UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

FACULDADE DE MEDICINA

Programa de Pós-Graduação em Psiquiatria e Ciências do Comportamento

Doctoral Thesis

TESTING DATA-DRIVEN THRESHOLDS FOR IRRITABILITITY IN ADOLESCENTS AND YOUNG ADULTS

PAOLA PAGANELLA LAPORTE

Advisor: Prof. Giovanni Abrahão Salum Junior

Porto Alegre, Junho de 2021

TESTING DATA-DRIVEN THRESHOLDS FOR IRRITABILITITY IN ADOLESCENTS AND YOUNG ADULTS

Doctoral thesis, by Paola Laporte.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

FACULDADE DE MEDICINA

Programa de Pós-Graduação em Psiquiatria e Ciências do Comportamento

Thesis submitted to the Graduate Program in Psychiatry and Behavioral Sciences as a partial requirement for obtaining a PhD

Porto Alegre, Brasil

2021

ABSTRACT

Justification: Irritability is the propensity to experience anger relative to peers. It is a common trait that affects children, adolescents and adults, presenting several adverse outcomes through the lifespan. Disruptive Mood Dysregulation Disorder (DMDD) is a novel diagnosis designed to classify children with severe and chronic manifestations of irritable mood and temper outbursts in childhood. The DMDD diagnosis was constructed from clinical descriptions, with very little empirical support for its operationalization in the Diagnostic and Statistical Manual for Mental Disorders, 5th edition (DSM-5). Objective: In this thesis, we explored new data-driven and clinically oriented operationalized criteria for DMDD diagnosis in pre-adolescents/early adolescents and also in young adults. Method: The clinical threshold was assessed in three stages: symptomatic, syndromic and clinical operationalization. The symptomatic threshold identified the response category in each "Development and Well-Being Behavior Assessment (DAWBA) item which separates normative misbehavior from a clinical indicator. The syndromic threshold identified the number of irritable mood and outbursts needed to capture pre-adolescents/early adolescents with high symptom levels. Clinical operationalization compared the impact of AND/OR rules for combining irritable mood and outbursts on impairment and levels of psychopathology. Results: *First,* we used data from 3,562 pre-adolescents/early adolescents from the 2004 Pelotas Birth Cohort with the DAWBA. We found out that most irritable mood items were normative in their lowest response categories and clinically significant in their highest response categories. For outbursts some indicated a symptom even when present at only a mild level, while others did not indicate symptoms at any level. At the syndromic level, a combination of 2 out of 7 irritable mood and 3 out of 8 outburst indicators accurately captured a cluster of individuals with high level of symptoms.

Analysis combining irritable mood and outbursts delineated non-overlapping aspects of DMDD, providing support for the OR rule in clinical operationalization. The best DMDD criteria resulted in a prevalence of 3%. *Second*, we expanded the same methodology to investigate irritability in 1705 adolescents and young adults from the 3rd wave of the Brazilian High-Risk Cohort Study for Mental Conditions. Measurement invariance analysis showed that thresholds for measuring irritability symptoms change between adolescence and adulthood, therefore requiring distinct operational strategies. Symptomatic threshold analyses showed that irritable mood items were considered problem indicators in their highest response category for both age groups. For outbursts, some results suggested a symptom to be significant even at mild levels, while other results found no evidence of clinical significance at any level, with some differences between age groups. At the syndromic level, a combination of 3 out of 8 symptoms of irritable mood and 1 out of 10 symptoms of outbursts accurately captured a cluster of individuals with high level of symptoms in the adolescent group; in young adults 1 out of 8 symptoms of irritable mood and 1 out of 9 symptoms of outbursts are required. Analysis combining irritable mood and outbursts delineated non-overlapping aspects of DMDD. We also presented the prevalence rates for the combination of distinct DMDD diagnostic rules and show a developmental follow-back analysis and experience momentary assessment. Discussion: In summary, this thesis advances in the construction of data-driven and clinically oriented criteria to classify youths and young adults with impairing levels of irritability that might require specialized clinical attention.

Keywords: Child Psychiatry; Irritability; DSM-5; Psychopathology.

RESUMO

Justificativa: A irritabilidade é definida como uma propensão a sentir raiva em relação aos pares. É um traço comum que afeta crianças, adolescentes e adultos, apresentando diversos desfechos adversos ao longo da vida. O Transtorno Disruptivo da Desregulação do Humor (TDDH) é um novo diagnóstico que fora desenvolvido para classificar as crianças com manifestações crônicas e graves de humor irritado e crises de raiva na infância. Este diagnóstico fora construído a partir de descrições clínicas, com pouco suporte empírico para sua operacionalização no Manual Diagnóstico e Estatístico para Transtornos Mentais, 5ª edição (DSM-5). Objetivo: Nesta tese exploramos novos critérios operacionalizados baseados em dados (coleta e análise de informações) e orientados clinicamente para o diagnóstico do TDDH em pré-adolescentes, adolescentes e também em adultos jovens. Método: O limiar clínico foi avaliado em três etapas: operacionalização sintomática, sindrômica e clínica. A operacionalização sintomática identificou a categoria de resposta de cada item do questionário "DAWBA" que separa um mau comportamento normativo de um indicador clínico (ou seja, sintoma). A operacionalização sindrômica identificou o número de itens de humor irritado e de crises de raiva que são necessários para identificar grupos de pré-adolescentes com altos níveis de sintomas. A operacionalização clínica comparou o impacto das regras E/OU para combinar humor irritado e crises de raiva nos níveis de prejuízo clínico e de psicopatologia. Resultados: Primeiramente, utilizamos dados do questionário "DAWBA" de 3.562 pré-adolescentes da "Coorte de Nascimentos de Pelotas", 2004. Na operacionalização sintomática, identificamos que a maioria dos itens de humor irritados são normativos em suas categorias de resposta mais baixas e clinicamente significativos em suas categorias de resposta mais altas. Referente as crises de raiva, alguns itens foram indicativos de sintoma mesmo quando presentes apenas em um nível moderado,

enquanto outros não indicaram sintomas em nenhum nível. Na operacionalização sindrômica, uma combinação de 2 de 7 indicadores de humor irritado e de 3 de 8 indicadores de crise de raiva capturaram com precisão um grupo de indivíduos com alto nível de sintomas. A análise da combinação de humor irritado e crises de raiva demonstrou aspectos não sobrepostos destas dimensões do TDDH, fornecendo suporte para a regra 'OU' (ou humor irritado OU crises de raiva) na operacionalização clínica. Nosso melhor critério para TDDH, estatisticamente construído, resultou em uma prevalência de 3% para este diagnóstico. Após, expandimos a mesma metodologia para investigar a irritabilidade em 1.705 adolescentes e adultos jovens da 3ª onda do Estudo Brasileiro chamado "Coorte de Alto Risco para Transtornos Mentais". A análise de invariância demonstrou que os limiares para medir os sintomas de irritabilidade variam entre a adolescência e a idade adulta, exigindo, portanto, estratégias operacionais distintas para estas faixas etárias. Na operacionalização sintomática, os itens de humor irritadol foram considerados indicadores de problemas em sua categoria de resposta mais alta para ambos os grupos de idade. Para as crises de raiva, alguns itens foram considerados indicadores de problema mesmo nas categorias de resposta mais leves, enquanto outros não foram indicadores de problema em qualquer nível de resposta, com algumas diferenças entre os grupos de idade. Na operacionalização sindrômica, uma combinação de 3 de 8 sintomas de humor irritado e 1 de 10 sintomas de crises de raiva capturou com precisão um grupo de indivíduos com alto nível de sintomas no grupo de adolescentes; em adultos jovens, foram necessários 1 de 8 sintomas de humor irritado e 1 de 9 sintomas de crises de raiva. A análise da combinação de humor irritado e crises de raiva também demonstrou aspectos não sobrepostos destas dimensões do TDDH, fornecendo suporte para a regra 'OU' (ou humor irritado OU crises de raiva) na operacionalização clínica. Ao final, apresentamos as taxas de prevalência para

as combinações de distintas regras diagnósticas para nossas propostas diagnósticas para TDDH. Realizamos também uma análise de sintomas de irritabilidade retrospectiva e uma avaliação momentânea de sintomas usando os participantes da mesma coorte. Discussão: Em resumo, esta tese avança na construção de critérios baseados em evidencia e orientados para classificar clinicamente jovens e adultos com níveis prejudiciais de irritabilidade que possam necessitar de atenção clínica especializada.

Palavras-chave: Psiquiatria da Infância e da Adolescência; Irritabilidade; DSM-5; Psicopatologia.

SUMMARY

Introduction

Systems of psychiatric classification

Psychiatric disorders are poorly defined constructs, which may change radically when viewed from different angles. They are extremely vulnerable to the "observer bias" [1] and, therefore, are likely to result in several low classification accuracies. This situation does not make these phenomena any less real for those who live with patients with problems related to behavior and emotions. The classification of mental disorders, also known as psychiatric nosology or psychiatric taxonomy attempts to categorize and organize this phenomenon. The classification of mental disorders represents a key aspect for psychiatrists and other mental health professionals and is required for fostering a better communication among clinicians and researchers, understanding etiology, testing treatment efficacy, knowing the prevalence of the disorders, healthcare planning and services organization.

The diagnostic criteria in psychiatry are classically based on descriptive-phenomenological models as operationalized in the two widely established systems of psychiatric classification: The Diagnostic and Statistical Manual of Mental Disorders (DSM) [2] and the International Classification of Diseases (ICD) [3]. DSM-5 is based on diagnostic validity and ICD-11 is based on clinical utility. The latter is reliant on the diagnostic validity. Furthermore, there is a considerable overlap between these systems. It is quite likely that clinicians who use the DSM nonetheless employ prototypic thinking in approaching diagnosis and evaluation in the clinic.

The DSM has provided a major foundation for nosology ever since the DSM-III [4], in 1980, defined categories using operationalized criteria that increased diagnostic

reliability. DSM-IV added changes based on evidence that further increased reliability of psychiatric diagnosis. This advance in psychiatric nosology was largely responsible for the improvement of the diagnostic system, allowing a more adequate communication between professionals and patients and progress in clinical research [5–7]. DSM-III and DSM-IV were widely adopted by clinicians, researchers, and regulatory authorities.

DSM-5 provided an opportunity for re-thinking diagnostic classification systems. As our classification systems are not prepared to adopt classifications based on etiological mechanisms, abandoning the phenomenological orientations, DSM-5 (now described with an Arabic numeral instead of the Romans, to allow more frequent updates, such as 5.1), adopted the "epistemic iteration". Epistemic iteration is defined as a process by which scientific knowledge claims are progressively altered and refined via self-correction or enrichment. As a consequence, DSM has become a model that, despite its failures, after every update it becomes closer to a true natural presentation of mental disorders. The constant evaluation and empirical testing of the DSM criteria is essential, given its eventual arbitrariness, and important for the progress of the diagnostic classification. The hope that neuroscience would develop to the point that categories were defined by biomarkers has not been realized and the process of revising the DSM-5 has had to proceed with other kinds of validation evidence. DMS-5 made incremental changes in the overall structure, and in the diagnostic criteria for some categories. Psychiatric diagnosis continues to suffer from relatively low reliability in clinical settings [8,9], and diagnoses continue to rely on clinical phenomenology rather than on biomarkers.

It seems increasingly likely that many psychiatric disorders are conditions with overlapping fuzzy boundaries with multiple interacting causes acting on multiple brain

mechanisms. Instead of essentialized diseases, they are best understood in terms of more central paradigms [10]. If this is the reality, then the DSM system may describe it fairly well. The problem is not our diagnostic criteria, the problem is that our expectations are based on an oversimplified medical model in which disorders are each imagined to be discrete with specific causes and biomarkers.

A classification based on more objective criteria may become an important tool for identifying individuals who are not being fully evaluated by current criteria. This classification would gain strength if it could be operationalized in such a way to be tested and replicated, if it had a predictive capacity for important deficits in functioning and if it were associated with biological correlates. The formulation of objective phenotypic classification seems to be a relevant strategy for testing whether this operationalization may influence clinical practice in the future.

Definition

Anger is a normal emotion. Like other emotions, it presents elements of subjective feelings, overt behavior of various kinds, and bodily changes. There are several lines for the development of anger during childhood and adolescence, such as the intensity and frequency of occurrence of the mood, the environmental features that evoke it, its expression in behavior, its effect on and modification by other people, and the extent to which the person controls it. Anger can be either functional or dysfunctional, according to context [11].

Irritability is defined by proneness to anger relative to peers [12]. It is a trait characterized by a proneness to feel anger, think and react aggressively towards peers. Irritability has at least three components. The first one is the affective component which encompasses irritable mood. The second is the behavioral component characterized by reactive aggression (anger-related behaviors towards others) and temper outbursts (anger-related behaviors directed towards objects or self). The third is the cognitive component, identified as irritable rumination. It is also seen as a tendency to experience anger and frustration [13,14] and as an increased sensitivity to environmental and internal stimuli [15]. Stringaris and Taylor [11] have noted that irritability is a unique trait. On one hand it shares a negative valence aspect with anxiety and depression; but on the other, contrary to anxiety and depression that typically lead to avoidance, anger frequently leads to approach-motivations. Irritability also crosses the boundaries of internalizing and externalizing symptoms [13].

Irritability in the clinic

Irritability is one of the most common reasons for referring children and adolescents to psychiatric care. Around 3% to 20% of youths are taken to services to assess irritability symptoms [16–18]. Also, it is particularly associated with adjustment problems, such as social difficulties [19–21] and it is present in the context of anxiety, depression and disruptive behaviors [22]. The symptoms of irritability are not only debilitating in childhood, but also have longitudinal associations with psychopathology and poor adjustment in several domains (such as education and peer relationships [23]) during adolescence and in adulthood [19,24].

Irritability is the core symptom of three diagnoses in DSM-5 [2]: Intermittent Explosive Disorder (IED), Oppositional Defiant Disorder (ODD) and Disruptive Mood Dysregulation Disorder (DMDD). Furthermore, irritability represents one of the most common symptoms reported in children with a range of psychiatric conditions, including major depressive disorder (MDD), attention-deficit/hyperactivity disorder (ADHD), conduct disorder, interpersonal trauma and exposure to traumatic events and bipolar disorder (BD) [13,25,26] .

A historical perspective on the developmental course of irritability and the operationalization of DMDD

In the mid two thousand, there was an important increase in the diagnosis of BD in children and adolescents. The number of pre-pubertal children diagnosed with BD in the USA had increased at rates close to 500% in that time [27]. This increase was thought to have resulted partially from counting severe and chronic irritability of early onset as a cardinal manic symptom of pediatric BD, analogous to the classical cardinal manic symptoms of elated mood or episodic irritability. Following that, an increase in the use of antipsychotics drugs was observed [28]. Within this framework, the "pediatric bipolar debate" arose. Irritability was considered by some researchers to be the most common early manifestation of BD. This line of research considered chronic irritability, in association with attention-deficit-hyperactivity disorder symptoms, as the prodrome of BD. This type of irritability would be a phenomenon that appeared before the hypomanic or manic episodes that defined BD. In opposition, another line of research understood BD in childhood as a rare condition, and chronic irritability needed further investigation. The work of Leibenluft et al. defining the precursor to DMDD [27,29], the syndrome of severe mood dysregulation (SMD),

was imperative in this context. SMD was developed to test the hypothesis whether chronic and severe irritability could be a phenotype of the development of pediatric BD [30]. SMD allowed the differentiation of youth with SMD (chronic, persistent, severe irritability) from youth with classic BD (distinct manic episodes characterized by elated mood). This line of research did not establish a deterministic association between SMD and BD in the future [31]. One of the main outcomes of this work was the introduction of DMDD as a new diagnosis in DSM-5 [27,32].

DMDD is defined by two cardinal symptoms – irritable mood and temper outbursts. Outbursts occur on average three or more times per week and are inconsistent with developmental level. Between outbursts, mood is persistently irritable, most of the day and nearly every day. The onset of symptoms must be before the age of 10, and children must be between 6 and 18 years of age to be diagnosed. Symptoms must have been present for 12 or more months and should not be absent for three or more consecutive months; they should be present in at least two of three settings and be severe in at least one setting. Symptoms are not better explained by another medical disorder, are not the manifestation of substance abuse or medical condition, criteria for manic/hypomanic episode have not been met for more than one day and behaviors do not occur solely during an episode of major depressive disorder. Prior to the age of 6 temper outbursts are considered normal and the boundaries of clinically concerning temper outbursts are being studied [33]. The complete lack of evidence in the adult population prevents researchers to investigate the homotypic continuity of irritability symptoms over development, given DMDD is typically not assessed in adults [34].

The introduction of DMDD as a nosologically category into DSM created a diagnostic home for children with chronic and severe irritability, who are currently not

misdiagnosed with BD. The main issues that still need further investigation are the following; 1- the conceptual foundations of DMMD diagnosis, as well as its reliability and validity; 2- the boundaries that separate irritable mood and temper outbursts normative misbehaviors from irritable mood and temper outbursts clinical indicators; 3 – the absence of a diagnostic classification of severe and chronic irritability in adults.

DSM-5 X ICD-11

DMDD was not included in the ICD-11 by the World Health Organization, leading to a remarkable schism between the 2 major diagnostic manuals. Lochman et al argued that ICD-11 should not incorporate the DMDD category and proposed to make a diagnosis of ODD adding a specifier indicating chronic irritability or anger instead [35]. This argument was bolstered by findings that 92% of children aged 6-12 years in a general population sample who met DMDD criteria also met criteria for ODD, and 66% of those with ODD also had DMDD symptoms [36]. Deveney et al reported that over 80% of their sample of adolescents meeting criteria for severe mood dysregulation also met criteria for ODD or ADHD [29]. However, other data indicate that DMDD and ODD show a more modest overlap (55%) and that DMDD predicts impairment over and above that due to ODD [37].

When considering chronic irritability, the DSM-5 represents irritability in DMDD or ODD, depending on its severity, time course, and number. Unlike DMDD, ODD emerged from a different theoretical foundation- focused on conduct issues. With regard to the ICD-11 stance, the separation and distinctive trajectory of irritability within ODD is a known issue for its construct homogeneity. In the DSM-5, meeting criteria for DMDD is exclusionary for ODD, establishing a formal link between the 2 diagnoses across a line of

severity. We understand the weakness of this thin boundary being a consequence of irritability symptoms overlapping. An important distinction between the 2 diagnoses is that ODD requires at least 1 defiance symptom. We believe that the DMDD diagnosis brings a home to irritability in our manuals, and that specifiers of the same would disentangle overlaps with other and more specific domains of ODD.

Epidemiology

Most of what we know about DMDD has been inferred from research on SMD. The upper age limit placed on its onset in DMDD (10 years) differs from SMD (12 years). A longitudinal study on SMD showed that 97% of youths who met diagnostic criteria for SMD also met criteria for DMDD; the age of onset of 10 or 11 years was the cause of those who did not meet criteria [29]. Besides, a diagnosis of SMD requires symptoms of hyperarousal, whereas DMDD does not. However, clinicians can provide a concomitant diagnosis of ADHD if justified.

A 3-month DMDD prevalence of 8.2% was observed in 6-year-old American children with no sex or ethnicity differences [37]. Copeland et al showed the prevalence of DMDD symptomology in a large epidemiological sample including preschoolers and scholars. Three-month prevalence rates for meeting criteria for DMDD ranged from 0.8% to 3.3%, with the highest rate in preschoolers [38].

Research using the CBCL "dysregulation profile" results in comparable estimates of prevalence: 1-2% in epidemiological samples [39–41], 6-7% in child psychiatric clinical samples, and 13-20% in children with ADHD [42].

The importance of defining a clinical threshold for people with severe irritability

Anger is seen as a normal, natural, and mature emotion experienced by all humans at times that has functional value for survival. On the other hand, when uncontrolled, intense or frequent, or depending on the developmental stage or context it may negatively impair personal and social well-being. As previously mentioned, the co-occurring irritability with several different disorders and the overlapping of diagnostic definitions may lead to confusion amongst clinicians. Furthermore, the thresholds concerning irritability per se are not clearly defined.

Defining a clinical threshold for youth with severe irritability will guide the clinician to better judge whether irritability is developmentally inappropriate or grossly disproportional to the provoking stimulus, recognizing the need for treatments. Very little specific guidance is available to clinicians to evaluate this threshold at any age. Our work provides a clear clinical guidance on how to identify preadolescents, adolescents and adults on the irritability continuum.

Precise, evidence-based knowledge on the course of irritability has several implications for research and clinical practice. These implications affect every other field related to irritability, from diagnosis and classification to neurobiology and clinical management.

Expanding contributions from developmental psychopathology to young adults

Irritability is a trait that has been explored in youth. However, few researches have explored irritable mood and outbursts in adults. Irritability as a core symptom in adults is categorized in the IED and ODD diagnosis in DSM-5The IED category involves extreme temper outbursts, whereas ODD and DMDD involve both disruptive behaviors and irritable mood [43]. IED typically assesses the adults who present outbursts, but not irritable mood. The ODD category has a minor mood component among several other dimensions. The inclusion of DMDD in DSM-5 provided a diagnostic home for youths with irritability, encompassing mood and behavior; however, adulthood was left aside. This has risen a potential bias in the adult literature, which categorized irritability only as a behavior. The diagnosis of major depression is another example in which irritable mood is seen as a core feature in children, but not in adults. Currently, the longitudinal evidence shows that irritability in children predicts generalized anxiety disorder, depression and dysthymia in adult life, but irritability in children predicting irritability in adult life has not been studied yet.

One possibility to solve this lack of integrative perspective in the development is to study the DMDD diagnosis in adults, which may provide an interesting alternative for capturing the mood component in this population and allow the literature to explore a pattern of homotypic continuity (i.e., irritability in children predicting irritability in adults); a pattern which is also common for other emotional disorders such as anxiety and depression.

Expanding the DMDD diagnostic category for adults with chronic and severe irritability will capture the mood component and will enable the derive of empirically clinical threshold for irritability symptoms in this population. Being this a different stage of

development, it is expected that the manifestations of irritability, as well as the thresholds, might differ from youths. Assuming that irritability in adults manifest in different manners, temper outbursts in childhood might manifest as self-harm behaviors in adolescence and as irritable mood in adulthood, for example.

Objectives

General objective

To contribute with the operationalization of DMDD diagnosis in youth

Specific objectives

a. To identify the most appropriate data-driven threshold for DMDD diagnosis in pre-adolescents aged 10-12 from Pelotas Birth Cohort and the impact of potential changes in diagnostic rules on prevalence levels in the community.
b. Expand this data-driven approach for the DMDD diagnosis in adolescents aged 14-17 and young adults aged 18-21 from Brazilian High-Risk Cohort Study for Mental Conditions and the impact of changes in diagnostic rules on prevalence levels in the community.

Hypotheses

Regarding each of the above-mentioned specific objectives, we had the following *a priori* hypotheses:

a. A psychometric data-driven approach may be used to refine the DMDD diagnostic criteria in preadolescents aged 10-12. Yet, external validation might be a challenging process.
b. The thresholds at which irritability becomes pathological may differ between adolescents and young adults; A psychometric data-driven approach might be used to refine the DMDD diagnostic criteria in adolescents aged 14-17 and start the discussion about a diagnostic home for adulthood irritability.

Ethical Considerations

The studies included in this thesis have been approved by their respective Institutional Review Boards before data collection and analysis. Original studies included samples of two cohorts: the 2004 Pelotas Birth Cohorts and the National Institute of Developmental Psychiatry High-Risk Cohort. All participants across all these samples provided written informed consent before inclusion in the study. Data were de-identified, and only raw data essential for analyses were shared with co-authors – therefore, attempts of identification of participants was not possible.

Articles

Article #1

With respect to specific objective a. To identify the most appropriate data-driven threshold for DMDD diagnosis in pre-adolescents aged 10-12 from Pelotas Birth Cohort and the impact of potential changes in diagnostic rules on prevalence levels in the community.

Disruptive Mood Dysregulation Disorder: Symptomatic and Syndromic Thresholds and Diagnostic Operationalization.

Laporte PP, Matijasevich A, Munhoz TN, Santos IS, Barros AJD, Pine DS, Rohde LA, Leibenluft E, Salum GA. J Am Acad Child Adolesc Psychiatry. 2021 Feb;60(2):286-295. **doi: 10.1016/j.jaac.2019.12.008**. Epub 2020 Jan 29. PMID: 32004697

Published in the Journal of the American Academy of Child and Adolescent Psychiatry

Abstract

Objective: The aim of this study is to identify the most appropriate threshold for Disruptive Mood Dysregulation Disorder (DMDD) diagnosis and the impact of potential changes in diagnostic rules on prevalence levels in the community.

Method: Trained psychologists evaluated 3,562 pre-adolescents/early adolescents from the 2004 Pelotas Birth Cohort with the Development and Well-Being Behavior Assessment (DAWBA). The clinical threshold was assessed in three stages: symptomatic, syndromic and clinical operationalization. The symptomatic threshold identified the response category in each DAWBA item which separates normative misbehavior from a clinical indicator. The syndromic threshold identified the number of irritable mood and outbursts needed to capture pre-adolescents/early adolescents with high symptom levels. Clinical operationalization compared the impact of AND/OR rules for combining irritable mood and outbursts on impairment and levels of psychopathology.

Results: At the symptomatic threshold, most irritable mood items were normative in their lowest response categories and clinically significant in their highest response categories. For outbursts some indicated a symptom even when present at only a mild level, while others did not indicate symptoms at any level. At the syndromic level, a combination of 2 out of 7 irritable mood and 3 out of 8 outburst indicators accurately captured a cluster of individuals with high level of symptoms. Analysis combining irritable mood and outbursts delineated non-overlapping aspects of DMDD, providing support for the OR rule in clinical operationalization. The best DMDD criteria resulted in a prevalence of 3%.

Conclusion: Results provide information for initiatives aiming to provide data-driven and clinically oriented operationalized criteria for DMDD.

Key words: developmental psychopathology, disruptive mood dysregulation disorder, irritability, temper outbursts, child/adolescent

Introduction

Temper outbursts and irritable mood are common manifestations of typical development. When outbursts and irritable mood are intense, frequent, last for significant periods, occur in several contexts, and are associated with behaviors not seen in typically developing children, they often require clinical attention [1–3]. Disruptive Mood Dysregulation Disorder (DMDD) is a new diagnosis designed to capture pathological manifestations of irritable mood and temper outbursts [4]. Given the newness of DMDD, data-driven approaches based on epidemiological evidence are needed to evaluate appropriate thresholds for DMDD and consider the need to refine criteria. The current report provides such data.

DMDD has its origins in the mid-2000s when Leibenluft and colleagues [5,6] defined a syndrome called severe mood dysregulation (SMD). SMD involved severe, chronic grouchy mood and heightened reactivity, along with symptoms of hyperarousal [6]. The syndrome was defined to distinguish children with severe irritability from those with classic bipolar disorder (BD), in light of increasing numbers of children diagnosed with BD [7,8]. The results of those studies converged to differentiate SMD from classic bipolar disorder based on course and familial aggregation [9–11]. For DSM-5, SMD was modified to create DMDD.

Alternative thresholds for defining DMDD have been only partially considered in the current literature. Some previous studies have focused on irritability as a dimensional trait, which is broader than DMDD as a diagnostic entity. These studies provide an important framework for investigating clinically-relevant thresholds for specific behaviors. Wakschlag and collaborators [12] used item response theory analysis to disentangle normative misbehavior from clinically significant problems by studying the 'symptomatic threshold', i.e., investigating which response category in each item from a questionnaire separates normative misbehavior from a clinical indicator. They found that some behaviors are normative and

only represent problems when their frequency is high or very high, whereas others always indicate a significant problem that requires clinical attention. This and similar research efforts in preschoolers [13] inform attempts to evaluate varying boundaries for the definition of DMDD. Other studies focused more specifically on varying DSM-5 criteria for DMDD in pre-adolescents [14,15] and adolescents [14–16]. They found the prevalence of temper outbursts and negative mood are much lower than what is found in preschoolers and that applying exclusion criteria such as frequency and hierarchical diagnostic rules affects DMDD prevalence rates considerably [14–16]. There was no evidence that clinical markers changed between pre-adolescents/early adolescents (9-12) to middle adolescents (13-16) [15]. Nonetheless, it is important to continue to identify appropriate diagnostic thresholds for distinct developmental periods, given that normative levels of irritability clearly vary across the lifespan [14–16].

Another important step towards evaluating such varying boundaries involves quantifying the number of abnormal behaviors required to characterize a valid diagnosis i.e., identifying the 'syndromic threshold' for a given diagnosis. Data-driven clustering approaches such as latent class analysis derive groups that differ in the number of clinical indicators endorsed [17] and thus inform attempts to set syndromic thresholds. Such efforts need to be balanced with clinical applicability in real world settings, which require practical decisions such as how to combine clinical indicators from distinct domains (i.e., irritable mood and outbursts). The latter can be achieved by investigating whether domains explain overlapping or distinct aspects of DMDD latent structure and related impairment, thus determining whether “and” or “or” rules should be used to provide a ‘clinical operationalization’ of the diagnosis. Previous research in pre-adolescents and adolescents suggests irritable mood and temper outbursts predict each other over time. However, while

each of them are associated with increased risk for disrupted functioning in adolescents [15], current criteria require both to be present for a diagnosis to be assigned.

The aim of this study is to evaluate alternative clinical thresholds for the DMDD diagnosis (see Figure 1 for an overview of the analytic strategy and Methods for details). We investigate 3,562 pre-adolescents/early adolescents aged 10-12. First, we used Confirmatory Factor Analysis (CFA) to identify item-level thresholds differentiating normative from clinical problems (the symptomatic threshold). This was used to dichotomize response levels as clinically significant or not. We next used these binary clinical indicators as input to a latent class analysis (LCA) that assigned individuals into clusters with high and low levels of clinical indicators for each domain. This was followed by receiver operating curves (ROC) to detect the number of clinical indicators needed to predict class membership from the LCA and to translate the data-driven results to DSM-5 symptom counts (the syndromic threshold). We then compare the impact of AND/OR rules on impairment and dimensional levels of psychopathology (clinical operationalization). Finally, we investigate the impact of varying definitions on DMDD prevalence and comorbidity profiles in a population-based sample.

Figure 1 around here

Method

Participants

Participants of this study were pre-adolescents/early adolescents aged 10-12 from the 2004 Pelotas Birth Cohort Study. All births occurring in the city of Pelotas, from January 1st to December 31st, 2004 were enrolled and followed over time. Pelotas is in southern Brazil and has a population of 328,000. For a full description of the methods, see [18]. Briefly, all

4231 live births in the city in 2004 whose mothers lived in the urban area and agreed to participate in the longitudinal study were considered eligible. Follow-up home visits were performed when the subjects had reached the ages of 3.0 months (SD=0.1), 11.9 months (SD=0.2), 23.9 months (SD=0.4), and 49.5 months (SD=1.7). When the subjects were, on average, 6.8 years old (SD=0.3) and 11.0 years old (SD=0.4), additional follow-up visits were conducted at a research clinic run by the Postgraduate Program of Epidemiology (Faculty of Medicine, Federal University of Pelotas, Brazil). Of the 4,231 subjects in the original birth cohort, 3,562 (84.1%) were included in our analysis, which used all available data from the 10-12 years of age assessment. The sample comprises 2,353 participants aged 10, 1,206 aged 11 and 4 aged 12. The prevalence of DMDD in this sample using current criteria associated with clinical ratings was 2.5% (95% CI=2.0–3.0) [19]. The study was approved by the Research Ethics Committee of the Federal University of Pelotas and by the Research Committee of the University of São Paulo School of Medicine. Written informed consent was obtained from all subjects.

