maturity, disruption, and gender differences. Journal of Personality and Social Psychology, *121*(5), 1095-1111. https://doi.org/10.1037/pspp0000367

*Bress, J. N., Foti, D., Kotov, R., Klein, D. N., & Hajcak, G. (2013). Blunted neural response to rewards prospectively predicts depression in adolescent girls. Psychophysiology, *50*(1), 74-81. https://doi.org/10.1111/j.1469-8986.2012.01485.x

*Bufferd, S. J., Dougherty, L. R., Olino, T. M., Dyson, M. W., Laptook, R. S., Carlson, G. A., & Klein, D. N. (2014). Predictors of the onset of depression in young children: A multi-method, multi-informant longitudinal study from ages 3 to 6. Journal of Child Psychology and Psychiatry, *55*(11), 1279-1287. https://doi.org/10.1111/jcpp.12252

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., Meier, M. H., Ramrakha, S., Shalev, I., Poulton, R., & Moffitt, T. E. (2014). The p Factor: One General Psychopathology Factor in the Structure of Psychiatric Disorders? *Clinical Psychological Science*, *2*(2), 119-137. https://doi.org/10.1177/2167702613497473

Charrois T. L. (2015). Systematic reviews: what do you need to know to get started?. The Canadian journal of hospital pharmacy, *68*(2), 144–148. https://doi.org/10.4212/cjhp.v68i2.1440

Chen, H. & Cohen, P. & Chen, S. (2010). How Big is a Big Odds Ratio? Interpreting the Magnitudes of Odds Ratios in Epidemiological Studies. Communications in Statistics

- Simulation and Computation. *39*. 860-864. https://doi.org/10.1080/03610911003650383

Chen, Z., Shen, S., & Dai, Q. (2021). Long-term and short-term psycho-social predictors of early-adulthood depression: Role of childhood trauma, neuroticism, social-support, resilience, and life-events. Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues. https://doi.org/10.1007/s12144-021-01570-5

Cipriani, A., Zhou, X., Del Giovane, C., Hetrick, S. E., Qin, B., Whittington, C., Coghill, D., Zhang, Y., Hazell, P., Leucht, S., Cuijpers, P., Pu, J., Cohen, D., Ravindran, A. V., Liu, Y., Michael, K. D., Yang, L., Liu, L., & Xie, P. (2016). Comparative efficacy and tolerability of antidepressants for major depressive disorder in children and adolescents: A network meta-analysis. The Lancet, *388*(10047), 881-890. https://doi.org/10.1016/S0140-6736(16)30385-3

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society*. Series B (Methodological), *34,* 187–220.

*Davey, C. G., Whittle, S., Harrison, B. J., Simmons, J. G., Byrne, M. L., Schwartz, O. S., & Allen, N. B. (2015). Functional brain-imaging correlates of negative affectivity and the onset of first-episode depression. Psychological Medicine, *45*(5), 1001-1009. https://doi.org/10.1017/S0033291714002001

Dawson, D. N., Jones, M. C., Fairbairn, C. E., & Laurent, H. K. (2022). Mindfulness and acceptance-based interventions with Black Americans: A meta-analysis of intervention efficacy for depressive symptoms. Journal of Consulting and Clinical Psychology, *90*(2), 123-136. https://doi.org/10.1037/ccp0000706

De Bolle, M., De Fruyt, F., McCrae, R. R., Löckenhoff, C. E., Costa, P. T., Jr., Aguilar-Vafaie, M. E., Ahn, C.-k., Ahn, H.-n., Alcalay, L., Allik, J., Avdeyeva, T. V., Bratko, D.,

Brunner-Sciarra, M., Cain, T. R., Chan, W., Chittcharat, N., Crawford, J. T., Fehr, R., Ficková, E., Terracciano, A. (2015). The emergence of sex differences in personality traits in early adolescence: A cross-sectional, cross-cultural study. Journal of Personality and Social Psychology, *108*(1), 171-185. https://doi.org/10.1037/a0038497

El-Monshed, A. H., Fayed, S. M., & Loutfy, A. (2021). Prevention and treatment of mental health and psychosocial problems in college students. Clinical Psychology: Science and Practice, *28*(3), 250-253. https://doi.org/10.1037/cps0000040