Instruments and Diagnostic Assessment

The parent-version of the DMDD section from the Development and Well-Being Assessment (DAWBA) questionnaire [20] was administered by certified psychologists. This questionnaire uses open and closed ended questions to identify the occurrence of clinical indicators in children and adolescents aged 5-17, based on the DSM criteria. The closed ended questions start with two skip questions about the frequency of temper outbursts and irritable mood. Parents who answered that temper outbursts and/or irritable mood occurred at least once a week were probed to answer specific questions that characterize all DSM-5 criteria for DMDD.

A total of 593 parents of participants answered the DMDD section on irritable mood, representing the top 17% of irritable mood frequency. This section includes 9 items characterizing the threshold for experiencing anger, intensity of anger if compared to peers of the same age, duration of anger during the day, whether irritable mood is perceived by others, setting in which anger occurs (at home, at school, with peers) and number of anger weeks throughout the year. A total of 425 parents answered the DMDD section on outbursts, representing the top 12% of frequency of outbursts. This section includes 15 items describing behavior during outbursts (slamming doors, shouting, swearing, saying mean things to others, saying negative things about self, physical aggression to others, deliberate self-harm, breaking things), setting in which outbursts occurred (at home, at school, with peers) and triggers (recognizable and easily triggered). We do not use the item "outbursts free-gap in the last year" in our analysis (DSM requires that there is not a period higher than 3 or more consecutive months without irritable mood and temper outbursts). The rationale for excluding this item is that it is unclear whether we would expect this item to be monotonically related to the overall latent construct given short periods of irritability with large gaps could also inform episodes of irritability (a marker of severity and bipolar disorder).

Lastly, 686 mothers or caregivers that completed either the outburst or irritable mood sections were asked to also complete 4 items about impairment (impact on family life, friendship, learning, and leisure activities). After the impairment questions, mothers or caregivers answered the open-ended questions that allow qualitative description of the symptoms, frequency, and other characteristics of the disorder. All questions and response categories from the DMDD section are depicted in Table S1, available online.

The DAWBA was administered to mothers or caregivers by trained psychologists. The forty-hour training included lectures, role playing, and supervised clinical interviews

with pediatric and mental health outpatients at the Federal University of Pelotas. The clinical evaluation of the total sample was performed by a psychologist, and a second independent psychologist evaluated 10% of the study sample. Both were trained in how to apply the DAWBA, in a standardized manner, by the child psychiatrist who had translated and validated the questionnaire for use in Brazil [21]. Rating procedures were used for assigning comorbidities given DMDD diagnosis was performed *a posteriori*. The inter-rater agreement was 91.2% for the presence of any psychiatric disorder, 75.9% for any anxiety disorder, 73.5% for any depressive disorder, 72.7% for ADHD, 72.9% for conduct disorder, 85.6% for any autism spectrum disorder, 59.5% for any eating disorder, and 52.4% for any tic disorder. Details of the questionnaire can be found online and in other studies [22].

The Strengths and Difficulties Questionnaire (SDQ) was used to measure dimensional psychopathology. The SDQ is a 25-item behavioral screening questionnaire with five domains, each of which contain five items (emotional, conduct, hyperkinetic, peer relationships, prosocial behaviors and impact scores). The overall SDQ total scores had a Cronbach's alpha of 0.82, which is considered high. Internal consistency for the SDQ subscales were low to moderate ranging from 0.48 (peer relationships) to 0.78 (hyperkinetic). Despite low reliability, we maintained results from subscales for their descriptive nature in Supplement 1, available online.

Statistical Analysis

Symptomatic threshold

The fourteen items on outbursts, eight items on irritable mood and four items on impairment were included in three CFAs testing unidimensional models for each construct (n= 593, 425 and 685 respectively). Details about the estimators and the model fit used in

this study can be found in Table S2, available online, which provides the fit indexes of the unidimensional models for irritability, outbursts and impairment items. CFA models estimate item level factor loadings (λ) and response category thresholds. Factor loadings represent the strength of the relationship between the latent trait and the item, i.e., they indicate how well each item discriminates different severity levels of a given construct. Category thresholds indicate the expected value of the latent factor at which there is a 50% probability of endorsing a given category or higher i.e., the category threshold indicates the severity level at which the transition from one response category to the next is likely to happen (e.g., from 'No' to 'A little' or higher, or from 'A little' to 'A lot').

To distinguish normative misbehavior from behavior that would meet a diagnostic criterion, we used category thresholds from the CFA. CFAs were performed only in subjects with a frequency of irritable mood and outbursts greater than once a week. In this sample, a value of 0.5 represents a half standard deviation above the mean of the distribution of subjects with a frequency of irritable mood and outbursts greater than once a week. Therefore, we interpreted values below 0.5 as typical development (normative) and values at or above 0.5 as 'clinical indicators' (a proxy for symptoms or problem indicators). The latter represents an approximation to the top 5% most symptomatic pre-adolescents/early adolescents in the population, which is a threshold used in other diagnostic investigations [12]. For details about the CFA, see Supplement 1, available online.

Syndromic threshold

Before data analysis, each questionnaire item was dichotomized at the value of the category threshold defined in the symptomatic threshold analysis described above in subjects with at least one clinical indicator. Dichotomized items were chosen to enter the LCA

analysis because our intention was not to characterize varying levels of irritability in the community, but to identify groups that differ in their number of clinical indicators. Three Latent Class Analyses (LCA) were used to create empirically derived groups with different levels of clinical indicators for irritable mood, outbursts and impairment. Next, we used three Receiving Operating Curves to predict the most accurate number of clinical indicators for detecting participants with high levels of symptoms (as defined by latent class analysis) with regard to irritable mood, outburst and impairment. ROC analysis was used as a way to translate results from the syndromic thresholds of the LCA to the reality of clinical practice, which uses symptom counts. Thus, the ROC identifies a simple rule to allow the identification of patients that are likely to be members of the cluster that exhibit a high level of clinical indicators. The optimal cut-off was estimated using the Younden's J Statistic, which maximizes both sensitivity and specificity [23].

Clinical operationalization

Four analytical strategies were used to determine the most appropriate rule for clinical operationalization: the 'OR' rule vs. the 'AND' rule. First, we compared the fit of CFA models (n=398), putting the selected dichotomized clinical indicators into a unidimensional model of irritability and a correlated model of irritability with two domains (irritable mood and outbursts). Second, we tested whether meeting criteria for the irritable mood group and/or for the outbursts group have distinct or overlapping associations with the impaired functioning group using a multiple logistic regression. Third, we used left censored regressions to compare skewed SDQ dimensional scores between subjects meeting criteria only for irritable mood, only outbursts, either, or both and compared with a group of participants with other DSM disorders except for DMDD and typically developing

comparisons. Fourth, for both OR and AND groups, we used a matching procedure to compare levels of SDQ scores between a group that differed in DMDD status (yes vs. no DMDD) but were otherwise fully matched for comorbidities.

Epidemiological impact

Finally, using the relative frequency, we investigated the impact of these AND/OR rules and combinations for impairment requirements on the prevalence rates of DMDD in the community and on the comorbidity profile.

All analysis were performed in R version 3.6.1 [24], including applications implemented in the packages lavaan 0.6-5 [25], poLCA 1.4.1 [26], pROC 1.15.3 [27], CensReg 0.5-26 [28] and MatchIt 3.0.2 [29]. The R markdown codes for the symptomatic, syndromic and clinical operationalization thresholds of current analysis can be found in Supplement 1, available online.

Results

Symptomatic threshold

All eight items of irritable mood were found to be normative in their lowest thresholds and clinical indicators (proxy for symptoms) in the highest thresholds, except "irritable mood that happens at home", which was found to be normative in all response categories. For the six items that describe intensity, the response option "A little" indicated normative behavior, while the response option "A lot" or "A great deal" indicated a symptom. For the duration item, irritable mood lasting less than an hour indicated normative behavior, whereas irritability lasting a few hours or most of the day indicated a symptom. For the frequency

item, irritable mood occurring less than 3 times a week indicated normative behavior, whereas irritable mood occurring every day indicated a symptom (Table S3, available online).

For outbursts, threshold varied substantially across items. For some items, their occurrence even at mild levels indicated a symptom, while other items did not indicate a symptom at any level. Outbursts that include self-harm, breaking things or saying negative things about self, or those that occurred in the classroom are indicative of a symptom if they occur at any level (i.e., "A little" or "A lot"). Outbursts that occurred with peers, include physical aggression, or are easily triggered indicate a symptom when they occurred "A lot", but were normative when they occurred "A little". Outbursts that occurred at home and included the pre-adolescents/early adolescents saying mean things, slamming doors, shouting, or swearing did not indicate a symptom irrespective of the level endorsed. Also, whether the triggers were recognizable or not was not relevant to symptom designation. Regarding frequency, only outbursts that occurred daily indicated a symptom (Table S4, available online).

For impairment, "impact on family life" is normative when "A little" and indicate a symptom when "A medium amount" or "A lot". Impairment that occurs in the other settings (friendship, learning or leisure) indicate a symptom at any level (Table S5, available online). See Figure 2.

Figure 2 around here

Syndromic threshold

The prevalence of each clinical indicator is presented in Table 1. The three LCA analyses (irritable mood, outbursts and impairment) indicated that the two-class solution was the best for each of the three domains (Table S6, available online). This indicates that, in each of the three domains, the population is divided into two groups characterized by high vs. low symptoms (Figure S1, available online). We next performed three Receiver Operating Curve analyses (irritable mood, outburst, impairment) to determine the best number of clinical indicators (i.e., those items identified by the CFA) to use to predict membership in the high vs. low symptom classes identified by the LCA. Younden's J demonstrates that subjects in the high symptom irritable mood and outburst classes are most accurately characterized by 2 out of 7 irritable mood symptoms and 3 out of 8 outburst symptoms. As for impairment, the subjects with high level of symptoms in LCA are most accurately characterized by significant impairment in at least two settings (Table 2 and Figure S2, available online).

Table 1 around here

Table 2 around here

Clinical operationalization

First, a model with two correlated domains (irritable mood and outbursts) provided a better fit than a unidimensional model encompassing both domains (χ^2_{diff}=7.3, df=1, p=0.007; Table S7, available online). Second, both irritable mood and outbursts were associated with clinical impairment in univariate models (irritable mood OR=41.71, p<0.001; outbursts OR=76.1, p<0.001) and in multiple models adjusted for the effects of including both predictors in the same model (irritable mood adjusted OR=18.2, p<0.001; outbursts

adjusted OR=23.63, p<0.001). Third, comparisons between irritable only, outbursts only and combined groups with typically developing comparisons and with a group of patients with other DSM disorders (except for DMDD) showed all three DMDD groups had higher scores on all SDQ scales than typically developing comparisons and higher total SDQ total scores than subjects with other DSM diagnosis (Figure S3 and Table S8, both available online). Fourth, left-censored regressions comparing groups matched for comorbidity (any anxiety, any mood, any hyperkinetic and any disruptive behavior disorder) showed that, using either the OR or the AND rule, the DMDD group showed higher total, emotional, conduct, hyperactivity, peer relationship and impact scores than did the non-DMDD group with matched comorbidities (Figure S4 and Figure S5, available online).

Epidemiological impact

When using an "OR" rule, the optimal criteria from the ROC analysis (2 of 7 irritable mood symptoms, 3 of 8 outburst symptoms, impairment in at least two settings) resulted in a prevalence of 3.0%: 1.12% have only irritable mood, 0.64% only outbursts and 1.23% have both irritable mood and outbursts (Table 3). Both the "OR rule" and the "AND rule" resulted in higher levels of psychiatric comorbidities compared to the current DMDD clinical criteria (Table S9, available online).

Table 3 around here

Discussion

This study provides important information to guide a revision of the diagnostic criteria for DMDD. Using CFA, we found that seven of the eight irritable mood items were normative when endorsed in the low response categories and clinical indicators in the high

response categories. The one exception was “irritable mood that happens at home”, which was always normative. For outbursts, the threshold for a clinical indicator varied substantially across items. For some items, such as outbursts with self-harm, their presence indicated a problem even at only mild levels. Others, such as shouting, were not clinical indicators even when present at the highest threshold. ROC analyses indicated that a combination of 2 of 7 irritable mood symptoms, 3 of 8 outburst symptoms and significant impairment in at least two settings would best predict membership in the “high” vs. “low” LCA-based symptom classes. The four clinical operationalization analysis converge to demonstrate that the two domains differ from a latent perspective; they are independently associated with impairment; and OR-rule groups show comparable or even higher levels of impairment then other DSM disorders. Matched analysis showed that results cannot be attributed to comorbidity. The most accurate solution resulted in a prevalence of 3% in the fully automated operationalized criteria (1.12% only irritable mood, 0.64% only outbursts and 1.23% combined).

Our findings are consistent with the limited literature examining irritability dimensionally in the population. Each set of findings suggest that normative outbursts differ from clinical indicators in frequency, duration, quality, context, and triggering events [30–33]. Wakschlag et al. [12] found that outbursts characterized by high frequency, “long duration”, or “aggressive components”, or those that occurred “with nonparental adults” or “out of the blue” were clinical indicators. Wiggins and collaborators [13] also used an empirical approach to identify irritable behaviors indicative of problems in preschoolers. They examined 22 temper loss behaviors from the criteria for oppositional defiant disorder, DMDD and other depressive disorders in the DSM-5 and found two informative items. Similar to our work, the item “easily frustrated” indicated a symptom only when present nearly every day, but the item “break/destroy” indicated a symptom even when at lower frequencies. Nevertheless,

those thresholds might vary substantially in distinct age ranges and cultures, which highlights the need for developmentally sensitive studies.

Clinical operationalization analysis suggests that an OR rule is most appropriate to capture cases in need of treatment. This algorithm identified pre-adolescents/early adolescents with either irritable mood or outbursts who manifested associated impairment, elevated symptoms, and functional impairment. This resulted in a prevalence rate of 3%, which is higher than the prevalence rate of 2.5% by the current diagnostic criteria. Of course, it is not possible to identify the "true" prevalence of DMDD in the population with one study; rather the current analyses inform nosologists' attempts to weigh the strengths and weaknesses of various diagnostic thresholds.

Advancing understanding about DMDD diagnostic criteria is a major concern in children and adolescent psychiatric practice. Our findings are a first step towards defining parameters to alert the clinician when to be (and when not to be) concerned with irritable mood and outbursts. Our approach suggests several refinements to the DSM-5 criteria. First, the new criteria provide a list of behaviors and a threshold for each that specifies when to consider that behavior to be a clinical indicator. This is more descriptive, precise, and data-based than the current criteria and provide a way 'calibrate' the severity of each clinical indicator composing the syndrome. Second, we suggest a syndromic threshold for the combination of such behaviors. This is a more practical way to separate normal from abnormal behaviors and considers that DMDD might present itself with distinct clinical indicators. Third, our data support an OR rule when combining irritable mood and outbursts, rather than the AND rule currently in the manual. Finally, our results support the importance of requiring two settings for the diagnosis, as in DSM-5. Specifically, our data indicate that, while the impact of symptoms on function needs to be at medium levels on family life to be

considered a clinical indicator, mild levels of impairment in friendship, learning or during leisure activities should suffice as a clinical indicator for the DMDD impairment criteria.

Our study has important strengths. *First*, we relied on a large representative population sample and implemented assessment methods that could mimic clinical assessment in the real world, as far as possible in an epidemiological investigation. *Second*, we applied Confirmatory Factor Analysis, Latent Class Analysis and Receiver Operating Curves Analysis, applying a similar framework used in other disorders [36,37] to a new syndrome that lacks empirical investigations to guide operationalization. However, this work has also some important limitations. First, our analysis is focused on internal validators. Further studies investigating course, family history, treatment response, and other external validators are needed to demonstrate the validity of the operationalized syndrome. Since associations between symptoms and irritability-related impairment were investigated using the same DAWBA DMDD section, the size of the associations is likely to be overestimated. However, the value of these odds ratios may be helpful in understanding whether the two aspects of irritability capture distinct or overlapping aspects of irritability-related impairment. Second, our subjects were all 10-12 years old, and our data might not be generalized to other developmental stages. Third, because of the skip rule questions, the CFA parameters were estimated for subjects with irritable mood or outbursts that occurred at least once a week. Analysis were modeled to consider these characteristics, but this might have biased the parameter estimates for some items. Also, our approach assumes irritable mood, outbursts and impairment are distinct domains, which is still an empirical question to be further tested. Fourth, our analysis is restricted to parent reports, and no information was acquired from pre-adolescents/early adolescents themselves. Lastly, our approach is

restricted to a single sample and it is unclear whether those results can be replicated in other samples.

To conclude, this is the first study in the field with this intent in this age range and thus is a first step towards refining the diagnostic criteria of DMDD. Future research to advance the field of DMDD should include replicating these findings; extending similar approaches to diagnostic instruments other than the DAWBA; examining symptomatic thresholds using measures that do not have skipping rules and are designed specifically to differentiate normative versus non-normative behaviors; investigating interrater reliability; and including developmentally sensitive items and external validators. Furthermore, prospective longitudinal investigation that applies this framework beginning at earlier ages can elucidate the origins of pathologic irritability, thus guiding the development of novel interventions and developmentally-based prevention.

References

1. Wakschlag LS, Estabrook R, Petitclerc A, et al. Clinical Implications of a Dimensional Approach: The Normal:Abnormal Spectrum of Early Irritability. *J Am Acad Child Adolesc Psychiatry*. 2015;54(8):626-634. doi:10.1016/j.jaac.2015.05.016
2. Petitclerc A, Briggs-Gowan MJ, Estabrook R, et al. Contextual variation in young children's observed disruptive behavior on the DB-DOS: implications for early identification. *J Child Psychol Psychiatry*. 2015;56(9):1008-1016. doi:10.1111/jcpp.12430
3. Perlman SB, Jones BM, Wakschlag LS, Axelson D, Birmaher B, Phillips ML. Neural substrates of child irritability in typically developing and psychiatric populations. *Dev Cogn Neurosci*. 2015;14:71-80. doi:10.1016/j.dcn.2015.07.003
4. DSM-5. https://www.psychiatry.org/psychiatrists/practice/dsm. Accessed July 1, 2018.
5. Leibenluft E. Severe mood dysregulation, irritability, and the diagnostic boundaries of bipolar disorder in youths. *Am J Psychiatry*. 2011;168(2):129-142. doi:10.1176/appi.ajp.2010.10050766
6. Leibenluft E, Blair RJR, Charney DS, Pine DS. Irritability in pediatric mania and other childhood psychopathology. *Ann N Y Acad Sci*. 2003;1008:201-218.
7. Blader JC, Carlson GA. Increased rates of bipolar disorder diagnoses among U.S. child, adolescent, and adult inpatients, 1996-2004. *Biol Psychiatry*. 2007;62(2):107-114. doi:10.1016/j.biopsych.2006.11.006
8. Olfson M, Blanco C, Liu L, Moreno C, Laje G. National trends in the outpatient treatment of children and adolescents with antipsychotic drugs. *Arch Gen Psychiatry*. 2006;63(6):679-685. doi:10.1001/archpsyc.63.6.679
9. Stringaris A, Cohen P, Pine DS, Leibenluft E. Adult outcomes of youth irritability: a 20-year prospective community-based study. *Am J Psychiatry*. 2009;166(9):1048-1054. doi:10.1176/appi.ajp.2009.08121849
10. Leibenluft E, Cohen P, Gorrindo T, Brook JS, Pine DS. Chronic versus episodic irritability in youth: a community-based, longitudinal study of clinical and diagnostic associations. *J Child Adolesc Psychopharmacol*. 2006;16(4):456-466. doi:10.1089/cap.2006.16.456
11. Stringaris A, Baroni A, Haimm C, et al. Pediatric bipolar disorder versus severe

mood dysregulation: risk for manic episodes on follow-up. *J Am Acad Child Adolesc Psychiatry*. 2010;49(4):397-405.

12. Wakschlag LS, Choi SW, Carter AS, et al. Defining the developmental parameters of temper loss in early childhood: implications for developmental psychopathology. *J Child Psychol Psychiatry*. 2012;53(11):1099-1108. doi:10.1111/j.1469-7610.2012.02595.x

13. Wiggins JL, Briggs-Gowan MJ, Estabrook R, et al. Identifying Clinically Significant Irritability in Early Childhood. *J Am Acad Child Adolesc Psychiatry*. 2018;57(3):191-199.e2. doi:10.1016/j.jaac.2017.12.008

14. Copeland WE, Angold A, Costello EJ, Egger H. Prevalence, comorbidity, and correlates of DSM-5 proposed disruptive mood dysregulation disorder. *Am J Psychiatry*. 2013;170(2):173-179. doi:10.1176/appi.ajp.2012.12010132

15. Copeland WE, Brotman MA, Costello EJ. Normative Irritability in Youth: Developmental Findings From the Great Smoky Mountains Study. *J Am Acad Child Adolesc Psychiatry*. 2015;54(8):635-642. doi:10.1016/j.jaac.2015.05.008

16. Althoff RR, Crehan ET, He J-P, Burstein M, Hudziak JJ, Merikangas KR. Disruptive Mood Dysregulation Disorder at Ages 13-18: Results from the National Comorbidity Survey-Adolescent Supplement. *J Child Adolesc Psychopharmacol*. 2016;26(2):107-113. doi:10.1089/cap.2015.0038

17. Borsboom D, Rhemtulla M, Cramer AOJ, van der Maas HLJ, Scheffer M, Dolan CV. Kinds versus continua: a review of psychometric approaches to uncover the structure of psychiatric constructs. *Psychol Med*. 2016;46(8):1567-1579. doi:10.1017/S0033291715001944

18. Santos IS, Barros AJD, Matijasevich A, et al. Cohort profile update: 2004 Pelotas (Brazil) Birth Cohort Study. Body composition, mental health and genetic assessment at the 6 years follow-up. *Int J Epidemiol*. 2014;43(5):1437-1437a-f. doi:10.1093/ije/dyu144

19. Munhoz TN, Santos IS, Barros AJD, Anselmi L, Barros FC, Matijasevich A. Perinatal and postnatal risk factors for disruptive mood dysregulation disorder at age 11: 2004 Pelotas Birth Cohort Study. *J Affect Disord*. 2017;215:263-268. doi:10.1016/j.jad.2017.03.040

20. Goodman R, Ford T, Richards H, Gatward R, Meltzer H. The Development and Well-Being Assessment: description and initial validation of an integrated assessment of

child and adolescent psychopathology. *J Child Psychol Psychiatry*. 2000;41(5):645-655.

21. Fleitlich-Bilyk B, Goodman R. Prevalence of child and adolescent psychiatric disorders in southeast Brazil. *J Am Acad Child Adolesc Psychiatry*. 2004;43(6):727-734. doi:10.1097/01.chi.0000120021.14101.ca

22. La Maison C, Munhoz TN, Santos IS, Anselmi L, Barros FC, Matijasevich A. Prevalence and risk factors of psychiatric disorders in early adolescence: 2004 Pelotas (Brazil) birth cohort. *Soc Psychiatry Psychiatr Epidemiol*. 2018;53(7):685-697. doi:10.1007/s00127-018-1516-z

23. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3(1):32-35.

24. R: The R Project for Statistical Computing. https://www.r-project.org/. Accessed August 26, 2018.

25. Rosseel Y, Oberski D, Byrnes J, et al. *Lavaan: Latent Variable Analysis.*; 2018. https://CRAN.R-project.org/package=lavaan. Accessed August 26, 2018.

26. poLCA.pdf. https://cran.r-project.org/web/packages/poLCA/poLCA.pdf. Accessed September 19, 2019.

27. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12:77. doi:10.1186/1471-2105-12-77

28. CRAN - Package censReg. https://cran.r-project.org/web/packages/censReg/index.html. Accessed May 27, 2019.

29. Ho D, Imai K, King G, Stuart E, Whitworth A. *MatchIt: Nonparametric Preprocessing for Parametric Causal Inference*.; 2018. https://CRAN.R-project.org/package=MatchIt. Accessed September 19, 2019.

30. Belden AC, Thomson NR, Luby JL. Temper tantrums in healthy versus depressed and disruptive preschoolers: defining tantrum behaviors associated with clinical problems. *J Pediatr*. 2008;152(1):117-122. doi:10.1016/j.jpeds.2007.06.030

31. Bhatia MS, Dhar NK, Singhal PK, Nigam VR, Malik SC, Mullick DN. Temper tantrums. Prevalence and etiology in a non-referral outpatient setting. *Clin Pediatr (Phila)*. 1990;29(6):311-315. doi:10.1177/000992289002900603

32. Osterman K, Björkqvist K. A cross-sectional study of onset, cessation, frequency, and duration of children's temper tantrums in a nonclinical sample. *Psychol Rep*.

2010;106(2):448-454. doi:10.2466/PR0.106.2.448-454

33. Wakschlag LS, Briggs-Gowan MJ, Carter AS, et al. A developmental framework for distinguishing disruptive behavior from normative misbehavior in preschool children. *J Child Psychol Psychiatry*. 2007;48(10):976-987. doi:10.1111/j.1469-7610.2007.01786.x

34. Hawes MT, Carlson GA, Finsaas MC, Olino TM, Seely JR, Klein DN. Dimensions of irritability in adolescents: longitudinal associations with psychopathology in adulthood. *Psychol Med*. October 2019:1-9. doi:10.1017/S0033291719002903

35. Leibenluft E, Cohen P, Gorrindo T, Brook JS, Pine DS. Chronic versus episodic irritability in youth: a community-based, longitudinal study of clinical and diagnostic associations. *J Child Adolesc Psychopharmacol*. 2006;16(4):456-466. doi:10.1089/cap.2006.16.456

36. Kessler RC, Green JG, Adler LA, et al. Structure and diagnosis of adult attention-deficit/hyperactivity disorder: analysis of expanded symptom criteria from the Adult ADHD Clinical Diagnostic Scale. *Arch Gen Psychiatry*. 2010;67(11):1168-1178. doi:10.1001/archgenpsychiatry.2010.146

37. Matte B, Anselmi L, Salum GA, et al. ADHD in DSM-5: a field trial in a large, representative sample of 18- to 19-year-old adults. *Psychol Med*. 2015;45(2):361-373. doi:10.1017/S0033291714001470

Table 1- Prevalence[a] of Each Disruptive Mood Dysregulation Disorder Item Written in Combination With the Response Category That Defines a Clinical Indicator

	Irritable Mood	Prevalence Estimation of Clinical Indicators (%)
	Frequency/duration	
1	Irritable mood ***occurring every day***	2.1
2	Irritable mood that ***lasts more than a few hours***	2.7
	Characteristics	
3	Easily irritated, annoyed or angry ***a lot***	3.6
4	Intense irritable mood ***a lot***	2.8
	Settings	
5	Irritable mood occurs in the classroom ***a lot***	1.1
6	Irritable mood occurs with peers ***a lot***	0.9
7	Irritable mood is evident to others ***a great deal***	1.7
	Temper Outbursts	Prevalence Estimation of Clinical Indicators (%)
	Frequency/duration	
1	Outbursts ***occurring every day***	1.5
	Characteristics of the outbursts	
2	Saying ***any*** negative thing about self	3.1
3	***Any*** physical aggression to others	1.7
4	***Any*** form of deliberate self-harm	1.3
5	Breaking things (***any***)	3.4

	Settings	
6	***Any*** outburst in the classroom	3.3
7	Outbursts occurs with peers ***a lot***	0.7
	Triggers	
8	Easily triggered ***a lot***	2.6

Note: [a] Prevalence estimates assume that pre-adolescents/early adolescents whose irritable mood and outbursts occurred less than once per week (and who therefore did not complete these items) do not have any of these problems to a significant degree.

Table 2. Receiver Operating Curves Parameters Investigating the Best Number of Clinical Indicators to Capture Latent Class Groups

Threshold	Prediction of Latent Class Groups																	
	Irritable Mood						Outbursts						Severity of Impairment					
	ACC	Sens	Spe	PPV	NPV	YI	ACC	Sens	Spe	PPV	NPV	YI	ACC	Sens	Spe	PPV	NPV	YI
0	0.25	1.00	0.00	0.25	-	0	0.12	1.00	0.00	0.12	-	0	0.36	1.00	0.00	0.12	-	0
1	0.56	1.00	0.41	0.36	1.00	0.4	0.40	1.00	0.32	0.17	1.00	0.32	0.73	1.00	0.58	0.57	1.00	0.58
2	**0.86**	**0.95**	**0.83**	**0.65**	**0.98**	**0.78**	0.69	1.00	0.64	0.28	1.00	0.64	**0.93**	**1.00**	**0.89**	**0.83**	**1.00**	**0.89**
3	0.90	0.64	0.99	0.95	0.89	0.63	**0.90**	**1.00**	**0.88**	**0.54**	**1.00**	**0.88**	0.90	0.72	1.00	1.00	0.87	0.87
4	0.83	0.32	1.00	1.00	0.82	0.32	0.98	0.86	0.99	0.98	0.98	0.85	0.74	0.28	1.00	1.00	0.71	0.28
5	0.78	0.11	1.00	1.00	0.77	0.11	0.93	0.43	1.00	1.00	0.92	0.43						
6	0.76	0.04	1.00	1.00	0.76	0.04	0.89	0.14	1.00	1.00	0.89	0.14						
7	0.75	0.00	1.00	-	0.75	0	0.89	0.12	1.00	1.00	0.89	0.12						
8							0.88	0.02	1.00	1.00	0.88	0.02						

Note: ACC = Accuracy; NPV = Negative Predictive Value; PPV = Positive Predictive Value; Sens = Sensitivity; Spe = Specificity; YI = Youden's Index.

Table 3 – Impact of Different Rules for Combining Irritable Mood and Temper Outburst Clinical Indicators and Impairment Requirements on Prevalence Rates of Disruptive Mood Dysregulation Disorder (DMDD)

	Irritable Mood	**Outbursts**	**AND Rule**	**OR Rule**
No impairment requirement	2.41	0.9	1.63	4.94
At least one setting	1.80	0.70	1.46	3.96
At least two settings (optimal)	**1.12**	**0.64**	**1.23**	**3.00**
At least three settings	0.61	0.48	0.87	1.96
All four settings	0.17	0.17	0.53	0.87

Note: Settings: 1= Impact on family life; 2 = impact on friendship; 3 = Impact on learning; 4 = Impact on leisure.