Fried, E. I., Nesse, R. M., Zivin, K., Guille, C., & Sen, S. (2014). Depression is more than the sum score of its parts: individual DSM symptoms have different risk factors. Psychological Medicine, *44*(10), 2067-2076. https://doi.org/10.1017/S0033291713002900

*Goldstein, B. L., Kotov, R., Perlman, G., Watson, D., & Klein, D. N. (2018). Trait and facet-level predictors of first-onset depressive and anxiety disorders in a community sample of adolescent girls. Psychological Medicine, *48*(8), 1282-1290. https://doi.org/10.1017/S0033291717002719

Gollan, J., Raffety, B., Gortner, E., & Dobson, K. (2005). Course profiles of early- and adult-onset depression. *Journal of Affective Disorders*, *86*(1), 81-86. https://doi.org/10.1016/j.jad.2004.12.009

Grist, C. L., & McCord, D. M. (2010). Individual differences in preschool children: Temperament or personality? Infant and Child Development, *19*, 264–274. https://doi.org/10.1002/icd.663 .

Gulley, L. D., Hankin, B. L., & Young, J. F. (2016). Risk for depression and anxiety in youth: The interaction between negative affectivity, effortful control, and stressors. Journal of

Abnormal Child Psychology, *44*(2), 207-218. https://doi.org/10.1007/s10802-015-9997-7

Haddad, M., & Boyce, P. (2023). Depression (5th ed., Ser. Fast facts). *Epidemiology and impact,* 39–64. Karger. February 8, 2024

Hakulinen, C., Elovainio, M., Pulkki, R. L., Virtanen, M., Kivimäki, M., & Jokela, M. (2015). Personality and depressive symptoms: Individual participant meta-analysis of 10 cohort studies. *Depression and Anxiety*, *32*(7), 461–470. https://doi-org.proxy-ub.rug.nl/10.1002/da.22376

Hammen, C. (1991). Generation of stress in the course of unipolar depression. *Journal of Abnormal Psychology*, *100*, 555–561.

Hammen, C., Brennan, P. A., Keenan-Miller, D., & Herr, N. R. (2008). Early onset recurrent subtype of adolescent depression: Clinical and psychosocial correlates. *Journal of Child Psychology and Psychiatry*, *49*(4), 433-440. https://doi.org/10.1111/j.1469-7610.2007.01850.x

Hasin, D. S., Goodwin, R. D., Stinson, F. S., & Grant, B. F. (2005). Epide- miology of major depressive disorder: Results from the National Epide- miologic Survey on Alcoholism and Related Conditions. Archives of General Psychiatry, *62*, 1097–1106.

Institute of Health Metrics and Evaluation, Global Burden of Disease (2020) – processed by Our World in Data. “Ages 5-14” [dataset]. IHME, Global Burden of Disease (2020) [original data]. Retrieved on February 15, 2024. https://ourworldindata.org/grapher/depressive-disorders-prevalence-by-age?time=2019

Jylhä, P., Melartin, T., Rytsälä, H., & Isometsä, E. (2009). Neuroticism, introversion, and major depressive disorder—traits, states, or scars? Depression and Anxiety, *26*(4), 325-334. https://doi.org/10.1002/da.20385

Katz, R., & McGuffin, P. (1987). Neuroticism in familial depression. *Psychological Medicine*, *17*(1), 155–161. https://doi-org.proxy-ub.rug.nl/10.1017/S0033291700013052

Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C., & Eaves, L. J. (1993). A longitudinal twin study of personality and major depression in women. Archives of General Psychiatry, *50*(11), 853-862. https://doi.org/10.1001/archpsyc.1993.01820230023002

*Kendall, A. D., Zinbarg, R. E., Mineka, S., Bobova, L., Prenoveau, J. M., Revelle, W., & Craske, M. G. (2015). Prospective associations of low positive emotionality with first onsets of depressive and anxiety disorders: Results from a 10-wave latent trait-state modeling study. Journal of Abnormal Psychology, *124*(4), 933-943. https://doi.org/10.1037/abn0000105

Klein, D. N., Kotov, R., & Bufferd, S. J. (2011). Personality and depression: Explanatory models and review of the evidence. Annual Review of Clinical Psychology, 7, 269-295. https://doi.org/10.1146/annurev-clinpsy-032210-104540