Figure 1 – Symptomatic and Syndromic Thresholds and Clinical Operationalizations

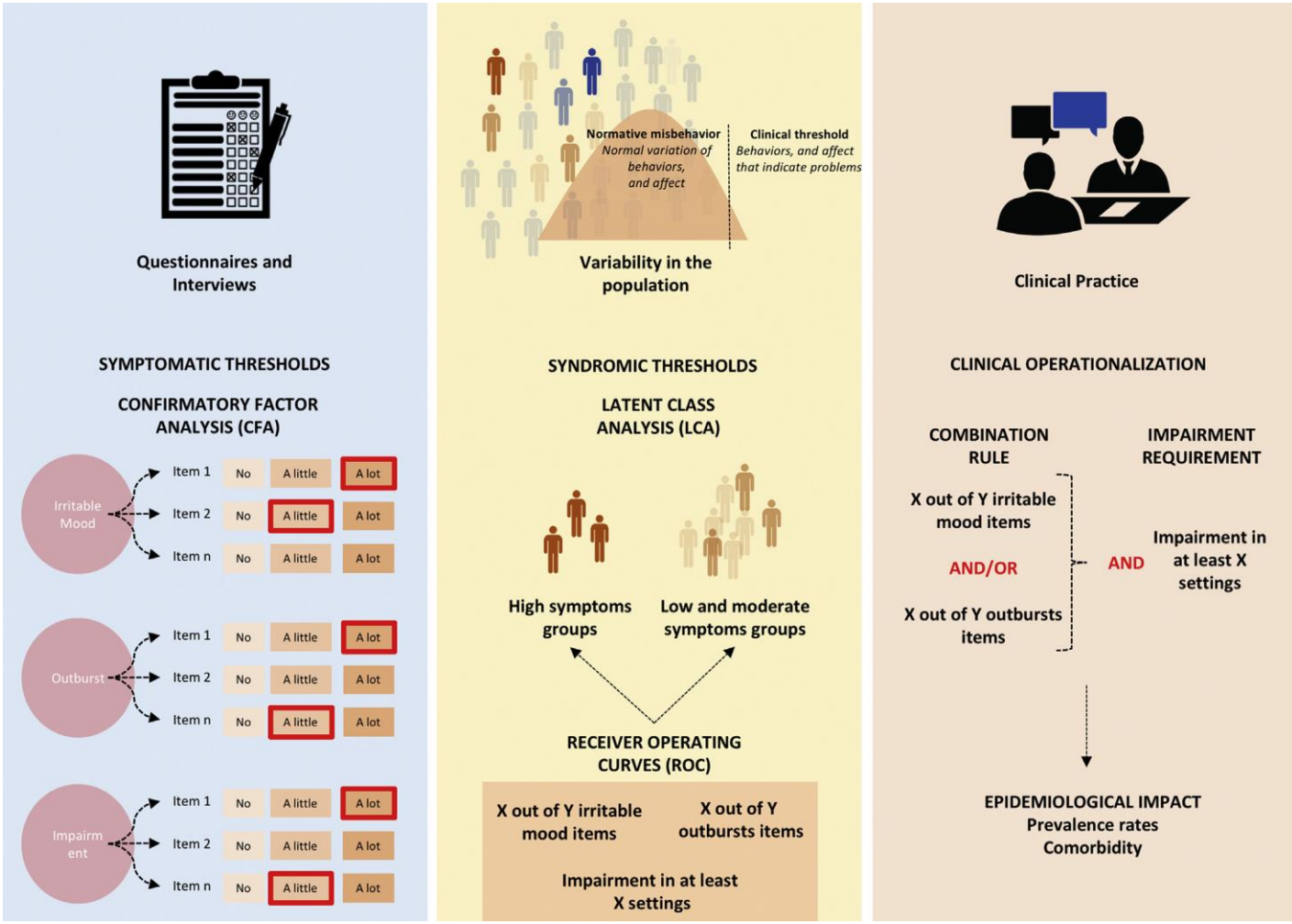
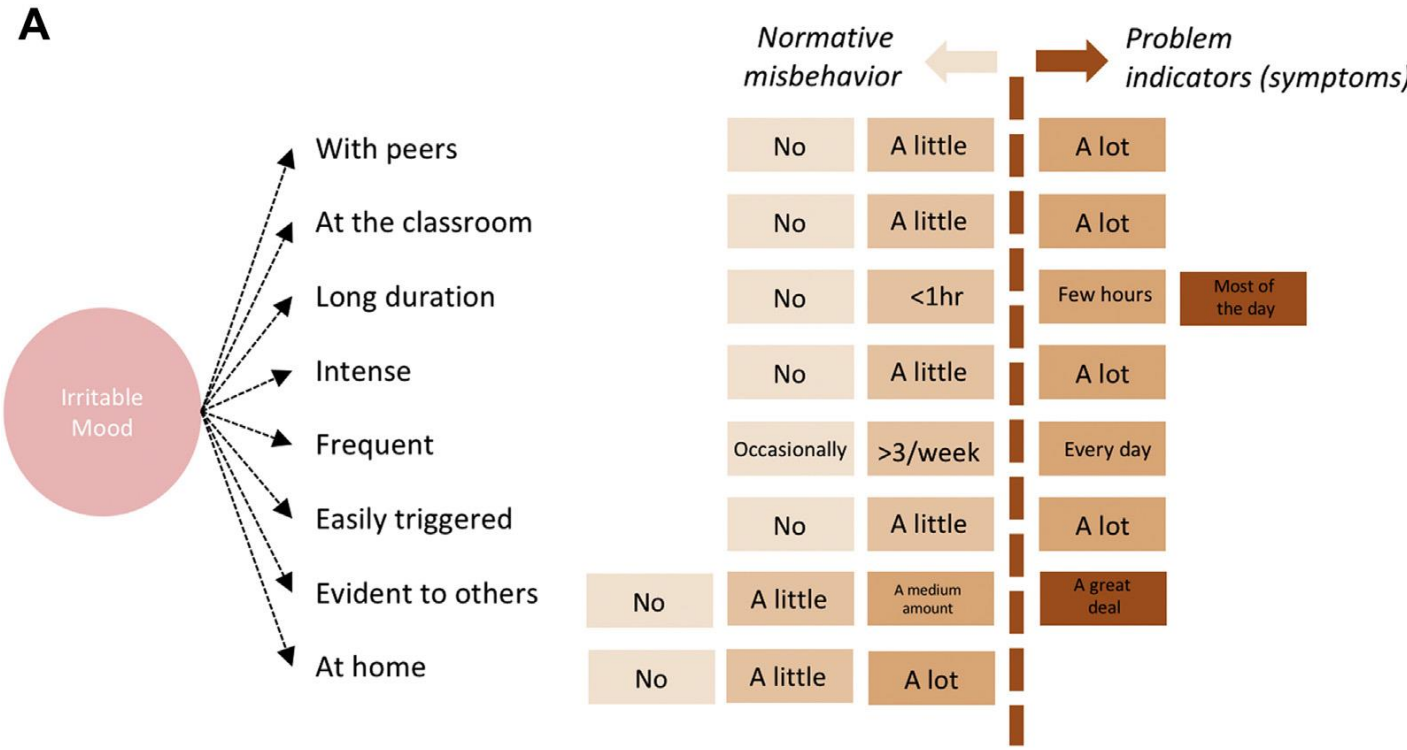
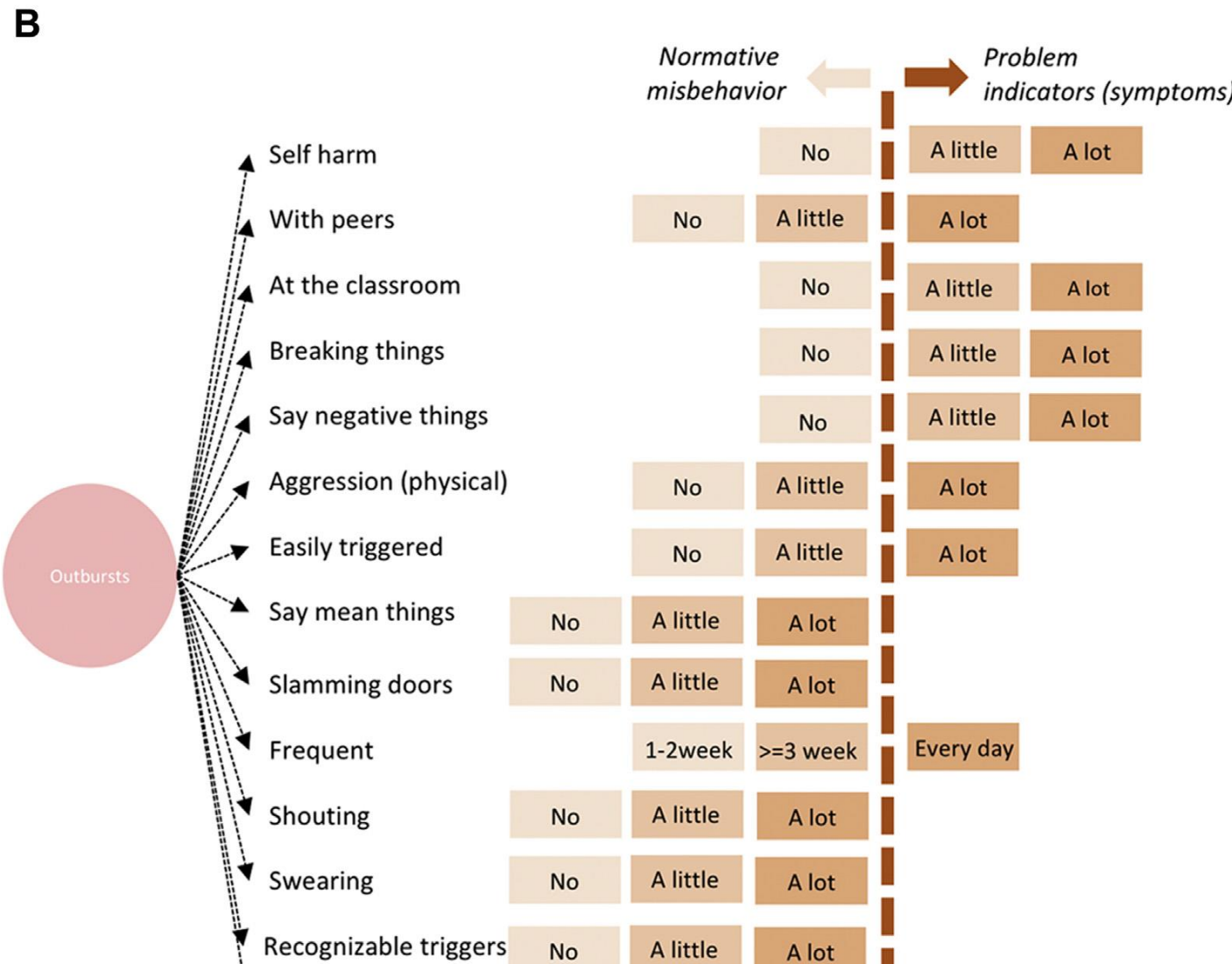


Figure 2 – Symptomatic Threshold for Each Irritable Mood and Outbursts Item in the Confirmatory Factor Analysis

Note: A) Irritable Mood. B) Outbursts.


A
Normative misbehavior
Problem indicators (symptoms)
Irritable Mood
With peers
No
A little
A lot
At the classroom
No
A little
A lot
Long duration
No
<1hr
Few hours
Most of the day
Intense
No
A little
A lot
Frequent
Occasionally
>3/week
Every day
Easily triggered
No
A little
A lot
Evident to others
No
A little
A medium amount
A great deal
At home
No
A little
A lot


B
Normative misbehavior
Problem indicators (symptoms)
Outbursts
Self harm
No
A little
A lot
With peers
No
A little
A lot
At the classroom
No
A little
A lot
Breaking things
No
A little
A lot
Say negative things
No
A little
A lot
Aggression (physical)
No
A little
A lot
Easily triggered
No
A little
A lot
Say mean things
No
A little
A lot
Slamming doors
No
A little
A lot
Frequent
1-2week
>=3 week
Every day
Shouting
No
A little
A lot
Swearing
No
A little
A lot
Recognizable triggers
No
A little
A lot

Supplement 1

Fit Indexes for Confirmatory Factor Analysis

For all CFA unidimensional models, we used delta parameterization and weighted least-square parameters using a diagonal weight matrix with standard errors and with mean and variance-adjusted chi-square test statistics (WLSMV) estimators. The measures of goodness of fit were assessed through the following fit indices: chi-square, SRMR (standardized root mean square residual), CFI (comparative fit index), TLI (Tucker-Lewis Index) and RMSEA, (root mean square error of approximation). To demonstrate good fit to the data, an estimated model should have a SRMR near or below 0.08, a RMSEA of near or below 0.06 and CFI and TLI near or above 0.95 (Hu & Bentler, 1999). Details about the estimators and the model fit used in this study can be found in Table S2, available online, which provides the fit indexes of the unidimensional models for irritability, outbursts and impairment items.

References

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Table S1 - Irritable mood, outburst and impairment items and their response categories as described in DMDD section of the DAWBA questionnaire

DAWBA Questionnaire Items	Response categories				
Irritable Mood Section					
Frequency of irritable/angry mood (p1y1)	Never	Occasionally	1-2/week	≥ 3/week	Every day
Easily irritated (p1y8)	No	A little	A lot		
Intense irritability (p1y9)	No	A little	A lot		
Long duration of irritability (p1y10)	No more than a few minutes	Less than an hour	A few hours	Most or all of the day	
Irritability evident to others (p1y11)	Not at all	A little	A medium amount	A great deal	
At home (p1y12a)	No	A little	A lot		
In the classroom (p1y12b)	No	A little	A lot		
With peers (p1y12c)	No	A little	A lot		
Angry weeks (irritable most of the day, nearly every day) [a] (p1y13)	No	Yes			
Outbursts Section					

Frequency of outbursts (p1y2)	Never	Occasionally	1-2/week	≥ 3/week	Every day
Slamming doors (p1y3a)	No	A little	A lot		
Shouting (p1y3b)	No	A little	A lot		
Swearing (p1y3c)	No	A little	A lot		
Saying mean things to others (p1y3d)	No	A little	A lot		
Saying negative things about himself (p1y3e)	No	A little	A lot		
Physical aggression to others (p1y3f)	No	A little	A lot		
Deliberate self-harm (p1y3g)	No	A little	A lot		
Breaking things (p1y3h)	No	A little	A lot		
At home (p1y4a)	No	A little	A lot		
In the classroom (p1y4b)	No	A little	A lot		
With peers (p1y4c)	No	A little	A lot		
Recognizable triggers (p1y5)	No	Perhaps	Definitely		
Easily triggered (p1y6)	No	A little	A lot		
Outburst-free gap in the last year [b] (p1y7)	Less than a day	Less than a week	Less than a month	1-3 months	More than 3 months
Impairment Section					
Impact on family life (p1y19a)	Not at all	A little	A medium amount	A great deal	
Impact on friendship (p1y19b)	Not at all	A little	A medium amount	A great deal	

Impact on learning (p1y19c)	Not at all	A little	A medium amount	A great deal
Impact on leisure (p1y19d)	Not at all	A little	A medium amount	A great deal

Note: [a] The item "angry weeks" was not included because the item "long duration of irritability" already contains the response option that captures "irritable most of the day, nearly every day"; [b] Item excluded of the analysis. The rationale for excluding this item is that it is unclear whether we would expect this item to be monotonically related to the overall latent construct given short periods of irritability with large gaps could also inform episodes of irritability (a marker of severity and bipolar disorder in children).

Table S2 - Fit indexes of the unidimensional irritable mood, outbursts and impairment Confirmatory Factor Analysis Models

Fit indexes	Irritable Mood [a]	Outbursts [b]	Impairment
FP	27	44	16
$\chi 2$	3254.236 (df=28; p<0.001)	3352.504 (df=91; p<0.001)	2865.135 (df=6; p<0.001)
RMSEA	0.057	0.057	0.032
CI 90%	0.040- 0.074	0.047-0.068	0.000-0.088
CFI	0.989	0.968	1.000
TLI	0.984	0.961	0.999
SRMR	0.057	0.085	0.019

Note: [a] Items p1y12b-p1y12c allowed to correlated due to similar item content/context; [b] Items p1y4b-p1y4c and p1y3b-p1y3c allowed to correlated due to correlated due to similar item content/context; CFI = Comparative Fit Index; CI = Confidence Interval; df = degrees of freedom; FP = Free Parameters; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual; TLI = Tucker-Lewis Index; χ^2 = Robust Chi Square Difference Test.

Table S3 – Factor Loadings and Category Threshold parameters for items in the Irritable Mood section

Items (item code in DAWBA)	Factor Loadings	T1	T2	T3	T4	Location (mean)	Location rank
Frequency/duration							
Frequency of irritable/angry mood		**Occasionally**	**1-2/week**	**≥ 3/week**	**Every day**		
	0.495			0.078	**1.143**	0.611	4
				Most/all of the			
Long duration of irritability		**<1hr**	**Few hours**	**day**			
	0.420	0.053	**0.98**	**1.726**		0.920	6
Characteristics		**A little**	**A lot**				
Easily irritated	0.905	-0.138	**0.775**			0.319	3
Intense irritability	0.910	0.281	**0.973**			0.627	5
Settings		**A little**	**A lot**				
At home	0.624	-2.123	0.303			-0.910	1
In the classroom	0.292	0.471	**1.521**			0.996	7

With peers	0.423	0.438	**1.608**			1.023	8
				A medium			
Irritability evident to others		**A little**	**amount**	**A great deal**			
	0.437	-1.345	-0.095	**1.256**		-0.061	2

Note: T1= First threshold; T2 = Second threshold; T3 = Third threshold; T4 = Fourth threshold.

Table S4 - Factor Loadings and Category Threshold parameters for items in the Temper Outbursts section

Items (item code in DAWBA)	**Factor Loadings**	**T1**	**T2**	**T3**	**T4**	**Location (mean)**	**Location rank**
Frequency/duration							
Frequency of outbursts		**Occasionally**	**1-2/week**	**≥ 3/week**	**Every day**		
	0.565			-0.068	**1.129**	-0.424	5
Characteristics		**A little**	**A lot**				
Slamming doors	0.525	-0.893	0.333			-0.274	6
Shouting	0.550	-1.004	0.062			-0.472	4
Swearing	0.693	-1.129	-0.015			-0.572	3
Saying mean things to others	0.743	-0.548	0.259			-0.143	7
Saying negative things about himself	0.481	**0.633**	**1.316**			0.981	10
Physical aggression to others	0.619	0.050	**1.086**			0.576	9
Deliberate self-harm	0.527	**1.211**	**2.030**			1.626	14
Breaking things	0.572	**0.562**	**1.471**			1.023	11

Settings		**A little**	**A lot**		
At home	0.722	-2.597	0.068	-1.264	1
In the classroom	0.351	**0.604**	**1.628**	1.123	13
With peers	0.372	0.448	**1.585**	1.024	12
Triggers					
Recognizable triggers		**Perhaps**	**Definitely**		
	-0.193	-0.815	-0.403	-0.610	2
Easily triggered		**A little**	**A lot**		
	0.664	-0.109	**0.784**	0.347	9

Note: T1 = First threshold; T2= Second threshold; T3= Third threshold; T4 = Fourth threshold.

Table S5 - Factor Loadings and Category Threshold parameters for items in the Impairment section

Items	Factor Loadings	T1	T2	T3	Location (mean)	Location rank
		A little	A medium amount	A great deal		
Impact on family life	0.709	0.018	**1.036**	**2.109**	1.054	1
Impact on friendship	0.903	**0.545**	**1.434**	**2.376**	1.452	3
Impact on learning	0.825	**0.601**	**1.238**	**2.443**	1.427	2
Impact on leisure	0.826	**0.672**	**1.521**	**2.757**	1.650	4

Note: T1 = First threshold; T2 = Second threshold; T3 = Third threshold.

Table S6 – Latent Class Model Fit Indexes and Entropy for each of the four latent class solutions for irritable mood, outbursts and impairment items

Irritable Mood

Model	log-likelihood	resid. Df	BIC	aBIC	cAIC	likelihood-ratio	Entropy
Model 1	-1.296.125	120	2.634.154	2.611.943	2.641.154	324.66270	-
Model 2	**-1.191.353**	**112**	**2472.502**	**2424.906**	**2487.502**	**115.11877**	**0.886**
Model 3	-1173.152	104	2483.992	2411.012	2506.992	78.71695	0.98
Model 4	-1167.479	96	2520.538	2422.173	2551.538	67.37120	0.948

Outbursts

Model	log-likelihood	resid. Df	BIC	aBIC	cAIC	likelihood-ratio	Entropy
Model 1	-1501.228	247	3050.347	3024.963	3058.347	335.1590	-
Model 2	**-1435.849**	**238**	**2973.468**	**2919.527**	**2990.468**	**204.4022**	**0.665**
Model 3	-1424.724	229	3005.096	2922.597	3031.096	182.1518	0.634
Model 4	-1412.391	220	3034.308	2923.251	3069.308	157.4852	0.719

Impairment

Model	log-likelihood	resid. Df	BIC	aBIC	cAIC	likelihood-ratio	Entropy
Model 1	-1015.8047	11	2055.555	2042.863	2059.555	307.99668133	-
Model 2	**-872.9925**	**6**	**1799.863**	**1771.306**	**1808.863**	**22.37226443**	**0.789**
Model 3	-866.3003	1	1816.411	1771.988	1830.411	8.98796565	0.661
Model 4	-861.8353	-4	1837.413	1777.125	1856.413	0.05789396	0.738

Note: aBIC = adjusted Bayesian Information Criteria; BIC = Bayesian Information Criterion; cAIC = Consistent Akaike Information Criterion; df = degrees of freedom.

Table S7 - Fit indexes of the unidimensional, correlated and bifactor models of irritable mood and temper outbursts (n=398)

Fit indexes	Unidimensional	Correlated with two dimensions	Bifactor [a]
FP	31	32	46
χ^2	241.4	231.625	124.387
RMSEA	0.066	0.064	0.041
CI 90%	0.056-0.076	0.054-0.074	0.028- 0.054
CFI	0.926	0.930	0.975
TLI	0.847	0.917	0.965
SRMR	0.136	0.134	0.103

Note: [a] The bifactor model resulted in negative residual variances and items with negative factor loadings in specific domains and, therefore, this model was not considered further in the analysis; CFI = Comparative Fit Index; CI = Confidence Interval; df = degrees of freedom; FP = Free Parameters; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual; TLI = Tucker-Lewis Index; χ^2 = Robust Chi Square Difference Test.

Table S8 - Left Censored regression comparing DMDD groups in SDQ scores

	Contrast with TDC			Contrast with Other DSM-5 Disorders		
	MD	SE	p-value	MD	SE	p-value
SDQ: Total						
Other disorders	9.54	0.35	<0.001	-	-	-
Irritable mood only	12.12	0.94	<0.001	2.57	1.07	0.016
Outbursts only	13.42	1.23	<0.001	3.87	1.37	0.005
Combined	15.56	0.89	<0.001	6.02	1.02	<0.001
SDQ: Emotional						
Other disorders	2.75	0.16	<0.001	-	-	-
Irritable mood only	2.94	0.42	<0.001	0.19	0.49	0.694
Outbursts only	2.46	0.55	<0.001	-0.28	0.63	0.657
Combined	3.31	0.40	<0.001	0.56	0.04	0.234
SDQ: Hyperkinetic						
Other disorders	3.65	0.21	<0.001	-	-	-
Irritable mood only	4.50	0.54	<0.001	0.85	0.53	0.113
Outbursts only	5.38	0.71	<0.001	1.73	0.68	0.011
Combined	5.35	0.52	<0.001	1.69	0.04	<0.001
SDQ: Conduct						
Other disorders	2.81	0.162	<0.001	-	-	-
Irritable mood only	3.96	0.416	<0.001	1.16	0.46	0.111
Outbursts only	5.13	0.547	<0.001	2.33	0.59	<0.001
Combined	5.71	0.396	<0.001	2.91	0.39	<0.001

SDQ: Peer relationships						
Other disorders	2.42	0.17	<0.001	-	-	-
Irritable mood only	3.11	0.43	<0.001	0.69	0.48	0.146
Outbursts only	2.82	0.57	<0.001	0.40	0.62	0.514
Combined	3.67	0.41	<0.001	1.25	0.46	<0.001
SDQ: Impact						
Other disorders	4.09	0.21	<0.001	-	-	-
Irritable mood only	5.35	0.47	<0.001	1.22	0.45	0.006
Outbursts only	4.51	0.63	<0.001	0.42	0.59	0.711
Combined	6.07	0.45	<0.001	1.94	0.43	<0.001

Note: MD = Mean Difference; SDQ = Strengths and Difficulties Questionnaire;

TDC = Typically Developing Comparisons.

Table S9 – Psychiatric comorbidities in Disruptive Mood Dysregulation Disorder defined using different combination rules for the irritable mood and temper outburst domains

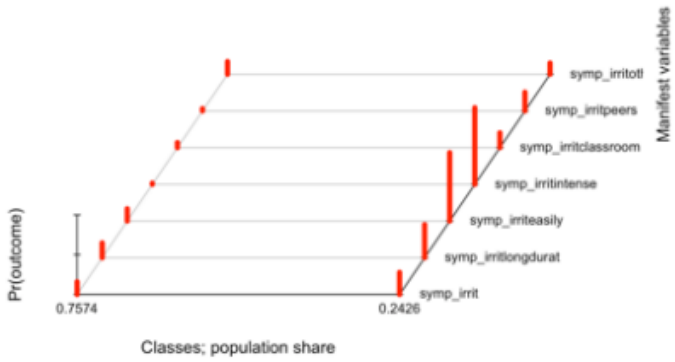
Psychiatric Comorbidities	**Current DMDD** (n=86)	**Irritable Mood** (n= 40)	**Outbursts** (n=22)	**AND Rule** (n=44)	**OR Rule** (n=107)
	Prevalence Estimation (%)				
Any disorder (except DMDD)	62.8	60.0	56.5	79.5	67.3
Any Anxiety Disorder	23.3	22.5	13.0	31.8	24.3
Generalized Anxiety Disorder	5.8	7.5	0	9.1	6.5
Separation Anxiety Disorder	7	2.5	4.3	13.6	7.5
Agoraphobia	3.5	2.5	0	4.5	2.8
Panic Disorder	1.2	0	0	2.3	0.9
Social Anxiety Disorder	3.5	10.0	0	4.5	5.6
Specific Phobia	11.6	2.5	8.7	15.9	9.3
Other Specified Anxiety Disorder	1.2	0	0	0	0
Any Mood Disorder	11.6	17.5	4.3	15.9	14.0
Major Depressive Disorder	8.1	12.5	4.3	11.4	10.3
Bipolar Disorder/Mania	3.5	7.5	0	2.3	3.7
Other Specified Depressive Disorder	1.2	2.3	0	2.5	1.9
Any Attention-Deficit/Hyperactivity Disorder	44.2	30.0	34.8	50.0	39.3
Combined	29.1	17.5	17.4	31.8	23.4
Hyperactive/impulsive	3.5	0	13.0	4.5	4.7
Inattentive	9.3	7.5	0	11.4	7.5

Another Specified ADHD	2.3	5	4.3	2.3	3.7
Any Disruptive Behavior Disorder	20.9	20.0	34.8	38.6	30.8
Oppositional Defiant Disorder	0	12.5	21.7	11.4	14.0
Conduct Disorder	18.6	10.0	17.4	27.3	18.7
Other Disruptive Behavior Disorder	2.3	0	4.3	4.5	2.8
Other Disorders					
Obsessive-Compulsive Disorder	1.2	2.5	0	2.3	1.9
Posttraumatic Stress Disorder	3.5	0	0	4.5	1.9
Autism Spectrum Disorder	3.5	0	4.3	6.8	3.7
Eating Disorders	1.2	0	0	2.3	0.9
Current DMDD	-	37.5	39.1	84.1	57.0

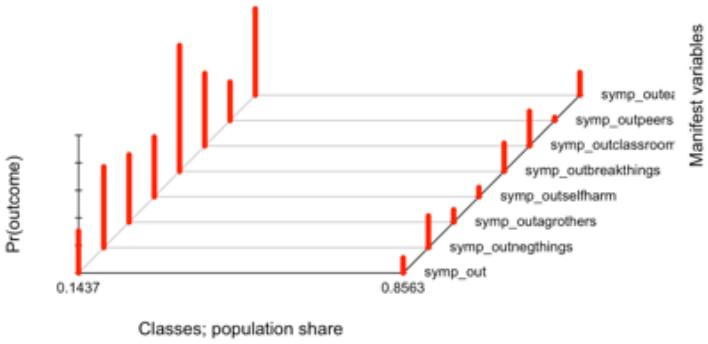
Note: DMDD = Disruptive Mood Dysregulation Disorder.

Figure S1 – Latent Class Clinical Indicator endorsement distribution between classes with low and high levels of irritable mood, outbursts and impairment

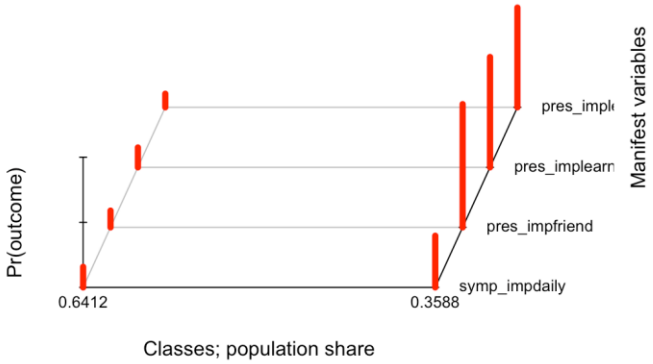
Irritable Mood



Outbursts



Impairment



Note: Graphs represents the chance of endorsement of each irritability, outburst and impairment clinical indicators between the two latent classes.

Figure S2 – Receiver Operating Curve for predicting latent classes of subjects with high levels of irritable mood, outbursts and impairment

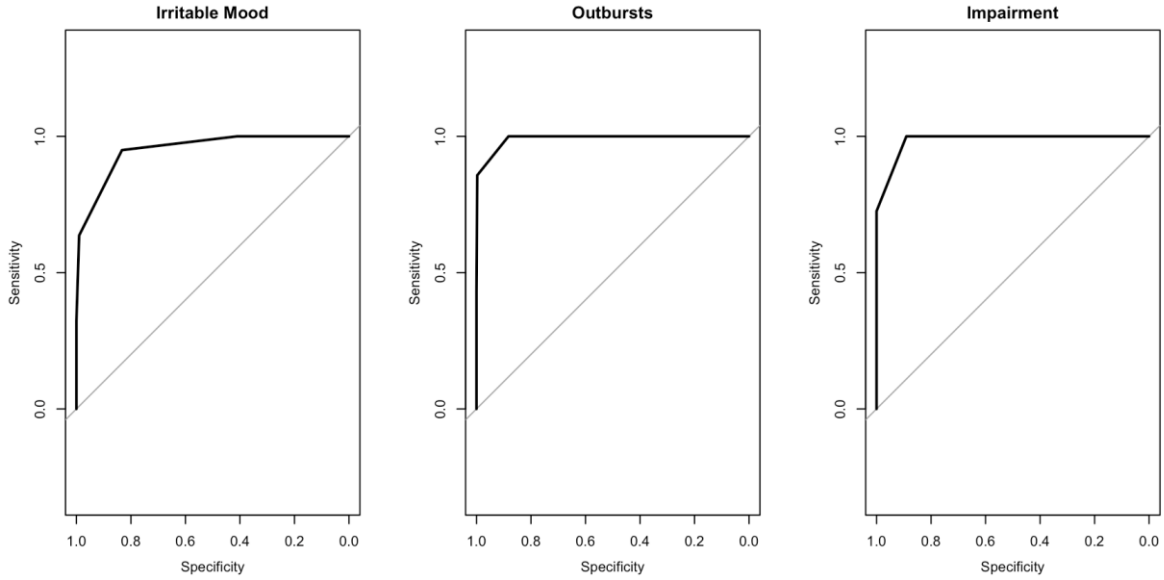


Figure S3 – Comparison of Strengths and Difficulties Questionnaire Scores among Typically Developing Pre-Adolescents/Early Adolescents, Pre-Adolescents/Early Adolescents with non-DMDD disorders, with irritable mood only, outbursts only and both irritable mood and outbursts (combined)

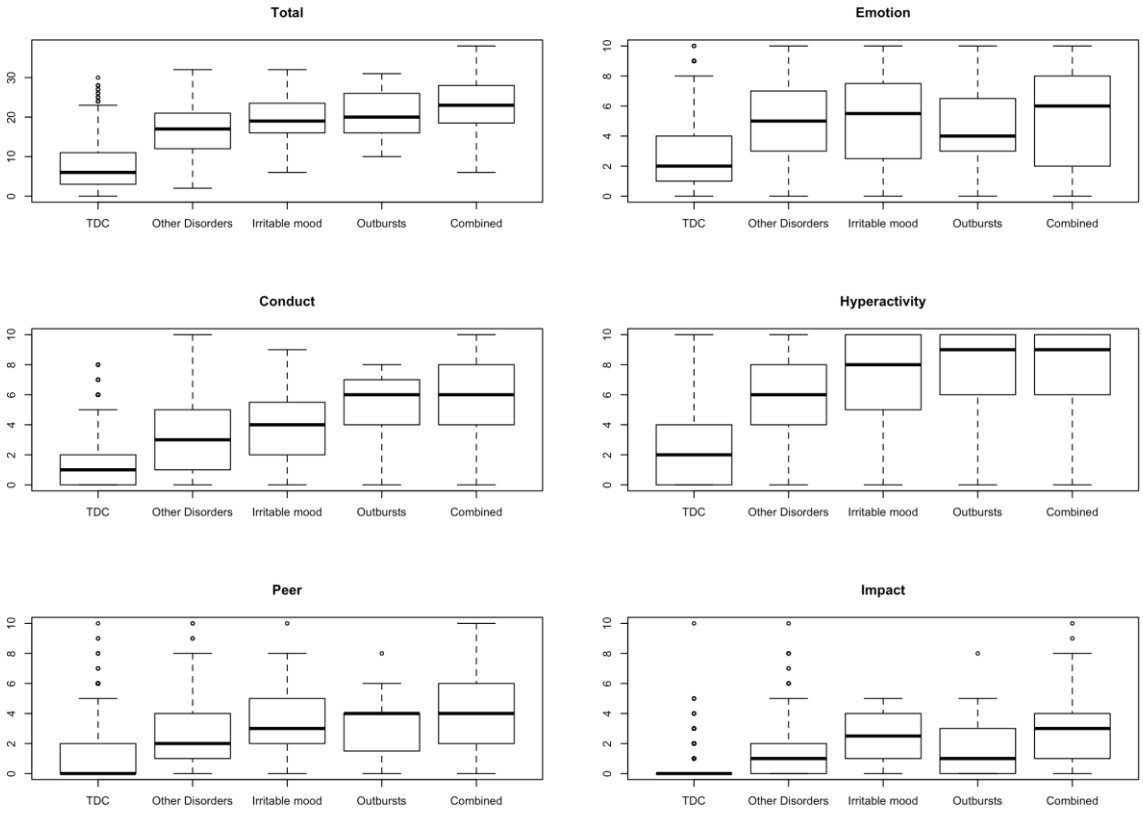


Figure S4 – Comparison of Strengths and Difficulties Questionnaire Scores between Disruptive Mood Dysregulation Disorder, defined using the OR rule, and subjects matched for psychiatric comorbidity

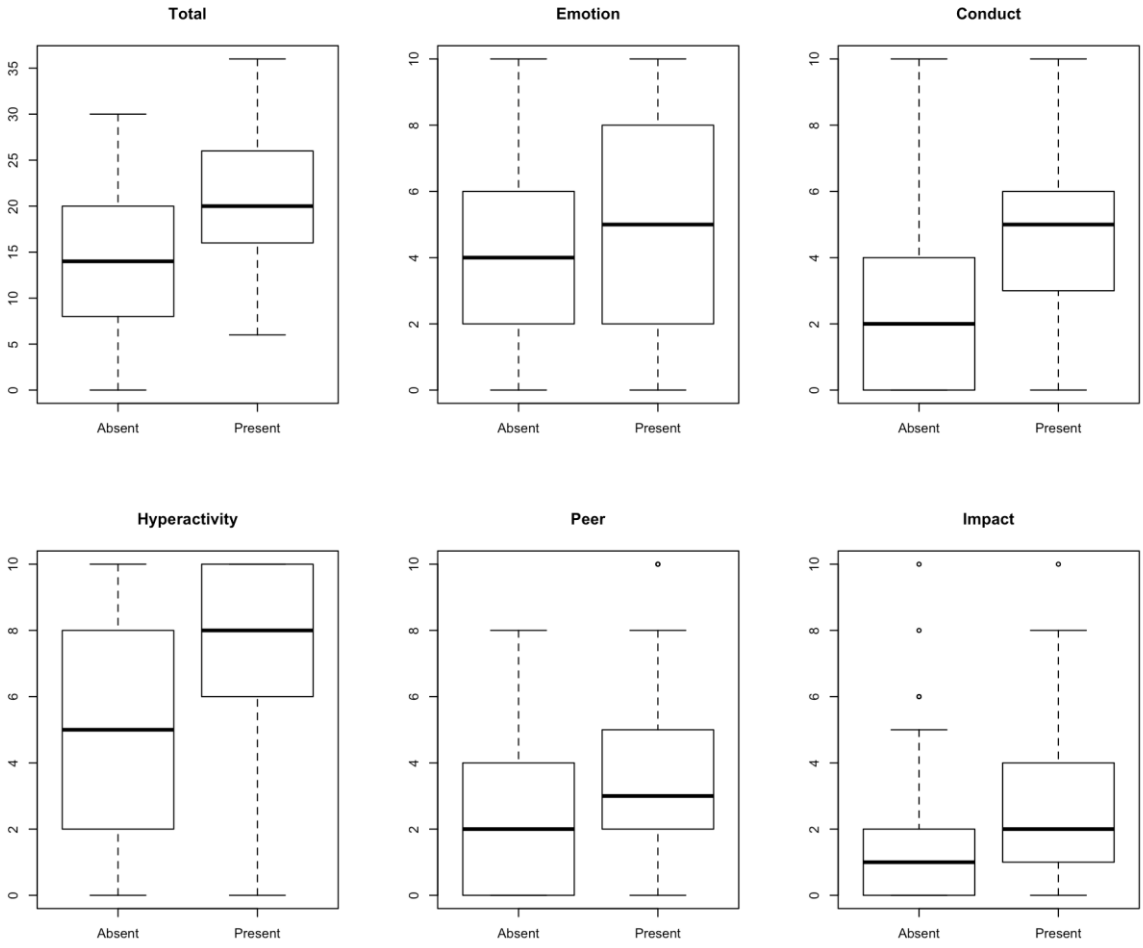
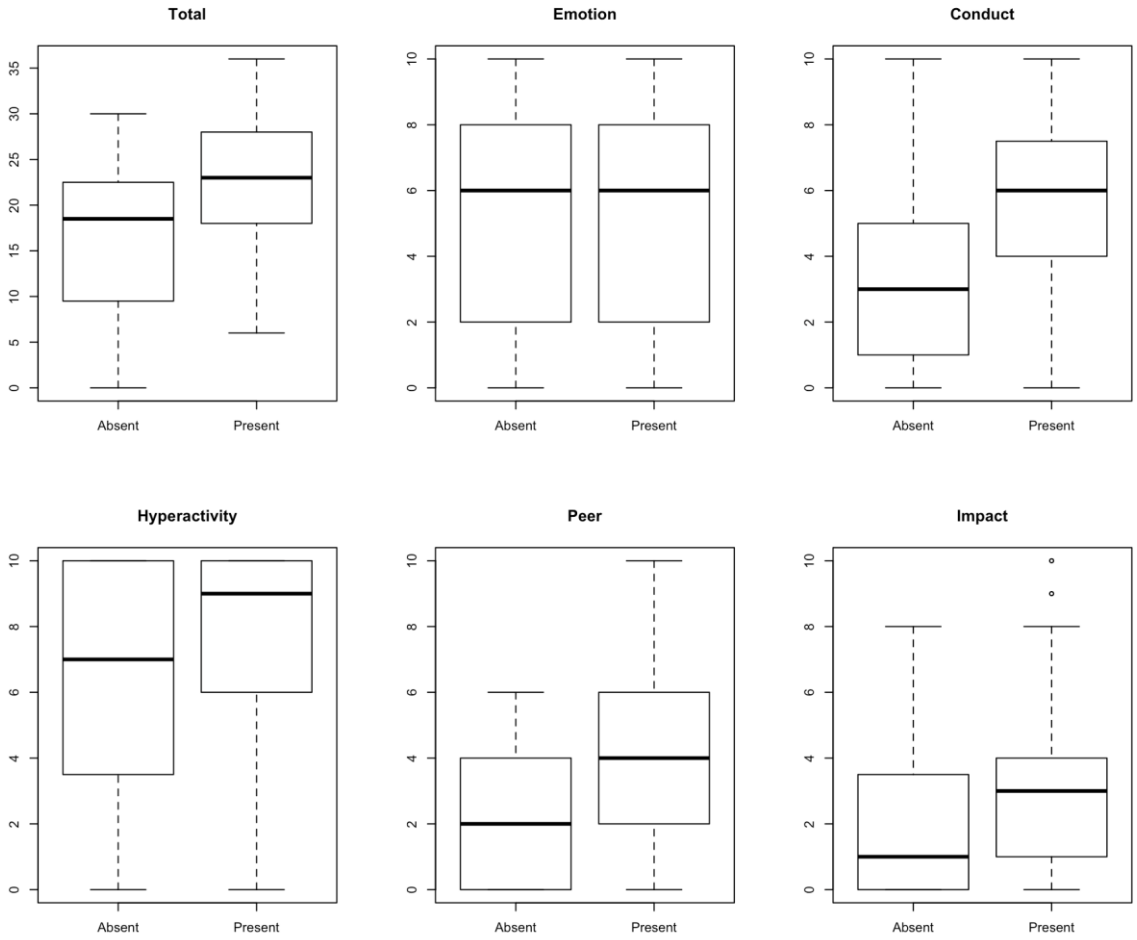


Figure S5 – Comparison between Strengths and Difficulties Questionnaire Scores between Disruptive Mood Dysregulation Disorder, defined using the AND rule, and subjects matched for psychiatric comorbidity



Article #2

With respect to specific objective b. To Expand the data-driven approach for the DMDD diagnosis in adolescents aged 14-17 and young adults ages 18-21 from Brazilian High-Risk Cohort

Ready for submission

The clinical threshold for Disruptive Mood Dysregulation Disorders in adolescents and young adults

Paola P. Laporte, Daniel S. Pine, Melissa Brotman, Katharina Kircanski, Maurício Hoffmann, [collaborators], Ellen Leibenluft, Giovanni A. Salum

Abstract

Objective: The aim of this study is to identify the most appropriate data-driven threshold for the Disruptive Mood Dysregulation Disorder (DMDD) diagnosis in adolescents and young adults and the impact of changes in diagnostic rules on prevalence.

Methods: Trained psychologists assessed 1705 adolescents (aged 14-17) and young adults (aged 18-23) with the DMDD module from the Development and Well-being Behavior Assessment (DAWBA). Participants are part of the Brazilian High-Risk Cohort Study for Mental Conditions (BHRCS, 3rd wave). First, we tested if the latent construct of irritability is comparable between groups of adolescents and adults using measurement invariance analysis. Second, we used a previously validated pipeline to assess the most appropriate symptomatic and syndromic thresholds for the diagnosis and the best way to perform clinical operationalization using AND/OR rules for combining irritable mood and outbursts.

Results: First, measurement invariance analysis showed that thresholds for measuring irritability symptoms change between adolescence and adulthood. Symptomatic threshold analyses showed irritable mood items were considered problem indicators in their highest response category for both age groups. For outbursts, some results suggested a symptom to be significant even at mild levels, while other results found no evidence of clinical significance at any level, with some differences between age groups. At the syndromic level, a combination of 3 out of 8 symptoms of irritable mood and 1 out of 10 symptoms of outbursts accurately captured a cluster of individuals with high level of symptoms in the adolescent group; in young adults 1 out of 8 symptoms of irritable mood and 1 out of 9 symptoms of outbursts are required. Analysis combining irritable mood and outbursts delineated non-overlapping aspects of DMDD. Finally, we

presented the prevalence rates for the combination of distinct DMDD diagnostic rules and show a developmental follow-back analysis and experience momentary assessment. **Conclusions:** Results provide data-driven information for future revisions of the DSM-5 and explore adaptations of the DMDD diagnosis to adulthood.

Introduction

Irritability, defined as an increased propensity to anger related to peers, is common in people seeing psychiatric treatment and associated with negative outcomes[1–3]. Irritability in childhood predicts problems in adolescence[4] and adulthood[5,6]. Disruptive Mood Dysregulation Disorder (DMDD), a new category included in the 5th edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5)[7], captures pathological manifestations of irritability in children. Epidemiological data on DMDD diagnostic thresholds are needed, specially beyond early development stages such as in adolescence and adulthood.

We recently suggested several refinements possibilities to the DMDD diagnostic criteria using a population of early adolescents of Pelotas birth cohort [12]. In our first step, the symptomatic threshold, we identified the irritable mood and outbursts items-level thresholds that differentiate normative from clinical problems in order to dichotomize response levels as clinically significant or not. In our second step, the syndromic threshold, we quantified the number of abnormal behaviors required to characterize a valid diagnosis using data-driven clustering approaches. In the last step, the clinical operationalization, we studied the irritable mood and outburst domains and demonstrated that the two differ from a latent perspective, that is, they are independently associated with impairment.

We aim now to expand this methodology for study the clinical threshold for severe and impairment irritability in adolescence and adulthood. Irritability is a symptom of three diagnoses in DSM-5[7]: Intermittent Explosive Disorder (IED), Oppositional Defiant Disorder (ODD) and DMDD. IED involves extreme temper outbursts, whereas ODD and DMDD involve both disruptive behaviors and irritable mood[8]. The DMDD diagnosis was based on the Severe Mood Dysregulation syndrome,

described by Leibenluft and colleagues in children, and made its way into the DSM-5 with some modifications. By means of a complete lack of evidence in the adult population, currently the DSM-5 requires the diagnosis of DMDD should not be performed for the first time after age 18. However, this impediment prevents research to really investigate homotypic continuity of irritability symptoms over development, given IED is typically not assessed in children and DMDD is typically not assessed in adults[9]. Also, ODD is deemed by many as too heterogeneous involving distinct developmental correlates[10,11].

In the current study, we studied 722 adolescents (14-17 years) and 983 young adults (18-23 years) who are part of the Brazilian High-Risk Cohort Study for Mental Conditions (BHRCS, 3rd wave). First, we considered whether the thresholds for DMDD in adolescents are comparable to those in young adults based on measurement invariance analysis. Then we used Confirmatory Factor Analysis (CFA) to identify item-level thresholds differentiating normative from clinical problems as assessed by the DMDD module of the DAWBA in both age groups. This was used to dichotomize response levels as clinically significant or not. We next used these binary clinical indicators as input to a Latent Class Analysis (LCA) that assigned individuals into clusters with high and low levels of clinical indicators for each domain, combined with a Receiver Operating Curves (ROC) analysis to detect the number of clinical indicators needed to predict class membership from the LCA and to translate the data-driven results to DSM-5 symptom counts. We then used AND/ OR rules to combine the irritable mood and temper outbursts symptomatic domains. Finally, we tested the impact of such rules in prevalence levels and associations with other mental health problems and included two data analysis assessing the validity of the best-matching operationalized diagnosis.

Methods

Participants

Participants of this study were 722 adolescents aged 14-17 and 983 young adults aged 18-23 which are part of the 3rd wave of the BHRCS. The BHRCS is a large and well-established school-based community cohort from two Brazilian cities: Porto Alegre and São Paulo which uses a two-stage design. We first assessed symptoms and family history of psychiatric disorders in a screening interview, collecting information from 9,937 participants at 57 schools in the cities of São Paulo and Porto Alegre, as well as from 45,394 family members. In the second stage, a random subsample (n = 958) and a high-risk subsample (children at increased risk for mental disorders, based on family risk and childhood symptoms, n = 1554) were selected for further evaluation. We evaluated those 2,511 participants using an extensive protocol. A total of 1705 participants were interviewed in the 3rd wave with a trained psychologist, which represents 68% of the original sample. The study was approved by the ethical committee of the University of Sao Paulo. Informed consent was obtained from the parents of all participants. Full details of the study design, measures, and sample have been published elsewhere[13].

Instruments and Diagnostic Assessment

The interview with the adolescents and young adults of the DMDD section from the DAWBA questionnaire[14] was administered by trained psychologists. The closed ended questions start with two skip questions about the frequency of irritable mood and temper outbursts. Participants who answered that irritable mood and/or temper outbursts

occurred at least once a week were probed to answer specific questions that characterize all DSM-5 criteria for DMDD.

A total of 347 adolescents and 495 young adults answered the DMDD section on irritable mood. This section includes 8 items characterizing the threshold for experiencing anger, intensity of anger if compared to peers of the same age, duration of anger during the day, whether irritable mood is perceived by others, setting in which anger occurs (at home, at school, with peers) and number of anger weeks throughout the year. A total of 196 adolescents and 258 young adults of the 3rd wave of the BHRCS answered the DMDD section on outbursts. This section includes 15 items describing behavior during outbursts (slamming doors, shouting, swearing, saying mean things to others, saying negative things about self, physical aggression to others, deliberate self-harm, breaking things), setting in which outbursts occurred (at home, at school, with peers) and triggers (recognizable and easily triggered). We do not use the item "outbursts free-gap in the last year" in our analysis (DSM requires that there is not a period higher than 3 or more consecutive months without irritable mood and temper outbursts). The rationale for excluding this item is that it is unclear whether we would expect this item to be monotonically related to the overall latent construct given short periods of irritability with large gaps could also inform episodes of irritability (a marker of severity and bipolar disorder).

Lastly, a total of 381 adolescents and 535 young adults that completed either the irritable mood or outburst sections were asked to also complete 4 items about impairment (impact on family life, friendship, learning, and leisure activities). After the impairment questions, participants answered the open-ended questions that allow qualitative description of the symptoms, frequency, and other characteristics of the disorder. All

questions and response categories from the DMDD section are depicted in Supplemental Table S1.

Categorical diagnoses of the main psychiatric higher-order groups (Any Anxiety Disorder, Any Depressive Disorder, Any Attention-Deficit/Hyperactivity Disorder and Any Disruptive Behavior Disorder) were performed by the DAWBA administered by trained psychologists to participants and by lay interviewers to parents. Final diagnosis was made by a psychiatrist reviewing both interviews and using a best-match approach.

The Strengths and Difficulties Questionnaire (SDQ) was used to measure dimensional psychopathology and impact of symptoms in everyday life[15,16]. The SDQ is a 25-item behavioral screening questionnaire with five domains, each of which contain five items (emotional, conduct, hyperkinetic, peer relationships, prosocial behaviors). We evaluated impact on different settings (education, family life and friendships). For all these settings, the impact was initially measured using the SDQ impact module. In this section of the questionnaire, participants were asked to what degree the difficulties interfere with the evaluated areas, classified as: "not at all", "only a little", "a medium amount", or "a great deal". We considered impairment to be present if difficulties interfered at least "a medium amount".

Statistical Analysis

Measurement invariance analysis

We used multigroup CFA (MG-CFA) to test the measurement invariance (MI) of the irritable mood, temper outbursts and impairment latent models using adolescents and young adults for group comparisons. To understand whether they are assessing equivalent constructs, a series of nested models with increasing levels of constraints

were compared. The first step was the configural invariance, which tests if the structural model fit different groups. If the same structural model fits the data well across groups, then configural invariance is supported, indicating that the factor structure is the same across groups in the MG-CFA. Our second step of invariance examination was to examine scalar invariance by constraining factor loadings and comparing model fit of this constrained model to the configural model. Scalar equivalence measure if thresholds are also equivalent across groups. Several indicators and cut-offs have been proposed to establish MI. As the $\chi 2$ is highly sensitive to sample size, we used two alternative fit indexes measures derived from the difference between the less and the more restricted models for the CFI and RMSEA. ΔCFI < .01 and ΔRMSEA < .015 or ΔSRMR < .010 between nested models with increasing levels of constrains indicate that the model is invariant[17].

Symptomatic threshold

The eight items on irritable mood, fourteen items on outbursts and four items on impairment of the adolescents and young adults were included in a MG-CFA testing unidimensional models for each construct for each age-group. Details about the estimators and the model fit used in this study can be found in Table S2, available online, which provides the fit indexes of the unidimensional models for irritability, outbursts and impairment items. CFA models estimate item level factor loadings (λ) and response category thresholds. Factor loadings represent the strength of the relationship between the latent trait and the item, i.e., they indicate how well each item discriminates different severity levels of a given construct. Category thresholds indicate the expected value of the latent factor at which there is a 50% probability of endorsing a given category or higher i.e., the category threshold indicates the severity level at which the transition from

one response category to the next is likely to happen (e.g., from 'No' to 'A little' or higher, or from 'A little' to 'A lot').

To distinguish normative misbehavior from behavior that would meet a diagnostic criterion, we used category thresholds from the CFA. CFAs were performed only in participants with a frequency of irritable mood and outbursts greater than once a week. We interpreted values below 0.5 as typical development (normative) and values at or above .5 as 'clinical indicators' (a proxy for symptoms or problem indicators). This limen was used in our model paper[12] and was also performed in other investigations[18]. The thresholds above .5 represents an approximation to the top 5% most symptomatic adolescents and young adults in the population.

Syndromic threshold

Before data analysis, each questionnaire item was dichotomized at the value of the category threshold defined in the symptomatic threshold analysis described above in participants with at least one clinical indicator. Dichotomized items were chosen to enter the latent class analysis (LCA) because our intention was not to characterize varying levels of irritability in the community, but to identify groups that differ in their number of clinical indicators. LCA were used to create empirically derived groups of levels of clinical indicators for irritable mood, temper outbursts and impairment for the adolescents and young adults' groups separately. Next, we used Receiving Operating Curves to predict the most accurate number of clinical indicators for detecting participants with high levels of symptoms (as defined by latent class analysis) with regard to irritable mood, outburst and impairment, in both age groups. ROC analysis was used as a way to translate results from the syndromic thresholds of the LCA to the reality of clinical practice, which uses symptom counts. Thus, the ROC identifies a simple rule to allow the identification of adolescents and young adults that are likely to be members of the cluster that exhibit a

high level of clinical indicators. The optimal cut-off was estimated using the Youden's J Statistic, which maximizes both sensitivity and specificity[19].

Clinical operationalization

To study the four solutions – irritable mood only, outbursts only, 'OR' rule (irritable mood or outbursts) and 'AND' rule (irritable mood and outbursts) - for clinical operationalization we used the following strategies for age groups separately: 1) we tested whether meeting criteria for the irritable mood group and/or for the outbursts group have distinct or overlapping associations with the impaired functioning group using a multiple logistic regression; 2) we used linear regressions adjusted for outcomes to compare SDQ total and impact dimensional scores between participants meeting criteria for the four solutions with a group of participants with other DSM disorders except for DMDD and typically developing comparisons; 3) for both OR and AND groups, we used a matching procedure to compare levels of SDQ scores between a group that differed in DMDD status (yes vs. no DMDD) but were otherwise fully matched for comorbidities.

Epidemiological Impact and validity

Finally, using the relative frequency, we investigated the impact of these diagnostic solutions (AND/OR rules and combinations with impairment requirements) on the prevalence rates of DMDD in the community. Prevalence rates were adjusted by attrition and corrects for the oversampling procedure using sampling weights that were constructed to represent the selected sample at baseline (which is representative of the community). Weights were constructed using propensity score weighting approach for the people investigated in the sample to match the randomly selected sample from the

ascertainment phase of the study. This strategy corrects both for attrition and for the oversampling strategy used in the study to increase the number of high-risk participants.

We have also included two data analysis assessing the validity of the best-matching operationalized diagnosis. *First*, a developmental follow-back retrospective analysis in which we compare trajectories of irritability symptoms measured by the parent-reported Child Behavior Checklist (CBCL) and Adult Behavior Checklist (ABCL) irritability items (items 3, 86, 87 and 95) in three timepoints (baseline, wave 2 and wave 3 – 4,794 observations from 2511 participants over the past 10 years) between subjects with and without DMDD in the last timepoint (wave 3). This analysis was performed using Generalized Additive Models Mixture Models, using age splines as random slopes for each participant. *Second*, a prospective Experience Momentary Assessment (EMA) collected in a subsample of participants during a COVID-19 pandemic. We selected participants with at least 7 valid responses among the 70 probes for data collection sent over a 2-week period (5 probes per day). This resulted in 8146 observations from 251 participants (25 participants with DMDD and 226 without DMDD). We compared the mean and standard deviation of probes with irritable mood (answered "somewhat angry", "angry" or "very angry") as opposed to non-irritable mood (answered "neutral", "somewhat calm", "calm", "very calm") in this 2-week period. This assessment occurred on average 2 years after the DMDD diagnosis.

The analysis were performed in Mplus version 8.6[20] and R version 3.6.1[21], including applications implemented in the packages lavaan 0.6-5[22], poLCA 1.4.1[23], pROC 1.15.3[24], CensReg 0.5-26[25], MatchIt 3.0.2[26] and data.table 1.14.0[27].

Results

Measurement invariance analysis

MI analysis are presented in Table 1. Results shows that the increasing level of constraint in irritable mood and impairment models simultaneously modified CFI and RMSEA or SRMR beyond the threshold proposed, contrary to what happened with the outburst model. We can conclude, for the irritable mood and impairment models, that we are measuring different thresholds of the latent construct of irritable mood and impairment in each age group. This result justifies conducting a separate investigation for adolescents and young adults.

Table 1 around here

Symptomatic threshold

The eight irritable mood items in adolescents were considered problem indicators (proxy for symptoms) in their highest category of response in the questionnaire. When we approach the items "frequency of irritable mood" and "long duration", the irritable mood has to be present every day and most of the day to be considered a problem indicator. The items "easily irritated" and "intensity" where considered problem indicators when a lot. The same applies to the settings "at home", "in the classroom" and "with peers". The item "irritable mood evident to others" has to be graded "a great deal" to be considered a symptom (Table S3, available online). Irritable mood in young adults showed similar results to the irritable mood in adolescents, except for the item "irritable mood that happens in the classroom" which is considered a symptom even when present as "a little" (Table S4, available online).

For outbursts some indicated a symptom even when present at only a mild level, while others did not indicate symptoms at any level, with differences in three items between age groups. When analyzing outbursts in adolescents, 10 out of 14 are considered symptoms. Outbursts have to occur every day to be considered problem indicators, being the items "shouting" and "swearing" always considered normative; "self-harm", "breaking things" and "physical aggression" considered problem indicators even when "a little"; and "saying negative things about himself", "saying mean things to others" and "slamming doors" considered problem indicators only when "a lot". The settings "with peers" and "at the classroom" are considered symptoms only when "a lot", while they are always normative when they occur "at home". The item "easily triggered" is only considered a problem when "a lot" while "recognizable trigger" is always normative (Table S5, available online). When comparing to young adults, 9 out of 14 are considered symptoms. The item "at the classroom" becomes a symptom even when "a little" and the item "breaking things" is symptom only when "a lot". As for the item "saying mean things" it becomes normative (Table S6, available online).

For impairment, the four settings (impact on family life, friendship, learning and leisure) are normative when "A little" and indicate a symptom when "a medium amount" or "a lot" in both populations (Tables S7 and S8, available online). See Figure 2.

(Table S3 –> Table S8, available online).

Figure 2 around here

Syndromic threshold

The prevalence of each clinical indicator is presented in Table 2. All LCA analyses indicated that the two-class solution was the best for each of the three domains for adolescents and young adults (Table S9, available online). This indicates that, in each of these domains, the population is divided into two groups characterized by high vs. low symptoms. We next performed ROC analyses to determine the best number of clinical indicators (i.e., those items identified by the CFA) to use to predict membership in the high vs. low symptom classes identified by the LCA. Youden's J demonstrates that for adolescents we need 3 out of 8 symptoms of irritable mood and 1 out of 10 symptoms of outbursts and 1 environment for that to become a diagnostic. For young adults we need 1 out of 8 symptoms of irritable mood and 1 out of 9 symptoms of outbursts and 1 environment for that to become a diagnostic. See Table 3 for details.

Clinical operationalization

To study our proposed solutions (combinations of the 'OR' and 'AND' rules), we have the following results for both age groups: (1) both irritable mood and outbursts were associated with clinical impairment in multiple models adjusted for the effects of including both predictors in the same model (adolescents: irritable mood adjusted OR= 8.87, $p<0.001$; outbursts adjusted OR= 5.93, $p<0.001$; young adults: irritable mood adjusted OR= 5.64, $p<0.001$; outbursts adjusted OR= 4.2, $p<0.001$); (2) comparisons between irritable only, outbursts only and combined groups with typically developing comparisons and with a group of patients with other DSM disorders (except for DMDD) showed all DMDD solutions had higher scores on total and impact SDQ scales than typically developing comparisons and the combined solution had higher total and impact SDQ total scores than participants with other DSM diagnosis (Table S10, available online); (3) left-censored regressions comparing groups matched for

comorbidity (any anxiety, any mood, any hyperkinetic and any disruptive behavior disorder) showed that, using either the OR or the AND rule, the DMDD group showed higher total, emotional, conduct, hyperactivity, peer relationship and impact scores than did the non-DMDD group with matched comorbidities (Figure S1 and Figure S2, available online).

Epidemiological Impact and validity

In adolescents, ROC analysis requires an impact in at least 1 setting and have the following prevalence rates: 1.1% only irritable mood, 4.8% only outbursts, 7.4% irritable mood or outbursts, and 3.0 % both irritable mood and outbursts. In young adults, ROC analysis requires an impact in at least 1 setting and have the following prevalence rates: 3.2% only irritable mood, 1.1% only outbursts, 7.4% irritable mood or outbursts, and 3.0 % both irritable mood and outbursts. Prevalence rates for other combinations can be found in Table 4.

In the developmental follow-back analysis, there was a significant time by DMDD status interactions (edf=1.772, F=10.3, $p<0.00005$), meaning that for children younger than 12 years of age irritability symptoms were lower than children without DMDD and for participants older than 12 years of age, participants have higher irritability symptoms as assessed by the CBCL. In the EMA analysis, on average the DMDD group spend 9.4% more probes with irritable mood than the non-DMDD group (t=3.54, p=0.00047) with standard deviations 11.2% higher (t=4.097, $p<0.00001$).

Discussion

This study aims to advance the study clinical threshold determination for DMDD in adolescents and young adults. We found that the thresholds for endorsing clinical

indicators for irritable mood and irritability-related impairment are different between adolescents and young adults. At symptomatic threshold, the CFA showed that all irritable mood items were normative in the low response categories and clinical indicators in the high response categories; irritable mood at the classroom in young adults was the only difference in relation to adolescents being a symptom even when mild. For outbursts, the threshold for a clinical indicator varied across items, but differences between age groups for those criteria were modest. At syndromic threshold, 3 irritable mood symptoms, 1 outburst symptom and impairment in at least one setting showed to be the best cut-off in adolescents; while 1 irritable mood symptom, 1 outburst symptom and impairment in at least one setting would be the best cut-offs for adults. Clinical operationalization revealed irritable mood and outbursts were independently associated with impairment; and the OR-rule groups showed comparable or even higher levels of impairment than other DSM disorders. Matched analysis showed that results cannot be attributed to comorbidity. The prevalence rates of all the diagnostic criteria for distinct combinations rules were presented. Finally, developmental follow-back analysis revealed it DMDD in adolescence and young adulthood is unlikely to be a continuity of childhood DMDD and prospective EMA analysis provided support for the diagnosis to predict irritability in intensive methods of data capture.