König, H., König, H.-H., & Konnopka, A. (2020). The excess costs of depression: A systematic review and meta-analysis. *Epidemiology and Psychiatric Sciences*, *29*. https://doi-org.proxy-ub.rug.nl/10.1017/S2045796019000180

Kopala-Sibley, D. C., Klein, D. N., Perlman, G., & Kotov, R. (2017). Self-criticism and dependency in female adolescents: Prediction of first onsets and disentangling the relationships between personality, stressful life events, and internalizing psychopathology. Journal of Abnormal Psychology, *126*(8), 1029-1043. https://doi.org/10.1037/abn0000297

Kushner, S. C. (2015). A review of the direct and interactive effects of life stressors and dispositional traits on youth psychopathology. *Child Psychiatry and Human*

Development, *46*(5), 810–819. https://doi-org.proxy-ub.rug.nl/10.1007/s10578-014-0523-x

Lahey, B. B. (2009) Public health significance of neuroticism. Am Psychol., *64*(4), 241-256. doi:10.1037/a0015309

Laidra, K., De Fruyt, F., & Konstabel, K. (2017). Assessing childhood personality with the Estonian short version of the Hierarchical Personality Inventory for Children (HiPIC). *Personality and Individual Differences*, *112*, 31-36. https://doi.org/10.1016/j.paid.2017.02.050

Lehto, K., Karlsson, I., Lundholm, C., & Pedersen, N. L. (2018). Genetic risk for neuroticism predicts emotional health depending on childhood adversity. Psychological Medicine. https://doi.org/10.1017/S0033291718000715

Lewinsohn, P. M, Steinmetz, J., Larson, D., & Franklin, J. (1981). Depression related cognitions: Antecedents or consequences? Journal of Abnormal Psychology, *90*, 213–219.

Lewis, M. D. (1997). Personality self-organization: Cascading constraints on cognition-emotion interaction. In A. Fogel, M. C. D. P. Lyra, & J. Valsiner (Eds.), *Dynamics and indeterminism in developmental and social processes,* 193–216. Hillsdale, NJ: Erlbaum

Lewis, A. J., Sae-Koew, J. H., Toumbourou, J. W., & Rowland, B. (2020). Gender differences in trajectories of depressive symptoms across childhood and adolescence: A multi-group growth mixture model. Journal of Affective Disorders, *260*, 463-472. https://doi.org/10.1016/j.jad.2019.09.027

Li, W., Yang, Y., An, F.-R., Zhang, L., Ungvari, G. S., Jackson, T., Yuan, Z., & Xiang, Y.-T. (2020). Prevalence of comorbid depression in schizophrenia: A meta-analysis of

observational studies. *Journal of Affective Disorders*, *273*, 524–531. https://doi-org.proxy-ub.rug.nl/10.1016/j.jad.2020.04.056

Liu, Y., Zhang, N., Bao, G., Huang, Y., Ji, B., Wu, Y., Liu, C., & Li, G. (2019). Predictors of depressive symptoms in college students: A systematic review and meta-analysis of cohort studies. *Journal of Affective Disorders*, *244*, 196–208. https://doi-org.proxy-ub.rug.nl/10.1016/j.jad.2018.10.084

LoPilato, A. M., Pike, M. R., Holton, J. E., & Craighead, W. E. (2023). Parent-based intervention to enhance treatment of adolescent depression. Practice Innovations. https://doi.org/10.1037/pri0000223

Luby, J. L., Heffelfinger, A. K., Mrakotsky, C., Hessler, M. J., Brown, K. M., & Hildebrand, T. (2002). Preschool major depressive disorder: Preliminary validation for developmentally modified DSM-V criteria. *Journal of the American Academy of Child & Adolescent Psychiatry*, *41*(8), 928–937. https://doi-org.proxy-ub.rug.nl/10.1097/00004583-200208000-00011

Lyon, K. A., Elliott, R., Ware, K., Juhasz, G., & Brown, L. J. E. (2021). Associations between facets and aspects of Big Five personality and affective disorders: A systematic review and best evidence synthesis. *Journal of Affective Disorders*, *288*, 175–188. https://doi-org.proxy-ub.rug.nl/10.1016/j.jad.2021.03.061