In our previous study with 11-12-year-old adolescents[12] we showed very similar results in terms of indicators of problematic behavior. Notable exceptions include irritable mood occurring at home that was normative in our previous work, but which is considered a symptom in our current work. Concerning outbursts, "slamming doors" and "saying mean things" were considered normative in our previous work and found to be symptoms in the current work (the last only for the adolescent group). Our previous work identified that a combination of 2 irritable mood, 3 outbursts and

significant impairment in at least two settings discriminated a group of affected individuals in this developmental period. Our current work showed the need for 3 irritable mood, 1 outburst and impairment in one setting for adolescents; and 1 irritable mood, 1 outburst and impairment in one setting for young adults. These results reveal the importance of fine tune diagnostic criteria to each developmental stage. We also suggested that either irritable mood or outbursts might be enough for capture significant impairment, which provide further evidence for the importance to discuss the possibility of diagnosing DMDD without the requirement of both manifestations.

While some considerable data exists for DMDD in youth, this is the first study to explore this diagnostic category in adults. Irritability in adults is listed as a cardinal symptom of IED, ODD and DMDD in the DSM-5. IED is characterized by the presence of disruptive behaviors (e.g., extreme temper outbursts, aggression); whereas ODD and DMDD are characterized by both disruptive behaviors and irritable mood[8]. All these categories are not reconciled throughout development and this imposes important difficulties for the field to move forward. Although this needs much discussion and debate, our results suggest the importance of finding a diagnostic home for adults with irritable mood without outbursts. One possibility is to adopt something similar to ADHD, in which DMDD could encompass irritable mood only presentations, outbursts only presentation and the combined presentation. Coccaro et al studied the relationship between IED and DMDD and acknowledged that phenomenologically they differ from each other in terms of persistent inter-outburst irritable mood[28].

Irritability was found to be a significant predictor of depression, anxiety, and ODD[29], exemplifying its role as part of a pattern of heterotypical comorbidity among these disorders. Our study investigates the possibility of DMDD diagnosis in adults, which provide an interesting alternative for capturing the irritable mood component in

this population. However, this category is unlikely to capture homotypic continuity (i.e., irritability in children predicting irritability in adults), given developmental follow-back analysis revealed that symptoms began to emerge in the adults with DMDD after age 12, and not before. The independence of irritable mood and outbursts components of irritability in a sample of adults with severe mental disorders was psychometrically studied and showed that the domains might have common and distinct patterns[30]. Our data support an OR rule when combining irritable mood and outbursts, rather than the AND rule currently in the manual.

Our study has important strengths. *First*, we are the first study aiming to studying DMDD criteria in adults with diagnostic instruments designed specifically for that purpose. *Second*, we relied on a large population sample and implemented assessment methods that could mimic clinical assessment in the real world, as far as possible in an epidemiological investigation. *Third*, we applied Confirmatory Factor Analysis, Latent Class Analysis and Receiver Operating Curves Analysis, applying a similar framework used in irritability in preadolescents and in other disorders given consistency to diagnostic panels to deliberate on best ways to characterize irritability over development[31,32]. *Fourth*, we assessed the validity of the best-matching operationalized diagnosis with a developmental follow-back retrospective analysis comparing trajectories of irritability symptoms and a prospective EMA collected in a subsample of the same cohort during the COVID-19 pandemic. However, this work has also some important limitations. *First*, our analysis is predominant focused on internal validators, and for the external validators we used instruments that are not specifically designed to diagnose DMDD. Further studies investigating course, family history, treatment response, genetics and neuroimaging are needed to demonstrate the validity of the operationalized syndrome. *Second*, because of the skip rule questions, the CFA parameters were estimated for participants with irritable

mood or outbursts that occurred at least once a week. Analysis were modeled to consider these characteristics, but this might have biased the parameter estimates for some items. Also, our approach assumes irritable mood, outbursts and impairment are distinct domains, which is still an empirical question to be further tested.

Finally, this study provides some clinical guidance on how to identify adolescents and young adults requiring clinical attention on the irritability continuum. Future research to advance the field of DMDD should include replicating these findings; extending similar approaches to diagnostic instruments other than the DAWBA; examining symptomatic thresholds using measures that do not have skipping rules and are designed specifically to differentiate normative versus non-normative behaviors; and include external validators. Better understanding of the phenomenology of irritability might allow for advances in classification of such symptoms, improve our current nosology and, perhaps, leverage future research to provide better care for these patients.

References

1. Wakschlag LS, Estabrook R, Petitclerc A, et al. Clinical Implications of a Dimensional Approach: The Normal:Abnormal Spectrum of Early Irritability. *J Am Acad Child Adolesc Psychiatry*. 2015;54(8):626-634. doi:10.1016/j.jaac.2015.05.016
2. Petitclerc A, Briggs-Gowan MJ, Estabrook R, et al. Contextual variation in young children's observed disruptive behavior on the DB-DOS: implications for early identification. *J Child Psychol Psychiatry*. 2015;56(9):1008-1016. doi:10.1111/jcpp.12430
3. Perlman SB, Jones BM, Wakschlag LS, Axelson D, Birmaher B, Phillips ML. Neural substrates of child irritability in typically developing and psychiatric populations. *Dev Cogn Neurosci*. 2015;14:71-80. doi:10.1016/j.dcn.2015.07.003
4. Cicchetti D. Developmental Psychopathology. In: *The Handbook of Life-Span Development*. American Cancer Society; 2010. doi:10.1002/9780470880166.hlsd002014
5. Copeland WE, Shanahan L, Egger H, Angold A, Costello EJ. Adult Diagnostic and Functional Outcomes of DSM-5 Disruptive Mood Dysregulation Disorder. *AJP*. 2014;171(6):668-674. doi:10.1176/appi.ajp.2014.13091213
6. Brotman MA, Schmajuk M, Rich BA, et al. Prevalence, Clinical Correlates, and Longitudinal Course of Severe Mood Dysregulation in Children. *Biological Psychiatry*. 2006;60(9):991-997. doi:10.1016/j.biopsych.2006.08.042
7. AMERICAN PSYCHIATRIC ASSOCIATION. *Diagnostic and Statistical Manual of Mental Disorders*. Fifth Edition (DSM-V).; 2013.
8. Brotman MA, Kircanski K, Stringaris A, Pine DS, Leibenluft E. Irritability in Youths: A Translational Model. *Am J Psychiatry*. 2017;174(6):520-532. doi:10.1176/appi.ajp.2016.16070839
9. Copeland WE, Shanahan L, Costello EJ, Angold A. Childhood and adolescent psychiatric disorders as predictors of young adult disorders. *Arch Gen Psychiatry*. 2009;66(7):764-772. doi:10.1001/archgenpsychiatry.2009.85
10. Stringaris A, Goodman R. Longitudinal outcome of youth oppositionality: irritable, headstrong, and hurtful behaviors have distinctive predictions. *J Am Acad Child Adolesc Psychiatry*. 2009;48(4):404-412. doi:10.1097/CHI.0b013e3181984f30
11. Boylan K. The many faces of oppositional defiant disorder. *J Can Acad Child Adolesc Psychiatry*. 2014;23(1):8-9.
12. Laporte PP, Matijasevich A, Munhoz TN, et al. Disruptive Mood Dysregulation Disorder: Symptomatic and Syndromic Thresholds and Diagnostic Operationalization. *J Am Acad Child Adolesc Psychiatry*. Published online January 28, 2020. doi:10.1016/j.jaac.2019.12.008
13. Salum GA, Gadelha A, Pan PM, et al. High risk cohort study for psychiatric disorders in childhood: rationale, design, methods and preliminary results. *Int J Methods Psychiatr Res*. 2015;24(1):58-73. doi:10.1002/mpr.1459
14. Goodman R, Ford T, Richards H, Gatward R, Meltzer H. The Development and Well-Being Assessment: description and initial validation of an integrated assessment of child and adolescent psychopathology. *J Child Psychol Psychiatry*. 2000;41(5):645-655.
15. Goodman R, Ford T, Simmons H, Gatward R, Meltzer H. Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *Int Rev Psychiatry*. 2003;15(1-2):166-172.

doi:10.1080/0954026021000046128
16. Stringaris A, Goodman R. The value of measuring impact alongside symptoms in children and adolescents: a longitudinal assessment in a community sample. *J Abnorm Child Psychol*. 2013;41(7):1109-1120. doi:10.1007/s10802-013-9744-x
17. Chen FF. Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*. 2007;14(3):464-504. doi:10.1080/10705510701301834
18. Wakschlag LS, Choi SW, Carter AS, et al. Defining the developmental parameters of temper loss in early childhood: implications for developmental psychopathology. *J Child Psychol Psychiatry*. 2012;53(11):1099-1108. doi:10.1111/j.1469-7610.2012.02595.x
19. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3(1):32-35.
20. Muthén & Muthén, Mplus Home Page. Accessed April 5, 2021. https://www.statmodel.com/
21. R: The R Project for Statistical Computing. Accessed August 26, 2018. https://www.r-project.org/
22. Rosseel Y, Oberski D, Byrnes J, et al. *Lavaan: Latent Variable Analysis*.; 2018. Accessed August 26, 2018. https://CRAN.R-project.org/package=lavaan
23. poLCA.pdf. Accessed September 19, 2019. https://cran.r-project.org/web/packages/poLCA/poLCA.pdf
24. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12:77. doi:10.1186/1471-2105-12-77
25. CRAN - Package censReg. Accessed May 27, 2019. https://cran.r-project.org/web/packages/censReg/index.html
26. Ho D, Imai K, King G, Stuart E, Whitworth A. *MatchIt: Nonparametric Preprocessing for Parametric Causal Inference*.; 2018. Accessed September 19, 2019. https://CRAN.R-project.org/package=MatchIt
27. Dowle M, Srinivasan A, Gorecki J, et al. *Data.Table: Extension of "Data.Frame."*; 2021. Accessed April 5, 2021. https://CRAN.R-project.org/package=data.table
28. Coccaro EF. DSM-5 intermittent explosive disorder: Relationship with Disruptive Mood Dysregulation Disorder. *Compr Psychiatry*. 2018;84:118-121. doi:10.1016/j.comppsych.2018.04.011
29. Vidal-Ribas P, Brotman MA, Valdivieso I, Leibenluft E, Stringaris A. The Status of Irritability in Psychiatry: A Conceptual and Quantitative Review. *J Am Acad Child Adolesc Psychiatry*. 2016;55(7):556-570. doi:10.1016/j.jaac.2016.04.014
30. Knackfuss ACU, Leibenluft E, Brotman MA, et al. Differentiating irritable mood and disruptive behavior in adults. *Trends Psychiatry Psychother*. 2020;42(4):375-386. doi:10.1590/2237-6089-2019-0078
31. Kessler RC, Green JG, Adler LA, et al. Structure and diagnosis of adult attention-deficit/hyperactivity disorder: analysis of expanded symptom criteria from the Adult ADHD Clinical Diagnostic Scale. *Arch Gen Psychiatry*. 2010;67(11):1168-1178. doi:10.1001/archgenpsychiatry.2010.146
32. Matte B, Anselmi L, Salum GA, et al. ADHD in DSM-5: a field trial in a large, representative sample of 18- to 19-year-old adults. *Psychol Med*. 2015;45(2):361-373. doi:10.1017/S0033291714001470

Table 1 - Measurement invariance of irritable mood, outbursts and impairment models across age groups

Model	Sample in each group	Invariance	$\chi 2$	*df*	RMSEA	CFI	TLI	SRMR	Model comparisson	$\Delta\chi 2$ (Δdf)	Δ CFI	Δ RMSEA	Δ SRMR	Decision
Irritable Mood	Adolescents = 347 Young Adults = 495	Configural	101.444***	38	0.063	0.969	0.954	0.047						
		Scalar	186.719***	54	0.076	0.935	0.933	0.060	Configural	77.661 (16)***	0.034	0.013	0.013	Reject
Outbursts	Adolescents = 196 Young Adults = 258	Configural	265.998***	150	0.058	0.937	0.924	0.074						
		Scalar	302.991***	176	0.056	0.931	0.929	0.081	Configural	46.63 (26)**	0.006	0.002	0.007	Accept
Impairment	Adolescents = 381 Young Adults = 535	Configural	2.178	4	0.000	1	1	0.008						
		Scalar	39.270***	14	0.063	0.984	0.986	0.023	Configural	34.505 (10)***	0.016	0.063	0.015	Reject

Note: Invariance decision is based on ΔCFI < 0.010 supplemented by ΔRMSEA < 0.015 or ΔSRMR < 0.010. $\chi 2$, Chi square test; DF, degrees of freedom; RMSEA, Root Mean Square Error of Approximation; CFI, Comparative Fit Index; TLI, Tucker-Lewis Index; SRMR, Standardized Root Mean-Square Residual; Δ, differences between fit index. **, $p < 0.01$; ***, $p < 0.001$

Table 2 – Prevalence of Each Disruptive Mood Dysregulation Disorder Item Written in Combination with the Response Category That Defines a Clinical Indicator for Adolescents and Young Adults weighted for attrition and BHRCS oversampling procedure

	Irritable Mood	Prevalence Estimation of Clinical Indicators (%)			
		Adolescents		Young Adults	
	Frequency/duration				
1	How often have you been in an angry or irritable mood?	Every day	3.9	Every day	4
2	When you become irritable or angry, how long do you typically stay that way?	Most of the day	4.5	Most of the day	5.6
3	Compared with others of your age, do you easily get annoyed, or become irritable or angry?	A lot	7.7	A lot	8.2
4	Do you get into seriously irritable or angry moods that are stronger and more intense than is usual for others of your age?	A lot	2.7	A lot	6.1
	Where do you become irritable or angry?				
5	At home?	A lot	12.6	A lot	11.7
6	In the classroom?	A lot	5.4	A little	13.1

7	When you are with other people of about your age?	A lot	21	A lot	2.2
8	*When you are in an angry or irritable mood, is this obvious to most other people?*	A great deal	4.4	A great deal	4.1
	Temper Outbursts	Prevalence Estimation of Clinical Indicators (%)			
		Adolescents		Young Adults	
	Frequency/duration				
1	How often have you had a temper outburst?	Every day	1.5	Every day	1.6
	When you have a temper outburst, do these involve the following:				
2	Saying negative things about yourself	A lot	6.7	A lot	6
3	Physical aggression directed to other people	A little	5.1	A little	5
4	Deliberately hurting yourself	A little	6.3	A little	4.9
5	Breaking things	A little	5.5	A lot	2.4

6	Slamming doors	A lot	5.2	A lot	3.5
7	Saying mean things to other people	A lot	6.7	-	
	Where do your temper outbursts occur?				
8	In the classroom?	A lot	1.5	A little	3.1
9	When you are with other people of about your age?	A lot	1.7	A lot	1
10	*Some people have temper outbursts that are triggered very easily. Compared with others of your age, is this true of you?*	A lot	2.2	A lot	3.8

Note: Prevalence estimates assume that adolescents and young adults whose irritable mood and outbursts occurred less than once per week (and who therefore did not complete these items) do not have any of these problems to a significant degree. Adjusted for weight.

Table 3 - Receiver Operating Curves Parameters Investigating the Best Number of Clinical Indicators to Capture Latent Class Groups

Threshold	Prediction of Latent Class Groups - Adolescents																	
	Irritable Mood						Temper Outbursts						Severity of Impairment					
	ACC	Sens	Spe	PPV	NPV	YI	ACC	Sens	Spe	PPV	NPV	YI	ACC	Sens	Spe	PPV	NPV	YI
0	0.2	1.00	0.00	0.22	-	0	0.82	0.76	1.00	1.00	0.62	0.76	0.21	1.00	0.21	0.21	-	0.21
1	0.36	1.00	0.18	0.26	1.00	0.18	**0.98**	**0.97**	**1.00**	**1.00**	**0.93**	**0.78**	**0.92**	**1.00**	**0.73**	**0.73**	**1.00**	**0.73**
2	0.76	1.00	0.69	0.48	1.00	0.69	0.91	1.00	0.69	0.89	1.00	0.69	0.93	0.65	1.00	1.00	0.91	0.65
3	**0.95**	**0.95**	**0.95**	**0.83**	**0.98**	**0.90**	0.84	1.00	0.43	0.82	1.00	0.43	0.84	0.21	1.00	1.00	0.83	0.21
4	0.90	0.54	0.99	0.94	0.88	0.53	0.78	1.00	0.22	0.76	1.00	0.22	0.81	0.07	1.00	1.00	0.80	0.07
5	0.84	0.26	1.00	1.00	0.82	0.26	0.74	1.00	0.10	0.74	1.00	0.10						
6	0.81	0.15	1.00	1.00	0.81	0.15	0.73	1.00	0.06	0.73	1.00	0.06						
7	0.79	0.06	1.00	1.00	0.79	0.06	0.72	1.00	0.02	0.72	1.00	0.02						
8							0.72	1.00	0.02	0.72	1.00	0.02						
Threshold	Prediction of Latent Class Groups - Young Adults																	
	Irritable Mood						Temper Outbursts						Severity of Impairment					
	ACC	Sens	Spe	PPV	NPV	YI	ACC	Sens	Spe	PPV	NPV	YI	ACC	Sens	Spe	PPV	NPV	YI
0	0.40	0.17	1.00	1.00	0.31	0.17	0.81	0.73	1.00	1.00	0.61	0.73	0.81	0.79	1.00	1.00	0.40	0.79
1	**0.82**	**0.75**	**1.00**	**1.00**	**0.60**	**0.75**	**0.98**	**0.97**	**1.00**	**1.00**	**0.94**	**0.97**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
2	0.90	0.95	0.77	0.92	0.87	0.72	0.91	1.00	0.70	0.89	1.00	0.70	0.93	1.00	0.42	0.92	1.00	0.42
3	0.85	0.99	0.46	0.83	0.98	0.45	0.81	1.00	0.35	0.78	1.00	0.35	0.89	1.00	0.12	0.88	1.00	0.12
4	0.79	1.00	0.26	0.78	1.00	0.26	0.74	1.00	0.12	0.73	1.00	0.12	0.87	1.00	0.00	0.87	-	0
5	0.75	1.00	0.09	0.74	1.00	0.09	0.72	1.00	0.07	0.72	1.00	0.07						

6	0.73	1.00	0.02	0.73	1.00	0.02	0.71	1.00	0.01	0.71	1.00	0.01
7	0.73	1.00	0.01	0.73	1.00	0.01	0.71	1.00	0.01	0.71	1.00	0.01
8							0.71	1.00	0.00	0.71	1.00	0

Note: ACC = Accuracy; NPV = Negative Predictive Value; PPV = Positive Predictive Value; Sens = Sensitivity; Spe = Specificity; YI = Youden's Index.

Table 4 – Impact of Different Rules for Combining Irritable Mood and Temper Outbursts Clinical Indicators and Impairment Requirements on Prevalence Rates of DMDD Solutions in Adolescents and Young Adults in a Weighted Analysis.

	Adolescents				Young Adults			
Number of symptoms	**1 setting**	**2 settings**	**3 settings**	**4 settings**	**1 setting**	**2 settings**	**3 settings**	**4 settings**
Best-match ROC Analysis								
Irritable mood only	**1.1**	0.9	0	0	**3.2**	1.1	0.6	0.2
Ourbursts only	**4.8**	0.2	0.1	0	**1.1**	0.3	0.3	0.2
AND rule	**3.0**	1.8	0.9	0.2	**4.5**	2.6	1.3	0.4
OR rule	**7.4**	3.3	0.9	0.2	**8.9**	3.9	2.2	0.7
1 symptom for each domain								
Irritable mood only	4.5	1.1	0	0		1.1	0.6	0.2
Ourbursts only	1.1	0.4	-	-	**1.1**	0.3	0.3	0.2
AND rule	5.2	2.0	0.9	0.2	**4.5**	2.6	1.3	0.4
OR rule	10.8	3.5	0.9	0.2	**8.9**	3.9	2.2	0.7
2 symptoms for each domain								
Irritable mood only	3.9	0.9	0.1	0.1	2.3	0.7	0.3	0.2
Ourbursts only	0.9	0	0	-	1.1	0.3	0.3	0.2
AND rule	3.5	2.0	0.9	0.1	3.6	2.4	1.3	0.4
OR rule	8.3	2.9	0.9	0.2	7.1	3.5	1.8	0.7
3 symptoms for each domain								
Irritable mood only	1.5	1.1	0.1	0.1	1.7	0.7	0.5	0.2
Ourbursts only	0.9	0.2	0.1	0	1.1	0.6	0.4	0.3

AND rule	2.6	1.6	0.8	0.1	2.7	1.9	0.9	0.3
OR rule	4.9	2.9	0.9	0.2	5.5	3.3	1.8	0.7
4 symptoms for each domain								
Irritable mood only	0.9	0.9	0.1	0	1.9	1.0	0.6	0.3
Ourbursts only	1.0	0.1	0.1	-	0.9	0.6	0.4	0.3
AND rule	1.3	1.0	0.7	1.8	1.3	1.1	0.7	0.1
OR rule	3.1	1.9	0.8	0.1	4	2.7	1.6	0.7
5 symptoms for each domain								
Irritable mood only	0.3	0.2	0.1	-	1.6	0.9	0.6	0.3
Ourbursts only	0.7	0.3	0.3	-	0.2	0.2	0.2	-
AND rule	0.6	0.5	0.3	0.1	0.4	0.4	0.2	1.6
OR rule	1.6	0.9	0.7	0.1	2.1	1.4	1	0.4

Note: Bold marks the optimal solution found in ROC analysis.

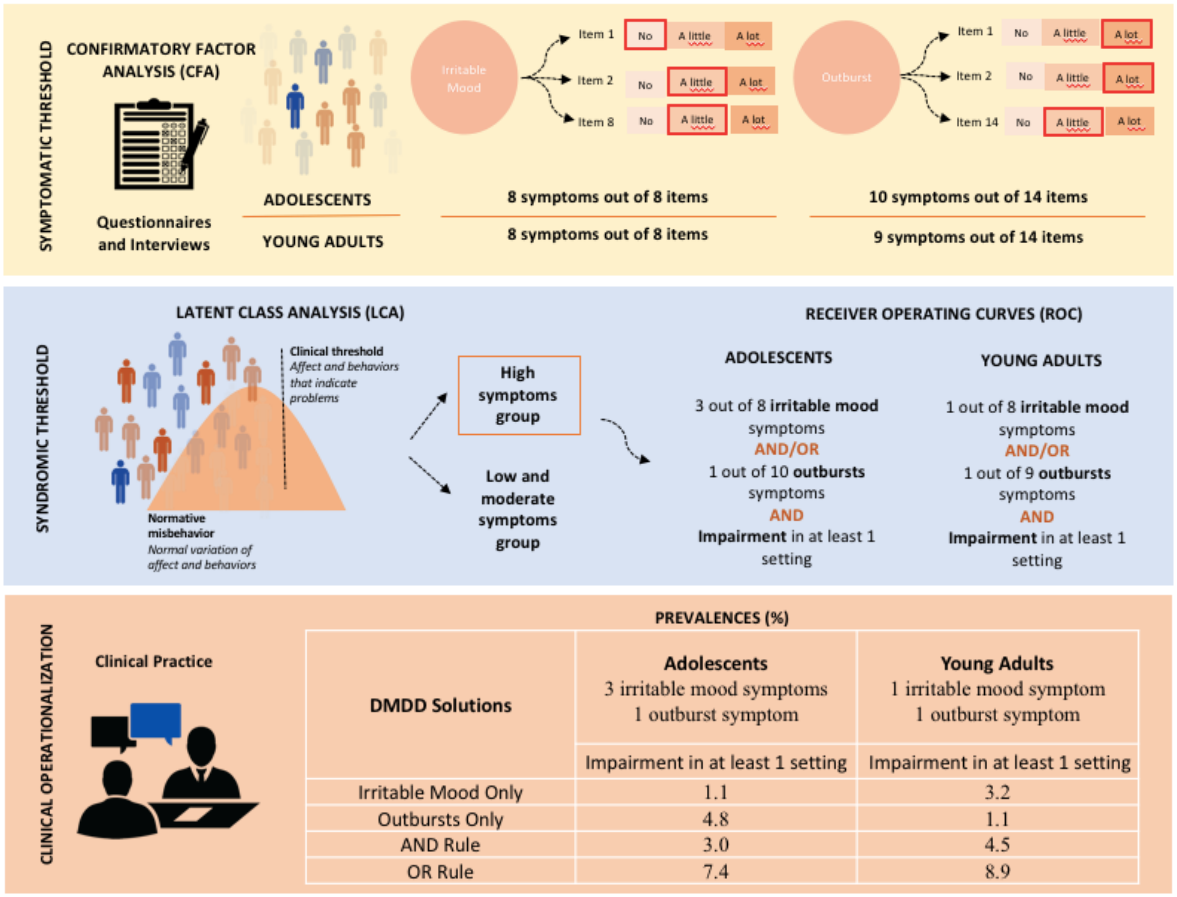
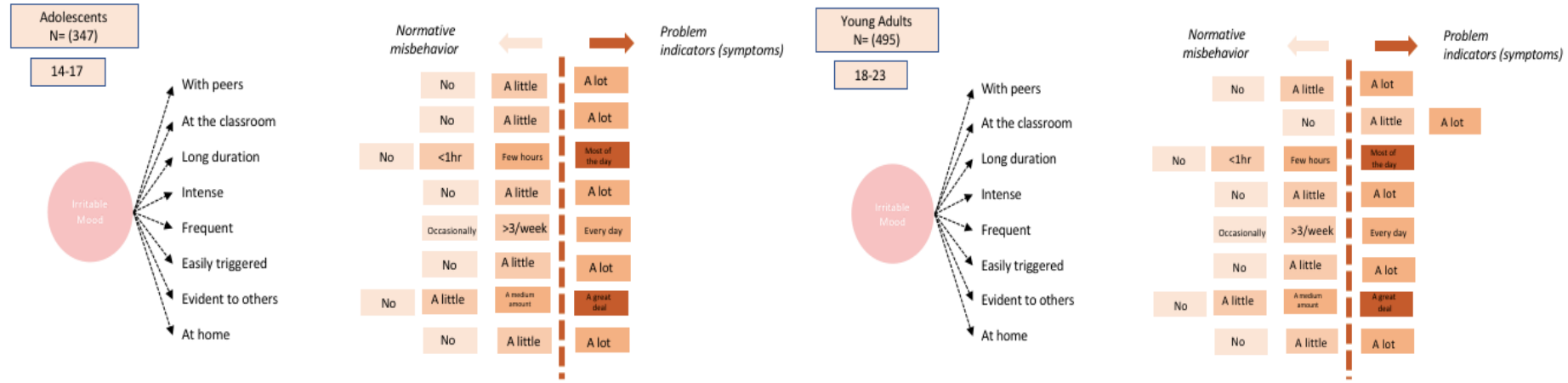


Figure 1 – Symptomatic and Syndromic Thresholds and Clinical Operationalizations

A


Adolescents
N= (347)
14-17
Irritable Mood
With peers
At the classroom
Long duration
Intense
Frequent
Easily triggered
Evident to others
At home
Normative misbehavior
Problem indicators (symptoms)
No
A little
A lot
No
A little
A lot
No
<1hr
Few hours
Most of the day
No
A little
A lot
Occasionally
>3/week
Every day
No
A little
A lot
No
A little
A medium amount
A great deal
No
A little
A lot
Young Adults
N= (495)
18-23
Irritable Mood
With peers
At the classroom
Long duration
Intense
Frequent
Easily triggered
Evident to others
At home
Normative misbehavior
Problem indicators (symptoms)
No
A little
A lot
No
A little
A lot
No
<1hr
Few hours
Most of the day
No
A little
A lot
Occasionally
>3/week
Every day
No
A little
A lot
No
A little
A medium amount
A great deal
No
A little
A lot

B

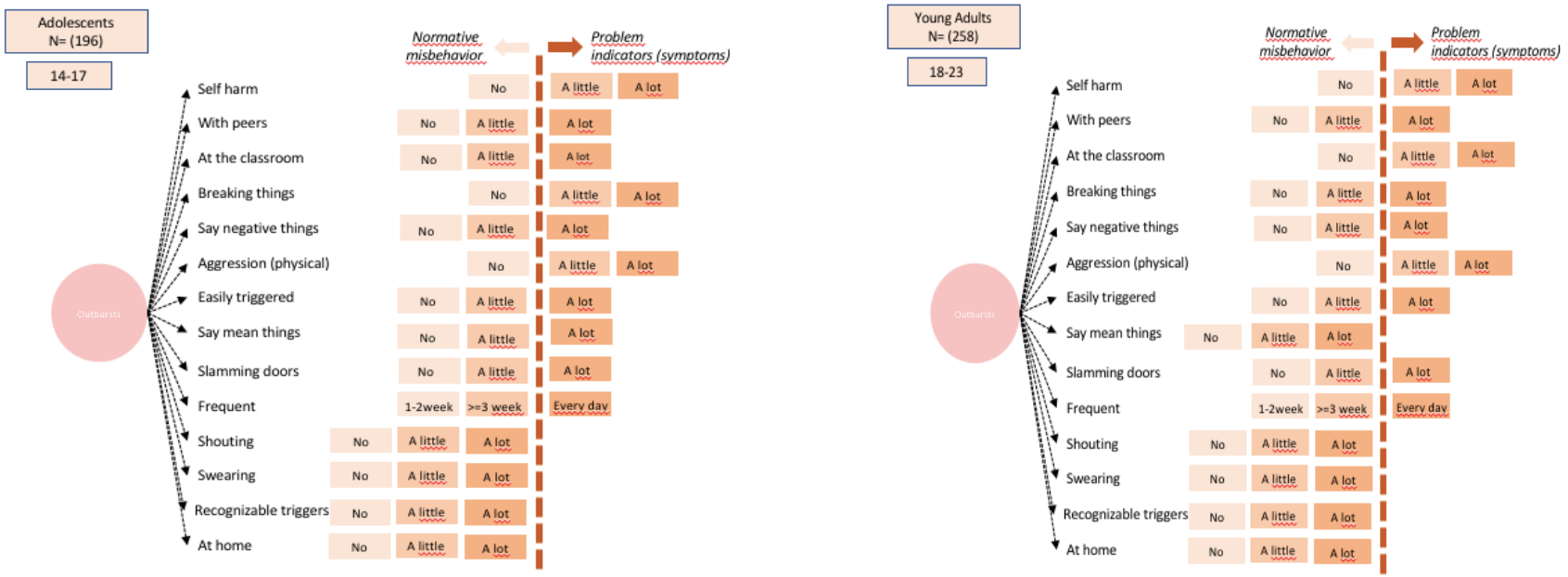


Figure 2 - Symptomatic Threshold for Each Irritable Mood and Outbursts Item in the Multiple Group Confirmatory Factor Analysis

Note: A) Irritable Mood. B) Temper Outbursts.