Mackin, D. M., Finsaas, M. C., Nelson, B. D., Perlman, G., Kotov, R., & Klein, D. N. (2022). Intergenerational transmission of depressive and anxiety disorders: Mediation via youth personality. Journal of Psychopathology and Clinical Science, *131*(5), 467-478. https://doi.org/10.1037/abn0000759

McCrae, R. R., & Costa, P. T., Jr. (2013). Introduction to the empirical and theoretical status of the five-factor model of personality traits. In T. A. Widiger & P. T. Costa, Jr. (Eds.),

Personality disorders and the five-factor model of personality., 3rd ed. (pp. 15-27). American Psychological Association. https://doi.org/10.1037/13939-002

McHugh M. L. (2012). Interrater reliability: the kappa statistic. Biochemia medica, *22*(3), 276–282. https://doi.org/10.11613/bm.2012.031

Metts, A., Zinbarg, R., Hammen, C., Mineka, S., & Craske, M. G. (2021). Extraversion and interpersonal support as risk, resource, and protective factors in the prediction of unipolar mood and anxiety disorders. Journal of Abnormal Psychology, *130*(1), 47-59. https://doi.org/10.1037/abn0000643

Michelini, G., Perlman, G., Tian, Y., Mackin, D. M., Nelson, B. D., Klein, D. N., & Kotov, R. (2021). Multiple domains of risk factors for first onset of depression in adolescent girls. Journal of Affective Disorders, 283, 20-29. https://doi.org/10.1016/j.jad.2021.01.036

Mu, W., Huang, C., Yao, N., Miao, J., Perlman, G., Watson, D., Klein, D. N., & Kotov, R. (2023a). Developmental pathway for first onset of depressive disorders in females: From adolescence to emerging adulthood. Psychological Medicine. https://doi.org/10.1017/S0033291723002441

Mu, W., Li, K., Tian, Y., Perlman, G., Michelini, G., Watson, D., Ormel, H., Klein, D. N., & Kotov, R. (2023b). Dynamic risk for first onset of depressive disorders in adolescence: Does change matter? Psychological Medicine. https://doi.org/10.1017/S0033291721004190

Muris, P., Meesters, C., & Diederen, R. (2005). Psychometric properties of the Big Five Questionnaire for Children (BFQ-C) in a Dutch sample of young adolescents. *Personality and Individual Differences*, *38*(8), 1757-1769. https://doi.org/10.1016/j.paid.2004.11.018

Olino, T. M., Mennies, R. J., & Wojcieszak, Z. K. (2020). Personality-stress vulnerability models. In K. L. Harkness & E. P. Hayden (Eds.), *The Oxford handbook of stress and mental health.* 397–413. Oxford University Press.

Olivier, M., & Herve, M. (2015). The Big Five Questionnaire for Children (BFQ-C): A French validation on 8- to 14-year-old children. *Personality and Individual Differences*, *87*, 55-58. https://doi.org/10.1016/j.paid.2015.07.030

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016) Rayyan — a web and mobile app for systematic reviews. Systematic Reviews, 5:210, https://doi.org/10.1186/s13643-016-0384-4

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron I., Hoffmann, T. C., Mulrow, C. D., et al.. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71

Patton, G. C., Coffey, C., Posterino, M., Carlin, J. B., & Bowes, G. (2003). Life events and early onset depression: Cause or consequence? Psychological Medicine, *33*(7), 1203-1210. https://doi.org/10.1017/S0033291703008626

Paus, T., Keshavan, M., & Giedd, J. N. (2008). Why do many psychiatric disorders emerge during adolescence? Nature Reviews Neuroscience, *9*(12), 947-957. https://doi.org/10.1038/nrn2513

Perlman, G., Gromatsky, M., Salis, K. L., Klein, D. N., & Kotov, R. (2018). Personality correlates of self-injury in adolescent girls: Disentangling the effects of lifetime psychopathology. Journal of Abnormal Child Psychology, *46*(8), 1677-1685. https://doi.org/10.1007/s10802-018-0403-0

Pettit, J. W., Lewinsohn, P. M., & Joiner, T. E., Jr. (2006). Propagation of major depressive disorder: Relationship between first episode symptoms and recurrence. Psychiatry Research, *141*(3), 271-278. https://doi.org/10.1016/j.psychres.2005.07.022