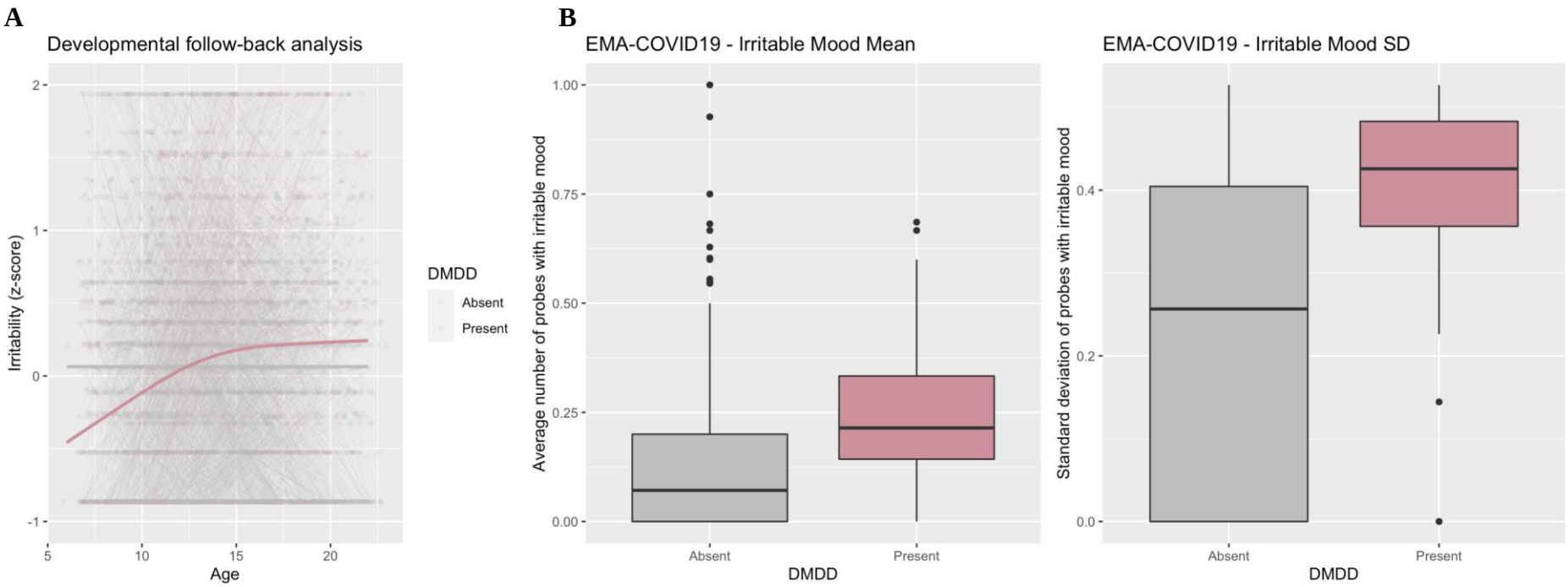


Figure 3 – The validity of the best-matching operationalized diagnosis. A) A developmental follow-back trajectory of irritability symptoms measured by the parent-reported Child Behavior Checklist and Adult Behavior Checklist ABCL irritability items. B) A prospective Experience Momentary Assessment (EMA) collected in a subsample of participants during the COVID-19 pandemic.

Supplement 1

Fit Indexes for Confirmatory Factor Analysis

For all CFA multiple-group unidimensional models, we used delta parameterization and weighted least-square parameters using a diagonal weight matrix with standard errors and with mean and variance-adjusted chi-square test statistics (WLSMV) estimators. The measures of goodness of fit were assessed through the following fit indices: chi-square, SRMR (standardized root mean square residual), CFI (comparative fit index), TLI (Tucker-Lewis Index) and RMSEA, (root mean square error of approximation). To demonstrate good fit to the data, an estimated model should have a SRMR near or below 0.08, a RMSEA of near or below 0.06 and CFI and TLI near or above 0.95 (Hu & Bentler, 1999). Details about the estimators and the model fit used in this study can be found in Tables S2 and S3, availables online, which provides the fit indexes of the unidimensional models for irritability, outbursts and impairment items in adolescents and young adults for both self and parent report samples.

References

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Table S1 - Irritable mood, outburst and impairment items and their response categories as described in DMDD section of the DAWBA questionnaire

DAWBA Questionnaire Items	Response categories				
Irritable Mood Section					
Frequency of irritable/angry mood (sy1)	Never	Occasionally	1-2/week	≥ 3/week	Every day
Easily irritated (sy8)	No	A little	A lot		
Intense irritability (sy9)	No	A little	A lot		
Long duration of irritability (sy10)	No more than a few minutes	Less than an hour	A few hours	Most or all of the day	
Irritability evident to others (sy11)	Not at all	A little	A medium amount	A great deal	
At home (sy12a)	No	A little	A lot		
In the classroom (sy12b)	No	A little	A lot		
With peers (sy12c)	No	A little	A lot		
Angry weeks (irritable most of the day, nearly every day) [a] (sy13)	No	Yes			
Outbursts Section					
Frequency of outbursts (sy2)	Never	Occasionally	1-2/week	≥ 3/week	Every day
Slamming doors (sy3a)	No	A little	A lot		

Shouting (sy3b)	No	A little	A lot		
Swearing (sy3c)	No	A little	A lot		
Saying mean things to others (sy3d)	No	A little	A lot		
Saying negative things about himself (sy3e)	No	A little	A lot		
Physical aggression to others (sy3f)	No	A little	A lot		
Deliberate self-harm (sy3g)	No	A little	A lot		
Breaking things (sy3h)	No	A little	A lot		
At home (sy4a)	No	A little	A lot		
In the classroom (sy4b)	No	A little	A lot		
With peers (sy4c)	No	A little	A lot		
Recognizable triggers (sy5)	No	Perhaps	Definitely		
Easily triggered (sy6)	No	A little	A lot		
Outburst-free gap in the last year [b] (sy7)	Less than a day	Less than a week	Less than a month	1-3 months	More than 3 months
Impairment Section					
Impact on family life (sy19a)	Not at all	A little	A medium amount	A great deal	
Impact on friendship (sy19b)	Not at all	A little	A medium amount	A great deal	
Impact on learning (sy19c)	Not at all	A little	A medium amount	A great deal	
Impact on leisure (sy19d)	Not at all	A little	A medium amount	A great deal	

Note: [a] The item "angry weeks" was not included because the item "long duration of irritability" already contains the response option that captures "irritable most of the day, nearly every day"; [b] Item excluded of the analysis. The rationale for excluding this item is that it is unclear whether we would expect this item to be monotonically related to the overall latent construct given short periods of irritability with large gaps could also inform episodes of irritability (a marker of severity and bipolar disorder in children).

Table S2 - Fit indexes of the unidimensional irritable mood, outbursts and impairment Multiple Group Confirmatory Factor Analysis Models for Adolescents and Young Adults.

Fit indexes	Irritable Mood [a]	Outbursts [b]	Impairment
FP	54	88	32
$\chi 2$	68.046 (df=38; p=0.002)	207.202 (df=150; p=0.001)	1.154 (df=4; p=0.886)
RMSEA	0.043	0.041	0.000
CI 90%	0.026- 0.060	0.026-0.054	0.000-0.033
CFI	0.989	0.981	1.000
TLI	0.984	0.977	1.004
SRMR	0.055	0.083	0.011

Note: [a] Items p1y12b-p1y12c allowed to correlated due to similar item content/context; [b] Items p1y4b-p1y4c and p1y3b-p1y3c allowed to correlated due to correlated due to similar item content/context; CFI = Comparative Fit Index; CI = Confidence Interval; df = degrees of freedom; FP = Free Parameters; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual; TLI = Tucker-Lewis Index; χ^2 = Robust Chi Square Difference Test. Number of observations per group: Irritable Mood - Adolescents 366, Young Adults 420; Outbursts - Adolescents 203, Young Adults 224; Impairment - Adolescents 230, Young Adults 250.

Table S3 – Factor Loadings and Category Threshold parameters for items in the Irritable Mood section for Adolescents, using self-report assessment.

Items (item code in DAWBA)	**Factor Loadings**	**T1**	**T2**	**T3**	**T4**	**Location (mean)**	**Location rank**
Frequency/duration							
Frequency of irritable/angry mood		**Occasionally**	**1-2/week**	**≥ 3/week**	**Every day**		
	0.652			0.366	**1.293**		7
Long duration of irritability		**<1hr**	**Few hours**	**Most/all of the day**			
	0.482	-0.428	0.350	**1.260**			4
Characteristics		**A little**	**A lot**				
Easily irritated	0.880	-0.389	**0.909**				5
Intense irritability	0.812	0.222	**1.327**				6
Settings		**A little**	**A lot**				

At home	0.431	-1.037	**0.619**		1
In the classroom	0.261	-0.230	**1.088**		3
With peers	0.410	0.098	**1.504**		8
			A medium		
Irritability evident to others		**A little**	**amount**	**A great deal**	
	0.360	-1.088	0.428	**1.327**	2

Note: T1= First threshold; T2 = Second threshold; T3 = Third threshold; T4 = Fourth threshold.

Table S4 – Factor Loadings and Category Threshold parameters for items in the Irritable Mood section for Young Adults, using self-report assessment.

Items (item code in DAWBA)	**Factor Loadings**	**T1**	**T2**	**T3**	**T4**	**Location (mean)**	**Location rank**
Frequency/duration							
Frequency of irritable/angry mood		**Occasionally**	**1-2/week**	**≥ 3/week**	**Every day**		
	0.590			0.196	**1.169**		5
Long duration of irritability		**<1hr**	**Few hours**	**Most/all of the day**			
	0.398	-0.629	0.196	**1.130**			6
Characteristics		**A little**	**A lot**				
Easily irritated	0.853	-0.420	**0.901**				3
Intense irritability	0.861	0.232	**1.169**				4
Settings		**A little**	**A lot**				

At home	0.552	-1.179	**0.568**		1
In the classroom	0.183	**0.660**	**1.820**		8
With peers	0.362	0.099	**1.723**		7
			A medium		
Irritability evident to others		**A little**	**amount**	**A great deal**	
	0.488	-1.179	0.381	**1.276**	2

Note: T1= First threshold; T2 = Second threshold; T3 = Third threshold; T4 = Fourth threshold.

Table S5 - Factor Loadings and Category Threshold parameters for items in the Temper Outbursts section for Adolescents, using self-report assessment.

Items (item code in DAWBA)	**Factor Loadings**	**T1**	**T2**	**T3**	**T4**	**Location (mean)**	**Location rank**
Frequency/duration							
Frequency of outbursts		**Occasionally**	**1-2/week**	**≥ 3/week**	**Every day**		
	0.475			0.312	**1.465**		11
Characteristics		**A little**	**A lot**				
Slamming doors	0.494	-0.090	**0.882**				7
Shouting	0.605	-0.612	0.380				4
Swearing	0.538	-0.536	0.421				3
Saying mean things to others	0.653	-0.259	**0.643**				5
Saying negative things about himself	0.674	-0.154	**0.691**				6
Physical aggression to others	0.487	**0.792**	**1.741**				14

Deliberate self-harm	0.710	**0.757**	**1.361**	12
Breaking things	0.519	**0.774**	**1.588**	13
Settings		**A little**	**A lot**	
At home	0.581	-1.215	0.464	1
In the classroom	0.299	0.193	**1.394**	9
With peers	0.209	0.026	**1.429**	10
Triggers				
Recognizable triggers		**Perhaps**	**Definitely**	
	0.042	-0.845	0.272	2
Easily triggered		**A little**	**A lot**	
	0.731	-0.167	**1.299**	8

Note: T1 = First threshold; T2= Second threshold; T3= Third threshold; T4 = Fourth threshold.

Table S6 - Factor Loadings and Category Threshold parameters for items in the Temper Outbursts section for Young Adults, using self-report assessment.

Items (item code in DAWBA)	**Factor Loadings**	**T1**	**T2**	**T3**	**T4**	**Location (mean)**	**Location rank**
Frequency/duration							
Frequency of outbursts		**Occasionally**	**1-2/week**	**≥ 3/week**	**Every day**		
	0.581			0.146	**1.322**		9
Characteristics		**A little**	**A lot**				
Slamming doors	0.545	-0.347	**0.922**				7
Shouting	0.644	-0.782	0.136				4
Swearing	0.756	-1.048	0.156				2
Saying mean things to others	0.703	-0.420	0.399				5
Saying negative things about himself	0.511	-0.205	**0.518**				6
Physical aggression to others	0.636	**0.681**	**1.422**				12

Deliberate self-harm	0.534	**0.705**	**1.422**	13
Breaking things	0.611	0.431	**1.194**	11
Settings		**A little**	**A lot**	
At home	0.675	-1.654	0.088	1
In the classroom	0.193	**1.031**	**2.067**	14
With peers	0.277	0.127	**1.813**	10
Triggers				
Recognizable triggers		**Perhaps**	**Definitely**	
	0.204	-0.809	0.205	3
Easily triggered		**A little**	**A lot**	
	0.642	-0.378	**0.952**	8

Note: T1 = First threshold; T2= Second threshold; T3= Third threshold; T4 = Fourth threshold.

Table S7 - Factor Loadings and Category Threshold parameters for items in the Impairment section for Adolescents, using self-report assessment.

Items	Factor Loadings	T1	T2	T3	Location (mean)	Location rank
		A little	A medium amount	A great deal		
Impact on family life	0.763	-0.155	**1.158**	**2.415**		2
Impact on friendship	0.805	0.283	**1.469**	**2.791**		1
Impact on learning	0.778	0.360	**1.268**	**2.151**		3
Impact on leisure	0.721	0.438	**1.552**	**2.415**		4

Note: T1 = First threshold; T2 = Second threshold; T3 = Third threshold.

Table S8- Factor Loadings and Category Threshold parameters for items in the Impairment section for Young Adults, using self-report assessment.

Items	Factor Loadings	T1	T2	T3	Location (mean)	Location rank
		A little	A medium amount	A great deal		
Impact on family life	0.745	-0.422	**0.896**	**1.623**		1
Impact on friendship	0.853	0.172	**1.511**	**2.124**		2
Impact on learning	0.606	0.287	**1.390**	**2.124**		3
Impact on leisure	0.668	0.392	**1.738**	**2.352**		4

Note: T1 = First threshold; T2 = Second threshold; T3 = Third threshold.

Table S9 – Latent Class Model Fit Indexes and Entropy for each of the four latent class solutions for Irritable Mood, Outbursts and Impairment items for Adolescents and Young Adults Groups

Adolescents							
Irritable Mood							
Model	log-likelihood	resid. Df	BIC	aBIC	cAIC	likelihood-ratio	Entropy
Model 1	-1045.5966	247	2136.127	2110.761	2144.127	314.0420	-
Model 2	**981.0416**	**238**	**2057.568**	**2003.665**	**2074.568**	**184.9319**	**0.744**
Model 3	-963.4490	229	2072.934	1990.493	2098.934	149.7467	0.777
Model 4	-952.3391	220	2101.265	1990.287	2136.265	127.5270	0.919
Outbursts							
Model	log-likelihood	resid. Df	BIC	aBIC	cAIC	likelihood-ratio	Entropy
Model 1	-970.7184	265	1997.604	1965.896	2007.604	546.6402	-
Model 2	**-836.9545**	**254**	**1791.861**	**1725.274**	**1812.861**	**279.1125**	**0.787**
Model 3	-825.5969	243	1830.931	1729.465	1862.931	256.3974	0.754
Model 4	-813.5432	232	1868.608	1732.263	1911.608	232.2899	0.728
Impairment							
Model	log-likelihood	resid. Df	BIC	aBIC	cAIC	likelihood-ratio	Entropy
Model 1	-395.4972	11	813.4615	800.7783	817.4615	90.954105	-
Model 2	**-352.7521**	**6**	**756.0552**	**727.5179**	**765.0552**	**5.463912**	**0.637**
Model 3	-350.8904	1	780.4156	736.0243	794.4156	1.740428	0.605
Model 4	-350.3643	-4	807.4472	747.2018	826.4472	0.688160	0.626
Young Adults							
Irritable Mood							
Model	log-likelihood	resid. Df	BIC	aBIC	cAIC	likelihood-ratio	Entropy
Model 1	-1350.190	247	2746.964	2721.587	2754.964	396.4972	-

Model 2	**-1262.580**	**238**	**2624.152**	**2570.225**	**2641.152**	**221.2773**	**0.759**
Model 3	-1245.138	229	2641.676	2559.200	2667.676	186.3938	0.855
Model 4	-1231.755	220	2667.316	2556.291	2702.316	159.6269	0.905
Outbursts							
Model	log-likelihood	resid. Df	BIC	aBIC	cAIC	likelihood-ratio	Entropy
Model 1	-1123.1412	329	2298.690	2270.140	2307.690	490.5987	-
Model 2	**-992.4368**	**319**	**2095.511**	**2035.241**	**2114.511**	**229.1899**	**0.757**
Model 3	-983.0418	309	2134.952	2042.959	2163.952	210.3999	0.74
Model 4	-973.6904	299	2174.480	2050.766	2213.480	191.6971	0.822
Impairment							
Model	log-likelihood	resid. Df	BIC	aBIC	cAIC	likelihood-ratio	Entropy
Model 1	-481.8328	11	986.9577	974.2691	990.9577	1.286718e+02	-
Model 2	**-420.5609**	**6**	**893.5292**	**864.9798**	**902.5292**	**6.128096e+00**	**0.724**
Model 3	-417.5733	1	916.6692	872.2591	930.6692	1.529144e-01	0.549
Model 4	-417.4968	-4	945.6316	885.3606	964.6316	1.198600e-08	-0.223

Note: aBIC = adjusted Bayesian Information Criteria; BIC = Bayesian Information Criterion; cAIC = Consistent Akaike Information Criterion; df = degrees of freedom.

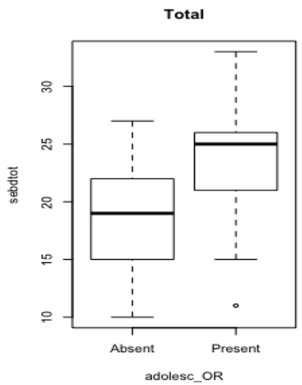
Table S10 - Linear regression comparing DMDD groups in SDQ total and impact scores

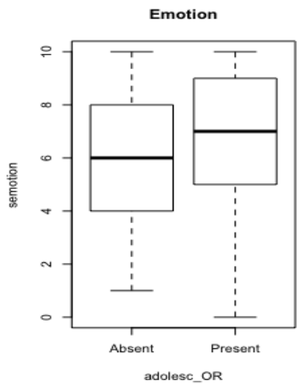
	Adolescents							Young Adults					
	Contrast with TDC			Contrast with Other DSM-5 Disorders				Contrast with TDC			Contrast with Other DSM-5 Disorders		
	MD	SE	p-value	MD	SE	p-value		MD	SE	p-value	MD	SE	p-value
SDQ: Total							**SDQ: Total**						
Other disorders	6.13	0.50	<0.001				Other disorders	5.76	0.43	<0.001			
Irritable mood only	9.68	1.54	<0.001	3.55	1.57	<0.05	Irritable mood only	6.13	0.93	<0.001	0.36	0.92	0.70
Outbursts only	8.21	1.15	<0.01	2.08	1.20	0.08	Outbursts only	7.00	1.77	<0.001	1.24	1.70	0.47
Combined	12.35	1.08	<0.001	6.22	1.13	<0.001	Combined	10.24	0.76	<0.001	4.47	0.78	<0.001
SDQ: Impact							**SDQ: Impact**						
Other disorders	1.65	0.11	<0.001				Other disorders	1.54	0.08	<0.001			
Irritable mood only	2.18	0.20	<0.001	0.53	0.20	<0.01	Irritable mood only	1.62	0.14	<0.001	0.08	0.13	0.54
Outbursts only	1.58	0.20	<0.001	-0.07	0.20	0.71	Outbursts only	1.82	0.22	<0.001	0.27	0.22	0.22
Combined	2.38	0.14	<0.001	0.73	0.14	<0.001	Combined	2.20	0.10	<0.001	0.65	0.09	<0.001

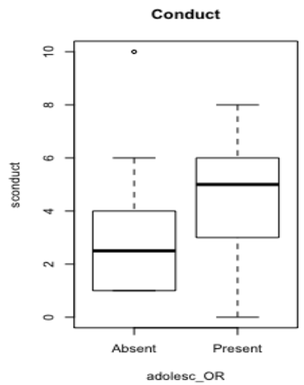
Note: MD = Mean Difference; SDQ = Strengths and Difficulties Questionnaire; TDC = Typically Developing Comparisons.

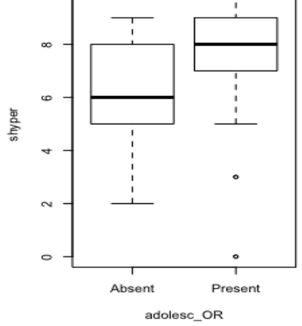
Figure S1 – Comparison of Strengths and Difficulties Questionnaire Scores between Disruptive Mood Dysregulation Disorder, defined using the OR rule, and subjects matched for psychiatric comorbidity

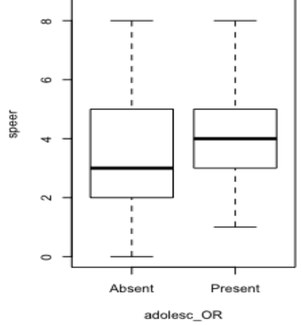
Panel A – Adolescents

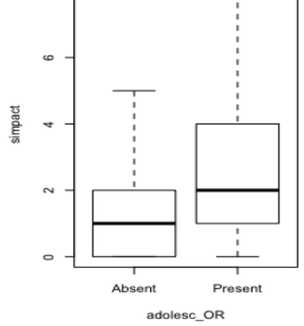




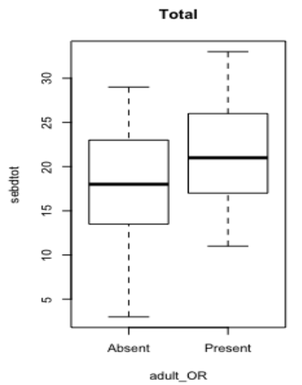


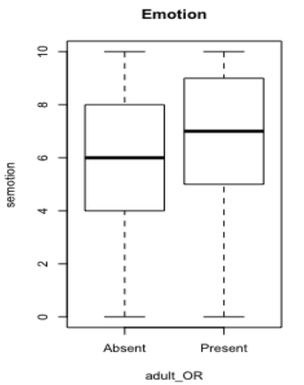


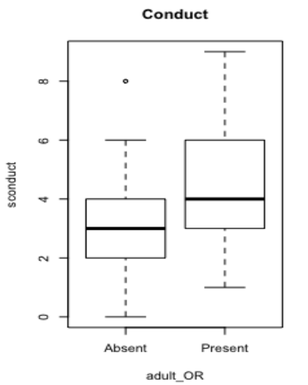


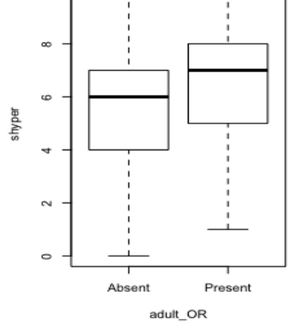


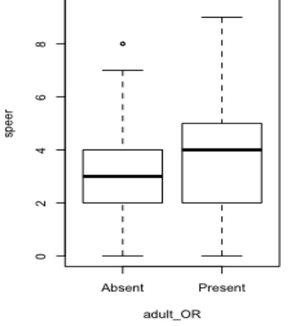
Panel B – Young Adults











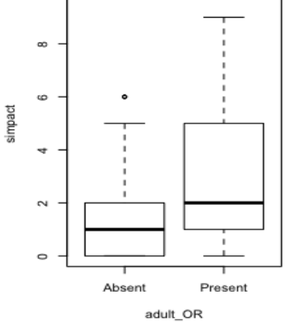
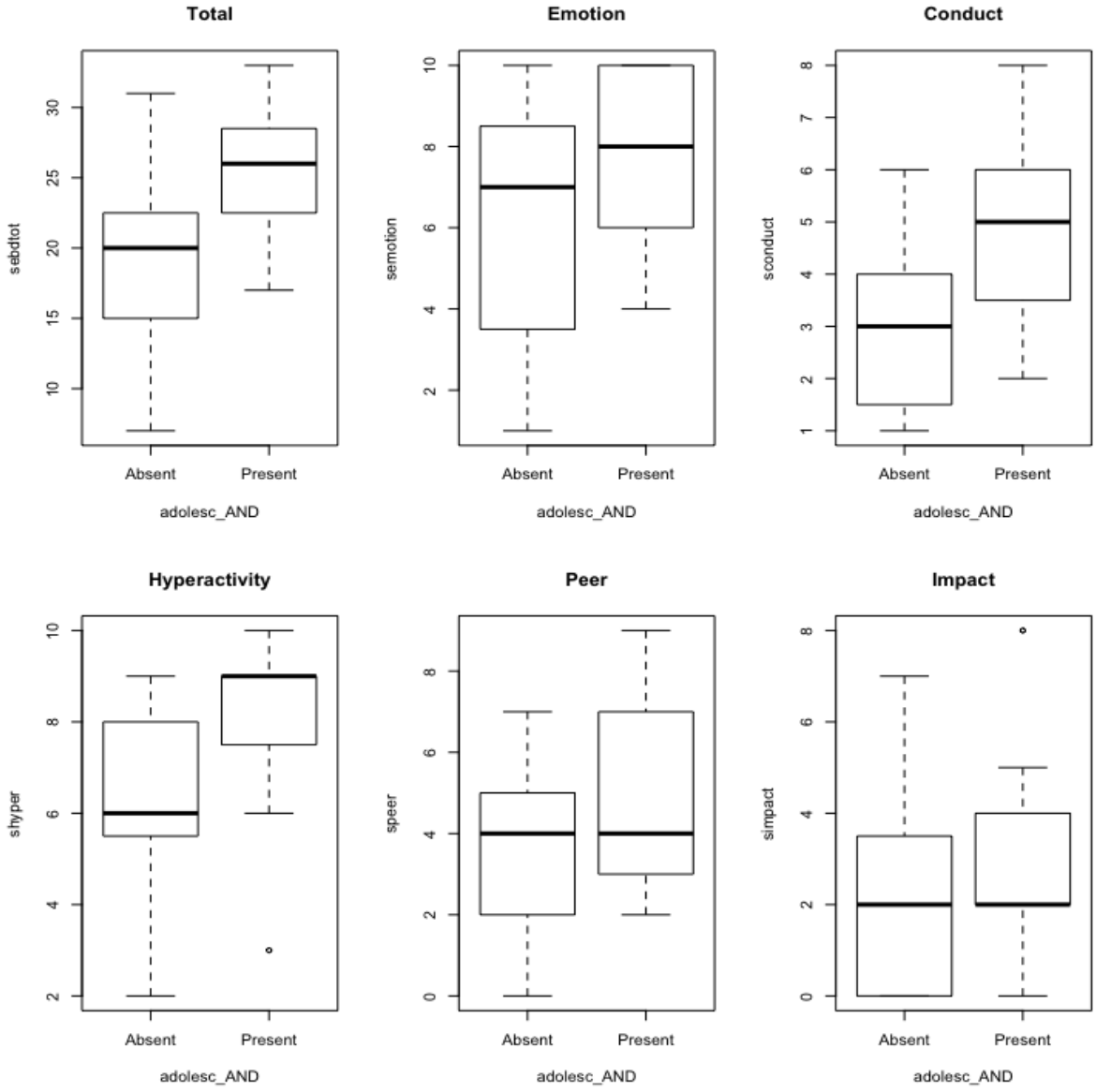
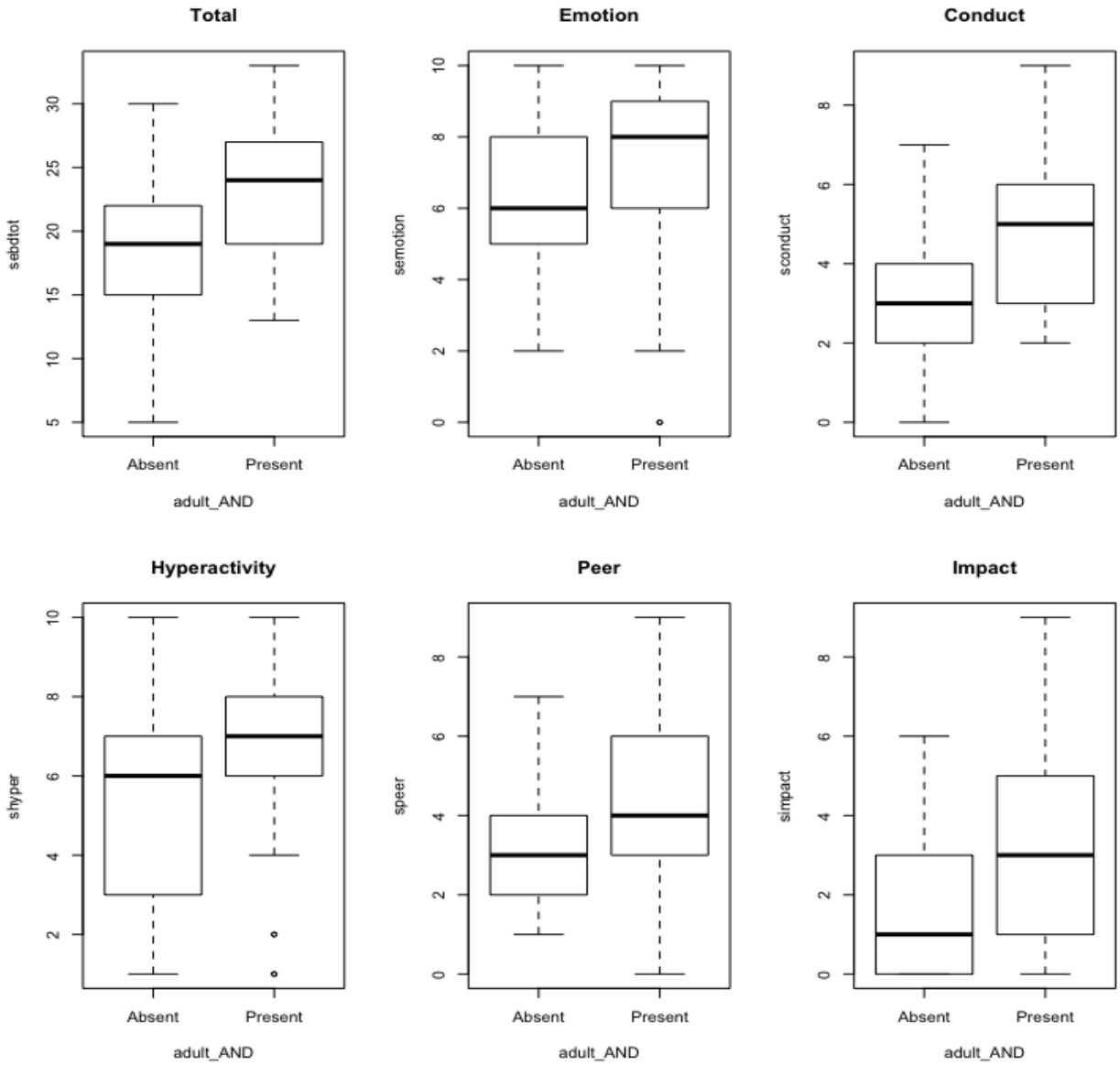


Figure S2 – Comparison of Strengths and Difficulties Questionnaire Scores between Disruptive Mood Dysregulation Disorder, defined using the AND rule, and subjects matched for psychiatric comorbidity

Panel A – Adolescents



Panel B – Young Adults


Total
sebdtot
5
10
15
20
25
30
Absent
Present
adult_AND
Emotion
semotion
0
2
4
6
8
10
Absent
Present
adult_AND
Conduct
sconduct
0
2
4
6
8
Absent
Present
adult_AND
Hyperactivity
shyper
2
4
6
8
10
Absent
Present
adult_AND
Peer
speer
0
2
4
6
8
Absent
Present
adult_AND
Impact
simpact
0
2
4
6
8
Absent
Present
adult_AND

Final considerations

Irritability is a clinically relevant phenomenon, transdiagnostic in nature and dimensionally distributed in the population [44]. It is associated with other dimensions of psychopathology, but dissociable from them. Furthermore, it seems to follow its own developmental course with considerable stability across time. Additionally, irritability predicts future psychopathology and impairment independently of the presence of other disorders**.** The boundaries between irritability and internalizing and externalizing syndromes are poorly defined, the thresholds are not empirically driven and there is a lack of an integrative perspective on development linking irritability in childhood with irritability in adulthood. This thesis aimed to add new understandings to the normal/abnormal spectrum of this clinical phenomenology and expanded the study of DMDD diagnosis to the adulthood field.