Prince, E. J., Siegel, D. J., Carroll, C. P., Sher, K. J., & Bienvenu, O. J. (2021). A longitudinal study of personality traits, anxiety, and depressive disorders in young adults. *Anxiety, Stress & Coping: An International Journal*, *34*(3), 299-307. https://doi.org/10.1080/10615806.2020.1845431

Ramos-Grille, I., Aragay, N., Valero, S., Garrido, G., Santacana, M., Guillamat, R., Vallès, V., & Gomà-i-Freixanet, M. (2022). Relationship between depressive disorders and personality traits: The value of the alternative five factor model. Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues, *41*(8), 5058-5064. https://doi.org/10.1007/s12144-020-01005-7

Ramskogler, K., & Bruckner, T. (2014). Beating the odds: An approach to the topic of resilience in children and adolescents. *neuropsychiatrie*, *28*. https://doi.org/10.1007/s40211-014-0125-7

Rohde, P., Lewinsohn, P. M., & Seeley, J. R. (1990). Are people changed by the experience of having an episode of depression? A further test of the scar hypothesis. *Journal of Abnormal Psychology*, *99*(3), 264-271. https://doi.org/10.1037/0021-843X.99.3.264

Rice, F., Riglin, L., Lomax, T., Souter, E., Potter, R., Smith, D. J., Thapar, A. K., & Thapar, A. (2019). Adolescent and adult differences in major depression symptom profiles. Journal of Affective Disorders, *243*, 175-181. https://doi.org/10.1016/j.jad.2018.09.015

Saha, S., Lim, C. C., Degenhardt, L., Cannon, D. L., Bremner, M., Prentis, F., Lawrence, Z., Heffernan, E., Meurk, C., Reilly, J., & McGrath, J. J. (2022). Comorbidity between

mood and substance-related disorders: A systematic review and meta-analysis. *Australian and New Zealand Journal of Psychiatry*, *56*(7), 757–770. https://doi-org.proxy-ub.rug.nl/10.1177/00048674211054740

Saha, S., Lim, C. C. W., Cannon, D. L., Burton, L., Bremner, M., Cosgrove, P., Huo, Y., & McGrath, J. (2021). Co-morbidity between mood and anxiety disorders: A systematic review and meta-analysis. *Depression and Anxiety*, *38*(3), 286–306. https://doi-org.proxy-ub.rug.nl/10.1002/da.23113

Sandstrom, A., Perroud, N., Alda, M., Uher, R., & Pavlova, B. (2021). Prevalence of attention-deficit/hyperactivity disorder in people with mood disorders: A systematic review and meta-analysis. *Acta Psychiatrica Scandinavica*, *143*(5), 380–391. https://doi-org.proxy-ub.rug.nl/10.1111/acps.13283

Silver, J., Carlson, G. A., Olino, T. M., Perlman, G., Mackin, D., Kotov, R., & Klein, D. N. (2021). Differential outcomes of tonic and phasic irritability in adolescent girls. Journal of Child Psychology and Psychiatry, *62*(10), 1220-1227. https://doi.org/10.1111/jcpp.13402

Sosnowska, J., Kuppens, P., De Fruyt, F., & Hofmans, J. (2020). New directions in the conceptualization and assessment of personality—A dynamic systems approach. European Journal of Personality, *34*(6), 988-998. https://doi.org/10.1002/per.2233

Steiger, A. E., Fend, H. A., & Allemand, M. (2015). Testing the vulnerability and scar models of self-esteem and depressive symptoms from adolescence to middle adulthood and across generations. *Developmental Psychology*, *51*(2), 236-247. https://doi.org/10.1037/a0038478

Subica, A. M., Fowler, J. C., Elhai, J. D., Frueh, B. C., Sharp, C., Kelly, E. L., & Allen, J. G. (2014). Factor structure and diagnostic validity of the Beck Depression Inventory–II

with adult clinical inpatients: Comparison to a gold-standard diagnostic interview. *Psychological Assessment*, *26*(4), 1106–1115. https://doi-org.proxy-ub.rug.nl/10.1037/a0036998.

Tachi, S., Asamizu, M., Uchida, Y., Katayama, S., Naruse, M., Masuya, J., Ichiki, M., & Inoue, T. (2019). Victimization in childhood affects depression in adulthood via neuroticism: A path analysis study. Neuropsychiatric Disease and Treatment, *15*. https://doi.org/10.2147/NDT.S220349

Todd, R. D., & Botteron, K. N. (2002). Etiology and genetics of early-onset mood disorders. *Child and Adolescent Psychiatric Clinics of North America*, *11*(3), 499-518. https://doi.org/10.1016/S1056-4993(02)00013-5

Toenders, Y. J., Kottaram, A., Dinga, R., Davey, C. G., Banaschewski, T., Bokde, A. L. W., Quinlan, E. B., Desrivières, S., Flor, H., Grigis, A., Garavan, H., Gowland, P., Heinz, A., Brühl, R., Martinot, J.-L., Paillère Martinot, M.-L., Nees, F., Orfanos, D. P., Lemaitre, H., . . . Schmaal, L. (2022). Predicting Depression Onset in Young People Based on Clinical, Cognitive, Environmental, and Neurobiological Data. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *7*(4), 376-384. https://doi.org/https://doi.org/10.1016/j.bpsc.2021.03.005

Tonarely, N. A., Sherman, J. A., Grossman, R. A., Shaw, A. M., & Ehrenreich-May, J. (2020). Neuroticism as an underlying construct in youth emotional disorders. Bulletin of the Menninger Clinic, *84*(3), 214-236. https://doi.org/10.1521/bumc.2020.84.3.214

Uliaszek, A. A., Zinbarg, R. E., Mineka, S., Craske, M. G., Sutton, J. M., Griffith, J. W., Rose, R., Waters, A., & Hammen, C. (2010). The role of neuroticism and extraversion in the stress-anxiety and stress-depression relationships. Anxiety, Stress & Coping: An International Journal, *23*(4), 363-381. https://doi.org/10.1080/10615800903377264

Valerio, M. P., Blasco, B., Tagni, F., Szmulewicz, A. G., & Martino, D. J. (2020). Personality disturbances in melancholic and nonmelancholic unipolar major depression: A systematic review and meta-analysis. *Journal of Nervous and Mental Disease*, *208*(10), 810–817. https://doi-org.proxy-ub.rug.nl/10.1097/NMD.0000000000001212

van Eeden, W. A., van Hemert, A. M., Carlier, I. V. E., Penninx, B. W., Spinhoven, P., & Giltay, E. J. (2019). Neuroticism and chronicity as predictors of 9-year course of individual depressive symptoms. Journal of Affective Disorders, *252*, 484-492. https://doi.org/10.1016/j.jad.2019.04.052

Vasey, M. W., Harbaugh, C. N., Lonigan, C. J., Phillips, B. M., Han- kin, B. L., Willem, L., & Bijttebier, P. (2013). Dimensions of temperament and depressive symptoms: Replicating a three-way interaction. Journal of Research in Personality, *47*, 908–921. http://dx.doi.org/10.1016/j.jrp.2013.09.001

Vikander, M., & Strand, S. (2023). Enhancing domestic violence risk assessments with children's perspectives: Exploring risk, vulnerability and protective factors through forensic interviews. Child & Family Social Work. https://doi.org/10.1111/cfs.13068

Vinkers, C. H., Joëls, M., Milaneschi, Y., Kahn, R. S., Penninx, B. W. J. H., & Boks, M. P. M. (2014). Stress exposure across the life span cumulatively increases depression risk and is moderated by neuroticism. Depression and Anxiety, *31*(9), 737-745. https://doi.org/10.1002/da.22262

Watson, D., & Clark, L. A. (1992). On traits and temperament: General and specific factors of emotional experience and their relation to the five-factor model. Journal of Personality, *60*(2), 441-476. https://doi.org/10.1111/j.1467-6494.1992.tb00980.x (The five-factor model: Issues and applications)

Wesselhoeft, R., Heiervang, E. R., Kragh-Sørensen, P., Juul Sørensen, M., & Bilenberg, N. (2016). Major depressive disorder and subthreshold depression in prepubertal children from the Danish National Birth Cohort. Comprehensive Psychiatry, *70*, 65-76. https://doi.org/10.1016/j.comppsych.2016.06.012