Our study presented two articles that have in common the intention of improving and expanding the diagnosis of DMDD. The studies are aligned to the vision that Kendler proposes for the advancement of research in this area, through successively disassembling and reassembling the empirical evidence [1,45–47]. He advocates that psychiatry must move beyond a prescientific "battle of paradigms" to embrace complexity and support empirically rigorous and pluralistic explanatory models. Studies that investigate this gap between innovative and classic approaches are essential. They intend to provide clinical sense to the investigated mental processes and direct research toward the mechanisms, prioritizing those most likely to inform the psychopathology of diseases.

Our contributions can be summarized as follows: One study provided important information to guide a data-driven and clinically oriented operationalized revision of the diagnostic criteria for DMDD in preadolescents

(Article #1). The other study advanced the empirically driven clinical threshold for DMDD to adolescents and young adults. Currently, there is no diagnostic criteria category for adults with chronic and severe irritability (Article #2).

Article #1 is of particular interest because it gives us a clear clinical guidance on how to identify a group of individuals at the preadolescent developmental stage on the irritability continuum. The article was built upon a large representative population sample and it implemented assessment methods that could mimic clinical evaluation in the real world, as far as possible in an epidemiological investigation. *First,* we utilized Confirmatory Factor Analysis and fixed up a list of behaviors and a threshold for each that indicated when each behavior was considered a problem. *Second*, Latent Class Analysis and Receiver Operating Curves Analysis suggested a syndromic threshold for the combination of such behaviors. *Third*, our data supported an OR rule when combining irritable mood and outbursts, rather than the AND rule currently found in the manual. *Finally*, our results sustained the requirement of at least two settings for the diagnosis, just as seen in DSM-5.

Although most psychiatric conditions present symptoms of irritability and research on irritability in childhood is increasing, the field in adults is still incipient. Article #2 studied participants that are part of the Brazilian High-Risk Cohort Study for Mental Conditions (BHRCS, 3rd wave) to provide data-driven thresholds for DMDD in adolescents and young adults and explore adaptations of DMDD diagnosis to adulthood. *First*, measurement invariance analysis showed that thresholds for measuring irritability symptoms changed from adolescence to adulthood. *Second*, we used a previously validated pipeline to assess the most appropriate symptomatic and syndromic thresholds for the diagnosis and the best way to perform clinical operationalization using AND/OR rules for combining irritable mood and

outbursts. Finally, we presented the prevalence rates for the combination of distinct DMDD diagnostic rules and presented a "developmental follow-back analysis" and an "experience momentary assessment".

In their everyday practices, clinicians will remain using the fuzzy constructs narratively depicted in DSM-5 and ICD-11. Nevertheless, the feeling is that we are advancing and refining this field. Regardless of diagnosis, both works are densely informative ~~for~~, describing at which level reports of irritable mood, outbursts, and their influence on domains of life are clinically significant. We were able to demonstrate that an objective-based classification may be operationalized in a clinically useful way. The thesis concludes that the adequacy of psychiatric diagnostic systems must be empirically tested, so that its limitations and/or potentialities can be properly addressed. Future practice will involve the consideration of multiple dimensions of symptoms, biology, and experience to triangulate an individual's clinical status and to determine a personalized treatment and prognosis.

We conclude this thesis acknowledging both, our small but significant contributions to the research in the field of DMDD diagnostic criteria, and the several open questions that are left to be answered. Although we have advanced in the field, the following challenges still remain: 1) investigate course, family history, treatment response, and other external validators needed to demonstrate the validity of the operationalized syndrome; 2) disentangle the role of heterotypic continuity on the emergence of irritability after childhood; 3) validate and explore the existence of the DMDD diagnosis in independent samples with further scrutiny than our study has provided until the present moment. The consideration of neurobiological features, for instance, with genetic and neuroimaging approaches, may further clarify the nature of this syndrome. In this sense, more research is needed to have a more comprehensive

understanding of the scope of the phenomenon and how our society should deal with it.

References

1. Kendler KS. An historical framework for psychiatric nosology. *Psychol Med*. 2009;39(12):1935-1941. doi:10.1017/S0033291709005753
2. AMERICAN PSYCHIATRIC ASSOCIATION. *Diagnostic and Statistical Manual of Mental Disorders*. Fifth Edition (DSM-V).; 2013.
3. ICD-11 - ICD-11 for Mortality and Morbidity Statistics. Accessed May 23, 2021. https://icd.who.int/browse11/l-m/en
4. KENDELL RE. Diagnostic and Statistical Manual of Mental Disorders, 3rd ed., revised (DSM-III-R). *American Journal of Psychiatry*. Published online April 1, 2006. doi:10.1176/ajp.145.10.1301
5. Insel T, Cuthbert B, Garvey M, et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry*. 2010;167(7):748-751. doi: 10.1176/appi.ajp.2010.09091379
6. Salum Junior GA, Gadelha A, Polanczyk GV, Miguel EC, Rohde LAP. Diagnostic operationalization and phenomenological heterogeneity in psychiatry: the case of attention deficit hyperactivity disorder. Published online 2018. Accessed May 23, 2021. https://lume.ufrgs.br/handle/10183/199782
7. Kapur S, Phillips AG, Insel TR. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol Psychiatry*. 2012;17(12):1174-1179. doi:10.1038/mp.2012.105
8. Regier DA, Narrow WE, Clarke DE, et al. DSM-5 field trials in the United States and Canada, Part II: test-retest reliability of selected categorical diagnoses. *Am J Psychiatry*. 2013;170(1):59-70. doi: 10.1176/appi.ajp.2012.12070999
9. Kraemer HC, Kupfer DJ, Clarke DE, Narrow WE, Regier DA. DSM-5: how reliable is reliable enough? *Am J Psychiatry*. 2012;169(1):13-15. doi: 10.1176/appi.ajp.2011.11010050
10. Kendler KS, Zachar P, Craver C. What kinds of things are psychiatric disorders? *Psychol Med*. 2011;41(6):1143-1150. doi:10.1017/S0033291710001844
11. Stringaris A, Taylor E. *Disruptive Mood: Irritability in Children and Adolescents*. Oxford University Press Accessed April 26, 2021. https://oxfordmedicine.com/view/10.1093/med/9780199674541.001.0001/med-9780199674541
12. Althoff RR, Kuny-Slock AV, Verhulst FC, Hudziak JJ, van der Ende J. Classes of oppositional-defiant behavior: concurrent and predictive validity. *J Child Psychol Psychiatry*. 2014;55(10):1162-1171. doi:10.1111/jcpp.12233
13. Stringaris A. Irritability in children and adolescents: a challenge for DSM-5. *Eur Child Adolesc Psychiatry*. 2011;20(2):61-66. doi:10.1007/s00787-010-0150-4
14. Deveney CM, Connolly ME, Haring CT, et al. Neural mechanisms of frustration in chronically irritable children. *Am J Psychiatry*. 2013;170(10):1186-1194. doi: 10.1176/appi.ajp.2013.12070917
15. Leibenluft E, Cohen P, Gorrindo T, Brook JS, Pine DS. Chronic versus episodic irritability in youth: a community-based, longitudinal study of clinical and diagnostic associations. *J Child Adolesc Psychopharmacol*. 2006;16(4):456-466. doi:10.1089/cap.2006.16.456
16. Copeland WE, Brotman MA, Costello EJ. Normative Irritability in Youth: Developmental Findings from the Great Smoky Mountains Study. *J Am Acad Child Adolesc Psychiatry*. 2015;54(8):635-642. doi: 10.1016/j.jaac.2015.05.008

17. Knopf A. Irritability in youth caused by reward-, threat-system dysfunction: New paradigm. *The Brown University Child and Adolescent Behavior Letter*. 2017;33(3):1-7. doi: https://doi.org/10.1002/cbl.30197
18. Roberson-Nay R, Leibenluft E, Brotman MA, et al. Longitudinal Stability of Genetic and Environmental Influences on Irritability: From Childhood to Young Adulthood. *Am J Psychiatry*. 2015;172(7):657-664. doi: 10.1176/appi.ajp.2015.14040509
19. Copeland WE, Shanahan L, Egger H, Angold A, Costello EJ. Adult Diagnostic and Functional Outcomes of DSM-5 Disruptive Mood Dysregulation Disorder. *AJP*. 2014;171(6):668-674. doi:10.1176/appi.ajp.2014.13091213
20. Dougherty LR, Schwartz KTG, Kryza-Lacombe M, Weisberg J, Spechler PA, Wiggins JL. Preschool- and School-Age Irritability Predict Reward-Related Brain Function. *J Am Acad Child Adolesc Psychiatry*. 2018;57(6):407-417.e2. doi:10.1016/j.jaac.2018.03.012
21. Smith JD, Wakschlag L, Krogh-Jespersen S, et al. Dysregulated Irritability as a Window on Young Children's Psychiatric Risk: Transdiagnostic Effects via the Family Check-Up. *Development and Psychopathology*. 2019;31(5):1887-1899. doi:10.1017/S0954579419000816
22. Toohey MJ, DiGiuseppe R. Defining and measuring irritability: Construct clarification and differentiation. *Clinical Psychology Review*. 2017;53:93-108. doi:10.1016/j.cpr.2017.01.009
23. Cicchetti D. Developmental Psychopathology. In: *The Handbook of Life-Span Development*. American Cancer Society; 2010. doi:10.1002/9780470880166.hlsd002014
24. Brotman MA, Schmajuk M, Rich BA, et al. Prevalence, Clinical Correlates, and Longitudinal Course of Severe Mood Dysregulation in Children. *Biological Psychiatry*. 2006;60(9):991-997. doi:10.1016/j.biopsych.2006.08.042
25. Stringaris A, Maughan B, Copeland WS, Costello EJ, Angold A. Irritable mood as a symptom of depression in youth: prevalence, developmental, and clinical correlates in the Great Smoky Mountains Study. *J Am Acad Child Adolesc Psychiatry*. 2013;52(8):831-840. doi:10.1016/j.jaac.2013.05.017
26. Rowe R, Costello EJ, Angold A, Copeland WE, Maughan B. Developmental pathways in oppositional defiant disorder and conduct disorder. *J Abnorm Psychol*. 2010;119(4):726-738. doi:10.1037/a0020798
27. Stringaris A, Goodman R. Longitudinal outcome of youth oppositionality: irritable, headstrong, and hurtful behaviors have distinctive predictions. *J Am Acad Child Adolesc Psychiatry*. 2009;48(4):404-412. doi:10.1097/CHI.0b013e3181984f30
28. Olfson M, King M, Schoenbaum M. Treatment of Young People With Antipsychotic Medications in the United States. *JAMA Psychiatry*. 2015;72(9):867-874. doi:10.1001/jamapsychiatry.2015.0500
29. Deveney CM, Hommer RE, Reeves E, et al. A prospective study of severe irritability in youths: 2- and 4-year follow-up. *Depress Anxiety*. 2015;32(5):364-372. doi:10.1002/da.22336
30. Leibenluft E, Charney DS, Towbin KE, Bhangoo RK, Pine DS. Defining clinical phenotypes of juvenile mania. *Am J Psychiatry*. 2003;160(3):430-437. doi:10.1176/appi.ajp.160.3.430
31. Towbin K, Axelson D, Leibenluft E, Birmaher B. Differentiating bipolar disorder-not otherwise specified and severe mood dysregulation. *J Am Acad Child Adolesc Psychiatry*. 2013;52(5):466-481. doi:10.1016/j.jaac.2013.02.006
32. Stringaris A, Cohen P, Pine DS, Leibenluft E. Adult outcomes of youth

irritability: a 20-year prospective community-based study. *Am J Psychiatry*. 2009;166(9):1048-1054. doi:10.1176/appi.ajp.2009.08121849
33. Wakschlag LS, Briggs-Gowan MJ, Carter AS, et al. A developmental framework for distinguishing disruptive behavior from normative misbehavior in preschool children. *J Child Psychol Psychiatry*. 2007;48(10):976-987. doi:10.1111/j.1469-7610.2007. 01786.x
34. Copeland WE, Shanahan L, Costello EJ, Angold A. Childhood and adolescent psychiatric disorders as predictors of young adult disorders. *Arch Gen Psychiatry*. 2009;66(7):764-772. doi:10.1001/archgenpsychiatry.2009.85
35. Lochman JE, Evans SC, Burke JD, et al. An empirically based alternative to DSM-5's disruptive mood dysregulation disorder for ICD-11. *World Psychiatry*. 2015;14(1):30-33. doi:10.1002/wps.20176
36. Mayes SD, Waxmonsky JD, Calhoun SL, Bixler EO. Disruptive Mood Dysregulation Disorder Symptoms and Association with Oppositional Defiant and Other Disorders in a General Population Child Sample. *J Child Adolesc Psychopharmacol*. 2016;26(2):101-106. doi:10.1089/cap.2015.0074
37. Dougherty LR, Smith VC, Bufferd SJ, et al. DSM-5 disruptive mood dysregulation disorder: correlates and predictors in young children. *Psychol Med*. 2014;44(11):2339-2350. doi:10.1017/S0033291713003115
38. Copeland WE, Angold A, Costello EJ, Egger H. Prevalence, comorbidity, and correlates of DSM-5 proposed disruptive mood dysregulation disorder. *Am J Psychiatry*. 2013;170(2):173-179. doi:10.1176/appi.ajp.2012.12010132
39. Holtmann M, Bölte S, Goth K, et al. Prevalence of the Child Behavior Checklist-pediatric bipolar disorder phenotype in a German general population sample. *Bipolar Disord*. 2007;9(8):895-900. doi:10.1111/j.1399-5618.2007.00463.x
40. Hudziak JJ, Althoff RR, Derks EM, Faraone SV, Boomsma DI. Prevalence and genetic architecture of Child Behavior Checklist-juvenile bipolar disorder. *Biol Psychiatry*. 2005;58(7):562-568. doi:10.1016/j.biopsych.2005.03.024
41. Volk HE, Todd RD. Does the Child Behavior Checklist juvenile bipolar disorder phenotype identify bipolar disorder? *Biol Psychiatry*. 2007;62(2):115-120. doi: 10.1016/j.biopsych.2006.05.036
42. Holtmann M, Bölte S, Poustka F. Rapid increase in rates of bipolar diagnosis in youth: "true" bipolarity or misdiagnosed severe disruptive behavior disorders? *Arch Gen Psychiatry*. 2008;65(4):477. doi:10.1001/archpsyc.65.4.477
43. Brotman MA, Kircanski K, Stringaris A, Pine DS, Leibenluft E. Irritability in Youths: A Translational Model. *Am J Psychiatry*. 2017;174(6):520-532. doi:10.1176/appi.ajp.2016.16070839
44. Wakschlag LS, Perlman SB, Blair RJ, Leibenluft E, Briggs-Gowan MJ, Pine DS. The Neurodevelopmental Basis of Early Childhood Disruptive Behavior: Irritable and Callous Phenotypes as Exemplars. *Am J Psychiatry*. 2018;175(2):114-130. doi:10.1176/appi.ajp.2017.17010045
45. Kendler KS. Explanatory models for psychiatric illness. *Am J Psychiatry*. 2008;165(6):695-702. doi:10.1176/appi.ajp.2008.07071061
46. Kendler KS. The dappled nature of causes of psychiatric illness: replacing the organic-functional/hardware-software dichotomy with empirically based pluralism. *Mol Psychiatry*. 2012;17(4):377-388. doi:10.1038/mp.2011.182
47. Kendler KS, First MB. Alternative futures for the DSM revision process: iteration v. paradigm shift. *Br J Psychiatry*. 2010;197(4):263-265. doi:10.1192/bjp.bp.109.076794

Appendix

In the following pages, we present the remaining publications that occurred during the doctorate of the PhD candidate, but that are either not directly related to the topic of the thesis, or that were not leaded by the PhD candidate.

Appendix #1

Published in the Revista Brasileira de Psiquiatria

Specific and social fears in childhood and early adolescence: separating normative fears from problem indicators

Paola Paganella Laporte[1,2], Pedro Mario Pan[1,3], Mauricio Scopel Hoffmann[1,2], Luis Augusto Rohde[1,2,4], Euripedes Constantino Miguel[1,2,4], Daniel Samuel Pine[5], Gisele Gus Manfro[1,2], Giovanni Abrahão Salum[1,2]

[1] National Institute of Developmental Psychiatry for Children and Adolescents - CNPq, Brazil

[2] Universidade Federal do Rio Grande do Sul, Brazil

[3] Universidade Federal de São Paulo, Brazil

[4] Universidade de São Paulo, Brazil

[5] National Institute of Mental Health Intramural Research Program, Bethesda, MD, USA

Conflict of interest statement

Dr Laporte and Hoffmann reports no biomedical financial interests or potential conflicts of interest. Dr. Pan reports personal fees from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Dr. Manfro and Dr. Miguel are in receipt of a senior research CNPq scholarship (304829/2013-7 and 302463/2011-9, respectively). Dr. Rohde was on the speakers' bureau and/or acted as consultant for Eli-Lilly, Janssen-Cilag, Novartis and Shire in the last three years. The ADHD and Juvenile Bipolar Disorder Outpatient Programs chaired by Dr. Rohde received unrestricted educational and research support from the following pharmaceutical companies in the last three years: Eli-Lilly, Janssen-Cilag, Novartis, Shire. Dr. Rohde received travel grants from Shire to take part of the 2014 APA and 2015 WFADHD congresses. He also receives authorship royalties from Oxford University Press and Artmed. Dr. Salum is in receipt of a post-doctoral scholarship from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES; Brazilian Federal Agency for Support and Evaluation of postgraduate education) and Fundação de Amparo a Pesquisa do Estado do Rio Grande do Sul (FAPERGS; Research Support Foundation from the State of Rio Grande do Sul).

Word count

Abstract: 331

Article body: 6759

Tables: (3 tables, 846 words)

Figures: (1 figure, 12 words)

Supplementary material: (4 tables, 1311 words)

Funding source:

Abstract

Background: Specific and social fears are common in children and adolescents. Whereas in the majority of children they are transient and non-problematic, in some they are intense and frequent, causing significant distress. The aim of this study is to investigate from a list of eighteen fears, which of them are likely to represent a normative fear and which are likely to represent a clinically significant problem.

Methods: We investigated children and adolescents aged 6-14 participating in the 'High Risk Cohort Study for Psychiatric Disorders' (n=2,512). Parent reports of eighteen fears (12 specific and 6 social) were investigated with the Specific Phobia and Social Phobia sections of the Development and Well-Being Behavior Assessment (DAWBA) questionnaire, which rates the presence of each fear across three levels: 'no', 'a little' and 'a lot' and provide diagnostic assessment. We used two analytical approaches: Confirmatory Factor Analysis (CFA)/Item Response Theory (IRT) and Non-parametric Receive Operator Curve (ROC).

Results: Both CFA/IRT and ROC analysis converge to demonstrate that generally social fears are more likely to indicate problems and psychiatric diagnosis than specific fears. Among specific fears, most of them were found to be normative when presented as little, with the exception of situational fears and fear of 'people who look unusual'. All specific fears indicate problems when presented as 'a lot'. In addition, the situational 'fear of toilets' and 'people who look unusual' were found to be highly indicative of specific phobia diagnosis. Among social fears, fears 'non-restricted to performance' and fear of writing in front of others indicate problems when presented as 'a little'. All social fears are likely to indicate problems and to be highly indicative of social phobia diagnosis when presented as 'a lot'.

Conclusion: Our findings might help clinicians and researchers to determine the boundaries that separate normative fears from problem indicators in children and early adolescents and indicate a marked differential severity threshold for specific and social fears.

Keywords: Developmental psychopathology; child/adolescent; phobia/phobic disorders; anxiety/anxiety disorders; specific/social fears.

Introduction

Fear can be defined broadly as a negative emotional state, which is triggered by the presence of a stimulus that has the potential to cause harm. It is an adaptive and essential emotion for survival. However, when intensity, duration, and/or frequency are not proportional to the eliciting threat, and thereby cause interference or excessive distress, such fears may indicate the presence of a mental disorder requiring treatment (*The Diagnostic and statistical manual of mental disorders* 2013). Children can fear a variety of things and situations, but more data are needed to identify fears most likely to indicate the presence of a significant clinical problem and the presence of specific and social phobia diagnosis.

Normative fears are observed in most children and adolescents and are typically transient. The prevalence of specific and social fears, considering both its mild and severe forms, reaches almost all children and adolescents (Benjet et al. 2012a; Ollendick e King 1994). In contrast, specific phobia and social phobia are characterized by one or more persistent fears, that causes distress and impairment to the child's life (Salum et al. 2013a; American Psychiatric Association 2000). The prevalence of specific and social phobia in children and adolescents was found to be 2.9% and 0.3% respectively (Salum et al. 2014). Therefore, whereas the presence of fears is typically normative, only a minority of children is significantly impaired by fears requiring clinical attention.

Previous studies investigated whereas a specific fear may indicate a disorder according to the number of feared objects/situations. These studies found that the number of fears was positively associated with a higher probability of meeting a diagnosis of specific phobia (Benjet et al. 2012b), meeting diagnosis of another anxiety disorder or having another psychiatric comorbidity (Burstein et al. 2012a; Benjet et al. 2012c). With respect to social fears, studies also investigated the role of the number of feared social situations. They found that the number of social fears is significantly associated with higher comorbidity rates, functional impairment and prevalence of lifetime treatment (Ruscio et al. 2008a). In addition, adolescents with generalized social phobia (i.e., those who fear most of the social situations) had earlier age of onset (Wittchen, Stein, e Kessler 1999a) and experienced a higher degree of clinical severity as compared to adolescents with *non-generalized* social phobia (Burstein et al. 2011a).

Therefore, so far the literature has been limited to investigate how the number of fear

situations and objects might inform about clinically relevant situations, but few studies have investigated the type of fear as an indicator of clinical severity. Among existing methods to discriminate the severity of different types of symptoms, two received particularly little attention in the fear literature: Confirmatory Factor Analysis (CFA) / Item Response Theory (IRT) and Signal Detection Theory (SDT). CFA with binary outcomes and two parameter normal ogive IRT are equivalent to each other (Timothy A. Brown PsyD 2006, 396) . They assume that items (i.e. symptoms) are endorsed by subjects (i.e. reported) as a function of their severity on a specific latent trait (i.e. how much fear they have). These techniques allows scaling items and people on the same underlying dimensional continuum (Reise e Waller 2009). Therefore, individual items are assigned to a severity score, and an individual score on the dimension can be used as an estimate of his/her overall severity of fears (Wakschlag et al. 2012a). In contrast, SDT aims to quantify the ability to discriminate between stimulus and random patterns that distract the true information (T. H. Wilmshurst 1990). SDT has wide applications in biomedical sciences including the use of ROC (Margolis et al. 2002). The ROC represents the relationship between sensitivity and specificity by plotting the true positive rate (such as phobia diagnosis) against the false positive rate at various threshold settings to detect a significant medical problem or diagnosis (Kessler et al. 2013a).

Previous literature is limited in a number of important ways. First, no previous study used modern psychometric analysis (CFA/IRT) and signal detection theory (diagnostic prediction) to inform about the thresholds of normative fear. *Second,* most of the current literature focuses on adolescent and adult populations (Blanco et al. 2011; Burstein et al. 2012b; Burstein et al. 2011b; Iza et al. 2014; Polo et al. 2011; Ruscio et al. 2008b; Wittchen, Stein, e Kessler 1999b; Kessler et al. 2013b), but there are few studies including children and early adolescents (Muris et al. 2000). *Finally,* the majority of studies are limited to small samples sizes or selected by a clinical condition, which limits the ability to use IRT and signal detection theory, respectively.

Here we address these issues by investigating 6-14 years of age children using CFA/IRT for problem prediction and ROC analysis for diagnostic prediction in a large community non-referred sample. These techniques were applied to investigate from a list of

eighteen fears, which are more likely to represent a normative fear and which ones are more likely to represent a clinically significant problem.

Methods

Sample description

The sample was obtained from a large community school-based study - the 'High Risk Study for Psychiatric Disorders' (Salum et al. 2014). Further details can be found elsewhere (Salum et al. 2014). Briefly, after a screening phase, 2,512 subjects were selected for further assessment by a simple randomization procedure that select a high risk sample (n=1,554) using a risk priorization procedure that consisted of selecting individuals with high family loading of symptoms and ongoing psychiatric symptoms and a random-selection subgroup (n=958). The study was previously approved by the Ethics Committee of the University of São Paulo and written informed consent was obtained from parents of all participants.

Instruments

Specific and Social Fears Assessment

Specific and Social Fears were investigated using 18 questions posed to a parental informant from the Specific Phobia and Social Phobia sections of the Development and Well-Being Assessment (DAWBA) (Goodman et al. 2000) instrument. The specific phobia session consists of 12 questions about specific fears: "Animals", "Storms, Thunder, Heights", "Dark", "Loud noises", "Blood, injection, injury", "Dentists, Doctors", "Vomiting, Choking, Diseases", "Types of transport", "Enclosed spaces", "Toilets", "People who look unusual", "Monsters" and "Other things". The social phobia section comprised 6 questions about social fears: "Meeting new people", "Meeting a lot of people", "Eating in front of others", "Speaking in class", "Reading aloud in front of others" and "Writing in front of others". For both sections, all questions have three response categories: "No", "A little", "A lot". In social phobia session, we removed the skipping rule of first question ("Does child particularly fear or avoid social situation?"), in order to have the six social fears for all participants.

Diagnostic Assessment

Psychiatric diagnosis was also investigated using the Brazilian Portuguese version of the DAWBA (Fleitlich-Bilyk e Goodman 2004) answered by the child's main caregiver. The DAWBA was performed by lay interviewers and had structured answers and recorded verbatim responses of any reported problem. Psychiatrists confirmed, refuted or altered initially computerized diagnosis after evaluating carefully the structured answers and the verbatim responses. All questions are based on the DSM-IV diagnostic criteria that for social and specific phobia diagnosis include: (b) 'phobic stimulus almost invariably provokes an immediate anxiety response'; (c) 'fear is excessive or unreasonable'; (d) 'phobic situation is avoided or endured with intense anxiety or distress'; (e) 'avoidance, anxious anticipation or distress in the feared situation interferes with the person's life'; (f) 'in individuals under age 18 years, the duration is at least 6 months'; (g) 'fear are not better accounted for by another mental disorder' (*Diagnostic and Statistical Manual of Mental Disorder DSM-IV*, [s.d.]).

The DAWBA is reliable, well suited for epidemiological research and has been translated into 22 languages. The Brazilian Portuguese version shows appropriate psychometric proprieties and high inter-rater reliability (Salum et al. 2014) and the reliability for emotional disorders diagnosis was good (kappa = 0.84) (Salum et al. 2014). Based on the information obtained during the interviews, diagnoses were assigned by one of nine trained psychiatrists using a computerized platform. These psychiatrists were trained by attending several meetings led by a senior child psychiatrist with significant experience rating the DAWBA.

Statistical Analysis

Confirmatory Factor Analysis and Item Response Theory

The eighteen fear items were included in a CFA testing unidimensional (all itens loading into a single factor), correlated (6 social items loading into a social latent factor; 12 specific items loading into a specific latent factor) and bifactor models (all items loading into a 'general factor' and the residuals not explained by the general factor loading into a two group factors – 'specific' and 'social'). CFA models were used to estimate factor loadings (λ) and item category thresholds (τ). In addition we also performed unidimensional IRT analysis using

Graded Response Model (Samejima F 1970, 35:139–139) to estimate GRM item parameters reflecting the item discrimination or slope (a) and item difficulty (b) for each item.

Item discrimination parameters from IRT are analogous to factor loadings from CFA and indicate how well the item discriminates different severity levels or represent the strength of the relationship between latent trait and item responses. Item difficulty parameters from IRT are analogous to category thresholds from CFA, in the way they are used to indicate the expected value of the latent factor at which there is a 50% probability of endorsing a given category of higher. The category thresholds roughly indicate the severity level at which the transition from one response category to the next is likely to happen (e.g., from 'No' to 'A little' or higher and from 'A little' to 'A lot'). Higher thresholds indicate that, to be severe, a given item must be endorsed at a given response category or higher. The mean threshold for each item was computed as the item location on the severity continuum that included all items. In this sample values of the latent trait at or above the 95^{th} of the sample distribution were chosen to define a 'problem indicator' as suggested by Wakschlag et al (Wakschlag et al. 2012b), which represents a b of 1.53 in IRT metric and a category thresholds of 0.78 in CFA metric for the bifactor model (general factor). A problem indicator was defined as those items meeting problem indicator criteria in both CFA and IRT analysis.

For all CFA models we used delta parameterization and weighted least-square parameters using a diagonal weight matrix with standard errors and with mean and variance-adjusted chi-square test statistics (WLSMV) estimators. Goodness of fit was assessed using four indices: Chi Square Test of model fit, Tucker Lewis Index (TLI), Comparative Fit Index (CFI), and Root Mean Square Error of Approximation (RMSEA). TLI and CFI values > 0.95 and RMSEA values < 0.06 represent good-to-excellent model fit (Jackson, Gillaspy, e Purc-Stephenson 2009). The comparison between these three nested models was tested using Chi-Square for Difference using DIFFTEST option from MPLUS 7.3 software (Muthén, B.O. 1998). To test the invariance of the best model parameters with respect to age and gender we conducted a multi-group CFA. The models were tested in a hierarchical form, i.e., a model was compared to a less restrictive previous model in order to test configural and scalar invariance. To assume measurement invariance, differences between the CFI of configural and scalar models should not be greater than 0.01 ($\Delta CFI<0.01$). IRT model was performed

using the marginal maximum likelihood estimator implemented in 'ltm' package (Rizopoulos, D 2006).