*Wilson, S., Vaidyanathan, U., Miller, M. B., McGue, M., & Iacono, W. G. (2014). Premorbid risk factors for major depressive disorder: Are they associated with early onset and recurrent course? *Development and Psychopathology*, *26*(4, Pt 2), 1477-1493. https://doi.org/10.1017/S0954579414001151

Wu, Y., Levis, B., Daray, F. M., Ioannidis, J. P. A., Patten, S. B., Cuijpers, P., Ziegelstein, R. C., Gilbody, S., Fischer, F. H., Fan, S., Sun, Y., He, C., Krishnan, A., Neupane, D., Bhandari, P. M., Negeri, Z., Riehm, K. E., Rice, D. B., Azar, M., . . . et al.. (2023). Comparison of the accuracy of the 7-item HADS Depression subscale and 14-item total HADS for screening for major depression: A systematic review and individual participant data meta-analysis. Psychological Assessment, *35*(2), 95-114. https://doi.org/10.1037/pas0001181

Zinbarg, R. E., Mineka, S., Craske, M. G., Griffith, J. W., Sutton, J., Rose, R. D., Nazarian, M., Mor, N., & Waters, A. M. (2010). The Northwestern-UCLA youth emotion project: Associations of cognitive vulnerabilities, neuroticism and gender with past diagnoses of emotional disorders in adolescents. Behaviour Research and Therapy, *48*(5), 347-358. https://doi.org/10.1016/j.brat.2009.12.008

Zisook, S., Rush, A. J., Lesser, I., Wisniewski, S. R., Trivedi, M., Husain, M. M., Balasubramani, G. K., Alpert, J. E., & Fava, M. (2007). Preadult onset vs adult onset of major depressive disorder: A replication study. Acta Psychiatrica Scandinavica, *115*(3), 196-205. https://doi.org/10.1111/j.1600-0447.2006.00868.x

Appendix A - Web of Science Search Terms

(((TI=(Depression or Depressive or "depressive symptom*" or MDD or "Major depressive disorder" or "mood disorder*" or "emotional disorder*") OR TI=(Anxiety or "Anxiety disorder")) OR KP=(Depression or Depressive* or MDD or "Major depressive disorder" or "mood disorder*" or "emotional disorder*" or Anxiety or "Anxiety disorder"))) AND (TS=(risk or risk factor* or contributing factor* or predisposing factor* or predict* or cause or vulnerability factor* or vulnerability) AND TS=(longitudinal or aetiolog* or "twin study" or "follow?up" or cohort or onset or "first onset" or "initial onset" or "first episode" or "initial onset" or "first episode")) AND (TS=(Child* or youth or adolescen* or "young people" or teen) OR TS=("age = 6" or "age = 7" or "age = 8" or "age = 9" or "age = 10" or "age = 11" or "age = 12" or "age = 13" or "age = 14" or "age = 15" or "age = 16" or "age = 17" or "age = 18")) NOT TI=("bipolar" or "*therapy" or "suicid*" or "post traumatic*" or autism or "attention deficit" or "*apnea" or "*somia" or "psychosis" or treatment or therapy or intervention) AND (DT==("ARTICLE") AND LA==("ENGLISH"))*.

Note: Search terms as were used in the Web of Science advanced search.
Web of Science search abbreviations: Title (TI); Keyword plus (KP); Topic (TS); Date (DT); Language (LA).
Web of Science symbols: Add multiple characters/suffixes (*; e.g. -atology, -s, -gic); Look for exact term ("); Add one character/symbol (?; e.g. -e, -d, -).

Appendix B - Reasons for Exclusion of Eligible Reports

Table 5

Reference and reasons for exclusion of eligible studies (n = 8) excluded from this systematic review of the predictive value of neuroticism for first onset depression.

Reference	Reason for Exclusion
Kopala-Sibley et al. (2017) Michelini et al. (2021) Mu et al. (2023a) Mu et al. (2023b) Perlman et al. (2018) Silver et al. (2021)	The work of Goldstein et al. (2018) was chosen over these other reports due to its focus on personality traits (as compared to Michelini et al, 2021) and a main focus on the personality trait neuroticism (as compared to Perlman et al., 2018; Mu et al, 2023a; Mu et al., 2023b; Silver et al., 2021) with additional insight in results (as compared to Kopala-Sibley et al, 2017)
Metts et al. (2021)	Main personality trait used in this report is Extraversion. Given that neuroticism is the main focus of this review, the works of Kendall et al. (2015) that focusses on neuroticism was picked for reviewing.
Zinbarg et al. (2010)	The work of Kendall et al. (2015) was chosen over this reports as the design used a multilayered hazard ratio that focussed multiple measurement waves. It therefore holds additional information that is not yet available from the works of Zinbarg et al. (2010).

Investigator 1	Investigator 2	Overlap start	Overlap end
Inv. A	Inv. B	1	40
Inv. A	Inv. C	261	300
Inv. A	Inv. D	521	560
Inv. A	Inv. E	781	820
Inv. A	Inv. F	1041	1080
Inv. B	Inv. C	1321	1360
Inv. B	Inv. D	1581	1620
Inv. B	Inv. E	1841	1880
Inv. B	Inv. F	2101	2140
Inv. C	Inv. D	2491	2530
Inv. C	Inv. E	2751	2790
Inv. C	Inv. F	3011	3050
Inv. D	Inv. E	3621	3660
Inv. D	Inv. F	4101	4140
Inv. E	Inv. F	4821	4860
Total reports		6000	
10% overlap		600	40
reports/investigor		1100	220

Note. Reports were rated by Rayyan AI from 0.5 stars (lowest) to 5 stars (highest). The highest rated reports (>2.5 stars) were included for review. In this table the order used was the highest rated report (nr. 1) up to the lowest rated report (nr. 6000).

			Inv. A		Inv. B		Inv. C		Inv. D		Inv. E		Inv. F			
	Numeric start	Numeric end	Overlap start	Overlap end	Overlap start	Overlap end	Overlap start	Overlap end	Overlap start	Overlap end	Overlap start	Overlap end	Overlap start	Overlap end	Reports screened	Total overlap
Inv. A	1	1100	-	-											1100	5
Inv. B	1101	2160	1	40	-	-									1100	44
Inv. C	2161	3180	261	300	1321	1360	-	-							1100	83
Inv. D	3181	4160	521	560	1581	1620	2491	2530	-	-					1100	122
Inv. E	4161	5100	781	820	1841	1880	2751	2790	3621	3660	-	-			1100	161
Inv. F	5101	6000	1041	1080	2101	2140	3011	3050	4101	4140	4821	4860	-	-	1100	200

Notes. Abreviation*:* Investigator (inv.); Numeric start and -end are given by the order of reports as (arbitrary) numbered by the Rayyan systematic reviewing program.

Investigator 1	Investigator 2	Overlap start	Overlap end
Inv. A	Inv. B	1	4
Inv. A	Inv. C	24	27
Inv. A	Inv. D	47	50
Inv. A	Inv. E	71	73
Inv. A	Inv. F	94	96
Inv. B	Inv. C	119	121
Inv. B	Inv. D	142	144
Inv. B	Inv. E	165	168
Inv. B	Inv. F	188	191
Inv. C	Inv. D	223	226
Inv. C	Inv. E	246	249
Inv. C	Inv. F	269	272
Inv. D	Inv. E	324	326
Inv. D	Inv. F	367	369
Inv. E	Inv. F	431	433
Total reports		535	
10% overlap		54	4
reports/investigor		98	20

Note. Approved reports after title and abstract selection procedure. In this table the order of reports was arbitrary.

			Inv. A		Inv. B		Inv. C		Inv. D		Inv. E		Inv. F			
	Numeric start	Numeric end	Overlap start	Overlap end	Overlap start	Overlap end	Overlap start	Overlap end	Overlap start	Overlap end	Overlap start	Overlap end	Overlap start	Overlap end	Reports screened	Total overlap
Inv. A	1	98	-	-											98	5
Inv. B	99	193	1	4	-	-									98	8
Inv. C	194	284	24	27	119	121	-	-							98	10
Inv. D	285	371	47	50	142	144	223	226	-	-					98	13
Inv. E	372	455	71	73	165	168	246	249	324	326	-	-			98	15
Inv. F	456	535	94	96	188	191	269	272	367	369	431	433	-	-	98	18

Notes. Abreviation*:* Investigator (inv.); Numeric start and -end are given by the order of reports as (arbitrary) numbered by the Rayyan systematic reviewing program.