Signal Detection Theory

To analyze associations between each specific and social fear with specific phobia and social phobia diagnosis we used non-parametric ROC available in Stata Software version 12.0. The ROC was used to estimate the area under the curve (AUC), sensitivity, specificity, positive and negative likelihood ratio. The AUC is the probability of correctly identifying cases of specific and social phobia from each fear of DAWBA questionnaire. The AUC is a function of sensitivity (SN; the percentage of true cases correctly classified) and specificity (SP; the percentage of true non-cases correctly classified), which are considered the fundamental parameters of agreement. AUC scores between 0.5 and 1.0 are considered as slight (AUC=0.5-0.6), fair (AUC=0.6-0.7), moderate (AUC=0.7-0.8), substantial (AUC=0.8-0.9) and almost perfect (AUC ≥0.9) (Landis and Koch, 1977). We also report measures of operating characteristics: LR positive [(LR+); SN/(1-SP)] and LR negative [LR-;(1-SN)/SP]. LR+ assesses the relative proportions of screened positives versus confirmed as cases (LR+) or non-cases (LR-). LR+ values ≥ 5 and LR values ≤ 0.2 are generally considered useful, whereas LR+ values ≥ 10 and LR values ≤ 0.1 are considered sufficient to rule in/out diagnoses (Kessler et al. 2013b; R. Brian Haynes et al. 2005).

Results

Our data suggests that fears are essentially normative and extremely prevalent in children and adolescents. Among the randomly selected subjects 86.5% (n= 829) presented at least one mild specific fear and 32.1% (n=308) presented at least one mild social fear. Specific fears were more common than social fears (the mean prevalence of specific fears answered as 'a little or higher' was 26% and the mean prevalence of social fears answered as 'a little or higher' was 11.7%) (Supplemental Table 1). Supplemental Table 2 shows the associations of age and gender in specific and social fears.

Confirmatory Factor Analysis and Item Response Theory

The bifactor model with 1 general factor ('fear') and 2 specific factors ('specific' and 'social') presented the best fit to the data (Supplemental Table 3). In addition, measurement invariance was demonstrated across age groups and gender (Supplemental Table 4**)**. Factor loadings and category thresholds for the best-fitting CFA model as well as discrimination and severity parameters from the unidimensional IRT model are depicted in Table 1.

An inspection of both CFA thresholds and IRT discrimination parameters converge to demonstrate that generally specific fears are more likely to be normative than social fears. For example, the mean of the severity parameter for specific fears was 1.77 (0.89 – 2.83) if compared to 2.57 (1.74 - 3.39) for social fears in IRT metric. The least severe specific fear ('blood, injection, injury') has a location of 0.86 whereas the least severe social fear ('reading in front of others') has a location of 1.74 (Table 1).

Among specific fears, animals, natural environments, blood, injection, injury and some fears classified as others ('monsters' and 'vomiting, choking and diseases'), were found to be normative when classified as a little. Situational fears ('types of transport', 'enclosed spaces' and 'toilets') and fear of 'people who look unusual' do indicate problems even when presented as a little. All specific fears indicate problems when presented as 'a lot'.

Among social fears, those classified as performance only ('speaking in class' and 'reading in front of others') were found to be normative when presented as little. Fears non-restricted to performance ('meeting new people', 'meeting a lot of people' and 'eating in front of others') and fear of 'writing in front of others' do indicate problems even when presented as a little. As in specific fears, all social fears indicate problems when presented as 'a lot'.

As an example, figure 1 depicts item response function curves for one item classified as normative (figure 1, panel A) and one item classified as a problem indicator (figure 1, panel B). As showed in panel A, 'fear of blood, injection and injury' starts to be endorsed as a little for subjects laying at the mild end of the fear latent trait, whereas only when classified as a lot is that it the item is more likely to be endorsed by those at the severe end of the latent trait. In contrast, panel B shows that for fear of 'meeting new people' the category 'a little' starts to be endorsed by subjects already lying at the severe end of the fear spectrum whereas those endorsing the 'a lot' category are at the extreme end of the fear latent trait.

Table 1 around here

Figure 1 around here

Diagnostic prediction (Receiver Operator Curves)

ROC analysis was used to predict the diagnosis of specific and social phobia from the corresponding fear items. Only fears characterized as ' a lot' were evaluated in this analysis because at least one of them is required for specific and social phobia diagnosis.

Just two out of twelve common fears answered as 'a lot' obtained LR+ considered useful for diagnosis (LR+ > 5), which were: 'toilets' and 'people who look unusual'. For all social fears, when classified as 'a lot', the LR+ was greater than 10 (17.7-47.8), indicating that screened positives are much more likely than screened negatives to be confirmed as cases of social phobia. It should be noted the low sensitivity, particularly among specific fears (5.6-55.1), indicating a low proportion of truly positive individuals among those diagnosed with specific phobia. The sensitivity increase for social fears, especially for 'performance only' fears (46.2-76.9) (Table 2).

Table 2 around here

Our analysis also showed that social fears tend to be more associated with phobias than specific fears. The AUC is also nominally higher for social fears predicting social phobia than for specific fears predicting specific phobia: AUC 0.775 (0.666-0.923) and AUC 0.609 (0.525-0.731), respectively (Table 2). Among the specific fears associated with greater severity, 'fear of toilets' and 'fear of loud noises' are positively associated with younger age and 'fear of vomiting, chocking, diseases' is positively associated with female gender. Among the social fears, fears classified as 'performance only' are associated with age and 'fear of writing in front of others' is associated with gender. Table 3 depicts a summary of our study findings.

Table 3 around here

Discussion

Our data is in accordance with the previous literature showing that fears are normative among children. In order to have a more precise picture of the clinical significance of specific and social fears, we investigated which of them are more likely to represent normative fears and problem indicators and which of them are more likely to ruling in and out the diagnosis of specific and social phobia. Our main findings can be summarized as follows. *First*, specific fears are generally less severe than social fears. *Second*, most of specific fears are normative when presented as 'a little', but 'situational fears' and 'people who look unusual' can indicate problems. When classified as 'a lot', all the specific fears indicate problems. Despite this, in the ROC analysis, only two fears were strong indicators of specific phobia diagnosis when answered as 'a lot' ('toilets' and 'people who look unusual'). *Third*, with respect to social fears, when classified as 'a little', fears 'non-restricted to performance' and 'writing in front of others' indicate problems. When classified as 'a lot', all social fears indicate problems. Furthermore, all social fears, when answered as 'a lot' were indicative of social phobia diagnosis in ROC analysis.

The developmental psychopathology theorizes that clinical patterns can be seen as deviations from normative patterns (Pine e Fox 2015). The fear system matures in the beginning of infant development and as a consequence normative fears are commonly present since very early in life (Blackford e Pine 2012; Salum et al. 2013b; Gullone 1999). Our data showed that common fears occur in most children and adolescents and that clinical syndromes can be associated with the increased severity of fears. It is noteworthy in our results the greatest severity of social fears in comparison with specific fears. Comparing two studies of Burstein *et al* (Burstein et al. 2012b; Burstein et al. 2011b) that evaluated the clinical correlates of specific and social phobia in the same population of adolescents, people with social phobia had worse rates on the Sheehan disability scale and days out of rule (Burstein et al. 2012b; Burstein et al. 2011b). Also there is evidence that social phobia is more co-morbid than specific phobia (Wittchen, Stein, e Kessler 1999a).

With respect to specific fears, our results are in agreement with Burstein et al (Burstein et al. 2012a) suggesting that the clinical relevance of specific phobia varies as a function of the nature of fears. However the clinical significance found in Burstein (Burstein et al. 2012a) was

somewhat different from the one found in ours. Whereas both studies showed that 'fear of animals' tend to be normative and 'fear of enclosed spaces' tend to indicate problems, our study found that fear of 'dark' and 'blood , injection, injury' tend to be normative, whereas Burstein (Burstein et al. 2012a) showed that such fears were associated with greater severity. Our study is also in the same direction as others showing that situational phobia was associated with high treatment-seeking behavior, use of medications, interference with daily and social life (LeBeau et al. 2010), as well as higher level of co-morbidity with other mental disorders (Park et al. 2013a; Depla et al. 2008a) in comparison with other types of specific phobia. In addition, we showed greater severity of 'people who look unusual', which are rarely studied in the previous literature.

With respect to social fears, contrasting fears non-restricted to performance and the performance only, co-morbidities were most common in subjects with fears non-restricted to performance than subjects with fears restricted to performance (Wittchen, Stein, e Kessler 1999a). Furthermore Burstein *et al* (Burstein et al. 2011a) showed that anxiety about non-performance situations demonstrates greater morbidity and clinical severity and have an earlier onset and higher degree of disability and impairment if compared to performance fears, which is in agreement with our findings. As the odds of social fear related to performance increases with age (Supplemental Table 2), performance fears may only become clinically significant in adulthood, when people can choose not to engage in public speaking activities, while in infancy and adolescence they are encouraged to participate on these situations (Burstein et al. 2011a).

The limitations of this work must be noted. *First*, our analysis is restricted to parent report fears. However, specific fears (as opposed to other forms of anxiety) have shown to be equally reported by parent and child reports (DeSousa et al. 2014). *Second*, our analysis is restricted to cross-sectional associations and important severity validators such as persistence could not be analyzed. *Finally*, in order to answer the complete DAWBA Specific Fear and Social Fears sections the subject must have at least one specific fear answered as 'A lot' (a condition also required for diagnosis) and, therefore, we were unable to use ROC analysis for fears classified as a little.

Advancing our understanding about the boundaries between normative symptoms from problem indicators is a major concern in children and adolescent psychiatric practice. Our

findings suggest parameters to alert the clinician when to be concerned with specific and social fears. Furthermore, we provide insights into the dimensionality of fear trait, showing that 'not all fears are created equally', meaning that they vary widely in terms of severity and might carry different information about a typical and an atypical development. Results also indicate that the classification by severity of fears provide clinically useful information for the diagnosis of specific and social phobia. Our outcomes might help clinicians and researchers to determine the boundaries that separate normative fears from problem indicators in children and early adolescents and indicate a differential severity threshold for specific and social fears.

Supplementary Material

Refer to Web version on Pubmed Central for supplementary material.

Acknowledgments

This work is supported by the National Institute of Developmental Psychiatry for Children and Adolescents, a science and technology institute funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; National Council for Scientific and Technological Development; grant number 573974/2008-0) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP; Research Support Foundation of the State of São Paulo; grant number 2008/57896-8). The authors thank the children and families for their participation, which made this research possible.

References

American Psychiatric Association. 2000. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed.

Benjet, Corina, Guilherme Borges, Dan J. Stein, Enrique Méndez, e María Elena Medina-Mora. 2012a. "Epidemiology of Fears and Specific Phobia in Adolescence: Results from the Mexican Adolescent Mental Health Survey". *The Journal of Clinical Psychiatry* 73 (2): 152–58. doi:10.4088/JCP.11m07442.

———. 2012b. "Epidemiology of Fears and Specific Phobia in Adolescence: Results from the Mexican Adolescent Mental Health Survey". *The Journal of Clinical Psychiatry* 73 (2): 152–58. doi:10.4088/JCP.11m07442.

———. 2012c. "Epidemiology of Fears and Specific Phobia in Adolescence: Results from the Mexican Adolescent Mental Health Survey". *The Journal of Clinical Psychiatry* 73 (2): 152–58. doi:10.4088/JCP.11m07442.

Blackford, Jennifer Urbano, e Daniel S. Pine. 2012. “Neural Substrates of Childhood Anxiety Disorders: A Review of Neuroimaging Findings”. *Child and Adolescent Psychiatric Clinics of North America* 21 (3): 501–25. doi:10.1016/j.chc.2012.05.002.

Blanco, Carlos, Yang Xu, Franklin R. Schneier, Mayumi Okuda, Shang-Min Liu, e Richard G. Heimberg. 2011. “Predictors of Persistence of Social Anxiety Disorder: A National Study”. *Journal of Psychiatric Research* 45 (12): 1557–63. doi:10.1016/j.jpsychires.2011.08.004.

Burstein, Marcy, Katholiki Georgiades, Jian-Ping He, Anja Schmitz, Emily Feig, Gabriela Kattan Khazanov, e Kathleen Merikangas. 2012a. “Specific Phobia among U.S. Adolescents: Phenomenology and Typology”. *Depression and Anxiety* 29 (12): 1072–82. doi:10.1002/da.22008.

———. 2012b. “Specific Phobia among U.S. Adolescents: Phenomenology and Typology”. *Depression and Anxiety* 29 (12): 1072–82. doi:10.1002/da.22008.

Burstein, Marcy, Jian-Ping He, Gabriela Kattan, Anne Marie Albano, Shelli Avenevoli, e Kathleen R. Merikangas. 2011a. “Social Phobia and Subtypes in the National Comorbidity Survey-Adolescent Supplement: Prevalence, Correlates, and Comorbidity”. *Journal of the American Academy of Child and Adolescent Psychiatry* 50 (9): 870–80. doi:10.1016/j.jaac.2011.06.005.

———. 2011b. “Social Phobia and Subtypes in the National Comorbidity Survey-Adolescent Supplement: Prevalence, Correlates, and Comorbidity”. *Journal of the American Academy of Child and Adolescent Psychiatry* 50 (9): 870–80. doi:10.1016/j.jaac.2011.06.005.

Depla, Marja F. I. A., Margreet L. ten Have, Anton J. L. M. van Balkom, e Ron de Graaf. 2008a. “Specific Fears and Phobias in the General Population: Results from the Netherlands Mental Health Survey and Incidence Study (NEMESIS)”. *Social Psychiatry and Psychiatric Epidemiology* 43 (3): 200–208. doi:10.1007/s00127-007-0291-z.

———. 2008b. “Specific Fears and Phobias in the General Population: Results from the Netherlands Mental Health Survey and Incidence Study (NEMESIS)”. *Social Psychiatry and Psychiatric Epidemiology* 43 (3): 200–208. doi:10.1007/s00127-007-0291-z.

DeSousa, Diogo A., Anderson S. Pereira, Circe S. Petersen, Gisele G. Manfro, Giovanni A. Salum, e Silvia H. Koller. 2014. “Psychometric Properties of the Brazilian-Portuguese Version of the Spence Children’s Anxiety Scale (SCAS): Self- and Parent-Report Versions”. *Journal of Anxiety Disorders* 28 (5): 427–36. doi:10.1016/j.janxdis.2014.03.006.

Diagnostic and Statistical Manual of Mental Disorder DSM-IV. [s.d.]. Draft, 1990.

Fleitlich-Bilyk, Bacy, e Robert Goodman. 2004. “Prevalence of Child and Adolescent Psychiatric Disorders in Southeast Brazil”. *Journal of the American Academy of Child and Adolescent Psychiatry* 43 (6): 727–34. doi:10.1097/01.chi.0000120021.14101.ca.

Goodman, R., T. Ford, H. Richards, R. Gatward, e H. Meltzer. 2000. “The Development and Well-Being Assessment: Description and Initial Validation of an Integrated Assessment of Child and Adolescent Psychopathology”. *Journal of Child Psychology and Psychiatry, and Allied Disciplines* 41 (5): 645–55.

Gullone, E. 1999. “The Assessment of Normal Fear in Children and Adolescents”. *Clinical Child and Family Psychology Review* 2 (2): 91–106.

Iza, M., M. M. Wall, R. G. Heimberg, T. L. Rodebaugh, F. R. Schneier, S.-M. Liu, e C. Blanco. 2014. “Latent Structure of Social Fears and Social Anxiety Disorders”. *Psychological Medicine* 44 (2): 361–70. doi:10.1017/S0033291713000408.

Jackson, Dennis L., J. Arthur Gillaspy, e Rebecca Purc-Stephenson. 2009. “Reporting Practices in Confirmatory Factor Analysis: An Overview and Some Recommendations”. *Psychological Methods* 14 (1): 6–23. doi:10.1037/a0014694.

Kessler, R. C., J. R. Calabrese, P. A. Farley, M. J. Gruber, M. A. Jewell, W. Katon, P. E. Keck, et al. 2013a. “Composite International Diagnostic Interview Screening Scales for DSM-IV Anxiety and Mood Disorders”. *Psychological Medicine* 43 (8): 1625–37. doi:10.1017/S0033291712002334.

———. 2013b. “Composite International Diagnostic Interview Screening Scales for DSM-IV Anxiety and Mood Disorders”. *Psychological Medicine* 43 (8): 1625–37. doi:10.1017/S0033291712002334.

LeBeau, Richard T., Daniel Glenn, Betty Liao, Hans-Ulrich Wittchen, Katja Beesdo-Baum, Thomas Ollendick, e Michelle G. Craske. 2010. “Specific Phobia: A Review of DSM-IV Specific Phobia and Preliminary Recommendations for DSM-V”. *Depression and Anxiety* 27 (2): 148–67. doi:10.1002/da.20655.

Margolis, David J., Warren Bilker, Raymond Boston, Russell Localio, e Jesse A. Berlin. 2002. “Statistical Characteristics of Area under the Receiver Operating Characteristic Curve for a Simple Prognostic Model Using Traditional and Bootstrapped Approaches”. *Journal of Clinical Epidemiology* 55 (5): 518–24.

Muris, P., H. Merckelbach, B. Mayer, e E. Prins. 2000. “How Serious Are Common Childhood Fears?”. *Behaviour Research and Therapy* 38 (3): 217–28.

Muthén, B.O., Muthén, L.K. 1998. *Mplus User’s Guide*. Fifth. Los Angeles, CA.

Ollendick, T. H., e N. J. King. 1994. “Fears and Their Level of Interference in Adolescents”. *Behaviour Research and Therapy* 32 (6): 635–38.

Park, Subin, Jee Hoon Sohn, Jin Pyo Hong, Sung Man Chang, Young Moon Lee, Hong Jin Jeon, Seong-Jin Cho, et al. 2013a. “Prevalence, Correlates, and Comorbidities of Four DSM-IV Specific Phobia Subtypes: Results from the Korean Epidemiological Catchment Area Study”. *Psychiatry Research* 209 (3): 596–603. doi:10.1016/j.psychres.2012.12.025.

———. 2013b. “Prevalence, Correlates, and Comorbidities of Four DSM-IV Specific Phobia Subtypes: Results from the Korean Epidemiological Catchment Area Study”. *Psychiatry Research* 209 (3): 596–603. doi:10.1016/j.psychres.2012.12.025.

Pine, Daniel S., e Nathan A. Fox. 2015. “Childhood Antecedents and Risk for Adult Mental Disorders”. *Annual Review of Psychology* 66 (janeiro): 459–85. doi:10.1146/annurev-psych-010814-015038.

Polo, Antonio J., Margarita Alegría, Chih-Nan Chen, e Carlos Blanco. 2011. “The Prevalence and Comorbidity of Social Anxiety Disorder among United States Latinos: A Retrospective Analysis of Data from 2 National Surveys”. *The Journal of Clinical Psychiatry* 72 (8): 1096–1105. doi:10.4088/JCP.08m04436.

R. Brian Haynes, David L. Sackett, Gordon H. Guyatt, e Peter Tugwell. 2005. *How to Do Clinical Practice Research*. 3rd ed.

Reise, Steven P., e Niels G. Waller. 2009. “Item Response Theory and Clinical Measurement”. *Annual Review of Clinical Psychology* 5: 27–48. doi:10.1146/annurev.clinpsy.032408.153553.

Rizopoulos, D. 2006. “ltm: An R Package for Latent Variable Modeling and Item Response Theory Analyses” 17 (5).

Ruscio, A. M., T. A. Brown, W. T. Chiu, J. Sareen, M. B. Stein, e R. C. Kessler. 2008a. "Social Fears and Social Phobia in the USA: Results from the National Comorbidity Survey Replication". *Psychological Medicine* 38 (1): 15–28. doi:10.1017/S0033291707001699.

———. 2008b. "Social Fears and Social Phobia in the USA: Results from the National Comorbidity Survey Replication". *Psychological Medicine* 38 (1): 15–28. doi:10.1017/S0033291707001699.

Salum, Giovanni Abrahão, Diogo Araújo Desousa, Maria Conceição do Rosário, Daniel Samuel Pine, e Gisele Gus Manfro. 2013a. "Pediatric Anxiety Disorders: From Neuroscience to Evidence-Based Clinical Practice". *Revista Brasileira De Psiquiatria (São Paulo, Brazil: 1999)* 35 Suppl 1: S03–21. doi:10.1590/1516-4446-2013-S108.

———. 2013b. "Pediatric Anxiety Disorders: From Neuroscience to Evidence-Based Clinical Practice". *Revista Brasileira De Psiquiatria (São Paulo, Brazil: 1999)* 35 Suppl 1: S03–21. doi:10.1590/1516-4446-2013-S108.

Salum, Giovanni Abrahão, Ary Gadelha, Pedro Mario Pan, Tais Silveira Moriyama, Ana Soledade Graeff-Martins, Ana Carina Tamanaha, Pedro Alvarenga, et al. 2014. "High Risk Cohort Study for Psychiatric Disorders in Childhood: Rationale, Design, Methods and Preliminary Results". *International Journal of Methods in Psychiatric Research*, dezembro. doi:10.1002/mpr.1459.

Samejima F. 1970. "Estimation of latent ability using a response pattern of graded scores". Psychometrika Monograph.

The Diagnostic and statistical manual of mental disorders. 2013. 5th ed.

T. H. Wilmshurst. 1990. *Signal Recovery from Noise in Electronic Instrumentation*.

Timothy A. Brown PsyD. 2006. *Confirmatory Factor Analysis for Applied Research*. 1st ed.

Wakschlag, Lauren S., Seung W. Choi, Alice S. Carter, Heide Hullsiek, James Burns, Kimberly McCarthy, Ellen Leibenluft, e Margaret J. Briggs-Gowan. 2012a. "Defining the Developmental Parameters of Temper Loss in Early Childhood: Implications for Developmental Psychopathology". *Journal of Child Psychology and Psychiatry, and Allied Disciplines* 53 (11): 1099–1108. doi:10.1111/j.1469-7610.2012.02595.x.

———. 2012b. "Defining the Developmental Parameters of Temper Loss in Early Childhood: Implications for Developmental Psychopathology". *Journal of Child Psychology and Psychiatry, and Allied Disciplines* 53 (11): 1099–1108. doi:10.1111/j.1469-7610.2012.02595.x.

Wittchen, H. U., M. B. Stein, e R. C. Kessler. 1999a. "Social Fears and Social Phobia in a Community Sample of Adolescents and Young Adults: Prevalence, Risk Factors and Co-Morbidity". *Psychological Medicine* 29 (2): 309–23.

———. 1999b. "Social Fears and Social Phobia in a Community Sample of Adolescents and Young Adults: Prevalence, Risk Factors and Co-Morbidity". *Psychological Medicine* 29 (2): 309–23.

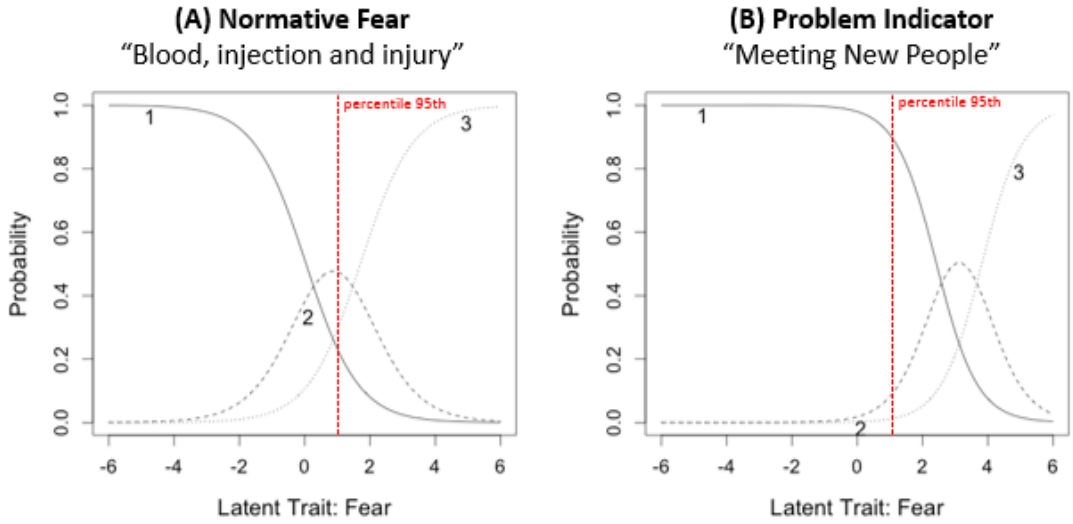


Figure 1. Illustrative Item Response Functions. (A) Normative Fear. (B) Problem Indicator.

Table 1 - Item Response Theory for specific and social fears in the total sample (n=2,512)

DSM-5 classification	Confirmatory Factor Analysis (Bifactor Model, General Factor)					Item Response Theory (unidimensional model)				
	Factor Loadings (λ)	Category Thresholds		Location (mean)	Location rank	Slope (a)	Category Thresholds		Item Location (mean b)	Location rank
		A Little	A lot				A Little (b1)	A lot (b2)		
Specific Fears										
Animals	0.221	-0.126	**0.996**	0.435	1	0.942	-0.254	**2.048**	0.897	2
Natural Environments										
Storms, thunder, Heights	0.262	0.103	**1.133**	0.618	4	1.286	0.168	**1.885**	1.0265	4
Dark	0.259	0.025	**0.978**	0.502	2	1.12	0.033	**1.766**	0.8995	3
Blood, injection, injury										
Blood, injection, injury	0.34	0.029	**1.017**	0.523	3	1.291	0.044	**1.677**	0.8605	1
Dentists, doctors	0.403	0.535	**1.392**	0.964	6	1.218	0.904	**2.426**	1.665	6
Situational										
Types of transport	0.402	**1.356**	**2.081**	1.719	15	1.444	**2.116**	**3.382**	2.749	13
Enclosed spaces	0.37	**1.109**	**1.849**	1.479	12	1.52	**1.666**	**2.872**	2.269	12
Toilets	0.356	**1.300**	**2.024**	1.662	13	1.316	**2.153**	**3.493**	2.823	14
Other fears										
Monsters etc	0.322	0.533	**1.384**	0.959	5	1.414	0.823	**2.191**	1.507	5
Vomiting, choking, diseases	0.388	0.770	**1.610**	1.190	8	1.414	1.196	**2.575**	1.8855	9

Loud noises	0.344	**0.886**	**1.643**	1.265	9	1.584	1.299	**2.469**	1.884	8
People who look unusual	0.351	**1.379**	**2.024**	1.702	14	1.357	**2.243**	**3.417**	2.83	15
Social fears										
Non-restricted to performance										
Meeting new people	0.783	**1.632**	**2.449**	2.041	18	1.409	**2.626**	**4.165**	3.3955	18
Meeting a lot of people	0.89	**1.524**	**2.259**	1.892	16	1.644	**2.234**	**3.431**	2.8325	16
Eating in front of others	0.804	**1.506**	**2.449**	1.978	17	1.555	**2.278**	**3.917**	3.0975	17
Performance only										
Speaking in class	0.747	**0.806**	**1.811**	1.309	10	1.399	1.262	**2.95**	2.106	10
Reading in front of others	0.727	0.566	**1.551**	1.059	7	1.33	0.913	**2.571**	1.742	7
Writing in front of others	0.739	**1.041**	**1.871**	1.456	11	1.504	**1.576**	**2.921**	2.2485	11

Note: All items loaded significantly on their respective latent factors ($p<0.0001$). **Bold mark** problem indicators.

Table 2 - Associations between each specific and social fear classified as 'A lot' with specific phobia and social phobia diagnosis in the total sample (n=2,512)

Specific fears	Sens	Spe	CC	LR+	LR-	ROC	CI95%
Animals	41.6	85	83.4	2.8	0.7	0.656	0.597-0.715
Natural Environments							
Storms, thunder, heights	37.1	88	86.2	3.1	0.7	0.654	0.595-0.714
Dark	55.1	85	84	3.7	0.5	0.731	0.677-0.785
Blood, injection, injury							
Blood, injection, injury	38.2	85.4	83.7	2.6	0.7	0.643	0.584-0.702
Dentists, doctors	20.2	92.2	89.7	2.6	0.9	0.612	0.556-0.668
Situational							
Types of transport	5.6	98.3	95	3.2	1	0.548	0.507-0.589
Enclosed spaces	6.7	96.9	93.7	2.2	1	0.525	0.484-0.567
Toilets	13.5	98.3	95.3	7.8	0.9	0.587	0.540-0.635
Others							
Monsters etc	29.2	92.5	90.2	3.9	0.8	0.642	0.583-0.700
Vomiting, diseases	12.3	94.9	92	2.4	0.9	0.587	0.534-0.639
Loud noises	16.9	95.4	92.6	3.7	0.9	0.585	0.532-0.637
People who look unusual	10.1	98.1	95	5.4	0.9	0.546	0.505-0.586
Social fears	Sens	Spe	CC	LR+	LR-	ROC	CI95%
Non-restricted to performance							
Meeting new people	23.1	99.5	98.7	47.8	0.8	0.769	0.671-0.867
Meeting a lot of people	30.8	99.1	98.4	34.8	0.7	0.688	0.588-0.789
Eating in front of others	19.2	99.5	98.7	36.8	0.8	0.666	0.567-0.763
Performance only							
Speaking in class	57.7	97.1	96.7	19.6	0.4	0.83	0.735-0.924
Reading aloud in front of others	76.9	94.7	94.1	14.5	0.2	0.923	0.873-0.972
Writing in front of others	46.2	97.4	96.9	17.7	0.5	0.779	0.678-0.880

Note: Fears were ordered by the DSM-5 classification.

Table 3 – Summary of Confirmatory Factor Analysis (CFA), Item response Theory (IRT) and Receiver Operator Curve (ROC) Analysis for each specific and social fears with respect to the severity and likelihood to indicate problems in the total sample (n=2,512) and their correlates with age and gender.

DSM-5 classification	CFA		IRT		ROC ANALYSIS	Associations with age	Associations with gender (male)
	A little (Theta>0.78)	A lot (Theta>0.78)	A little (Theta>1.53)	A lot (Theta>1.53)	A lot (LR>5)		
Specific fears							
Animals	No	Yes	No	Yes	No	No	Yes
Natural environments							
Storms, thunder, heights	No	Yes	No	Yes	No	No	Yes
Dark	No	Yes	No	Yes	No	Yes	No
Blood, injection, injury							
Blood, injection, injury	No	Yes	No	Yes	No	No	Yes
Dentists, doctors	No	Yes	No	Yes	No	No	No
Situational							
Types of transport	Yes	Yes	Yes	Yes	No	No	No
Enclosed spaces	Yes	Yes	Yes	Yes	No	No	No
Toilets	Yes	Yes	Yes	Yes	Yes	Yes	No
Other fears							
Monsters etc	No	Yes	No	Yes	No	Yes	No
Vomiting, choking, diseases	No	Yes	No	Yes	No	No	Yes
Loud noises	Yes	Yes	No	Yes	No	Yes	No

People who look unusual	Yes	Yes	Yes	Yes	Yes	No	No
Social fears							
Non-restricted to performance							
Meeting new people	Yes	Yes	Yes	Yes	Yes	No	No
Meeting a lot of people	Yes	Yes	Yes	Yes	Yes	No	No
Eating in front of others	Yes	Yes	Yes	Yes	Yes	No	No
Performance only							
Speaking in class	Yes	Yes	No	Yes	Yes	Yes	No
Reading aloud in front of others	No	Yes	No	Yes	Yes	Yes	No
Writing in front of others	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Fears were ordered by the DSM-V classification.

Supplemental Table 1 - Prevalence of common fears in the randomly selected sample (n=958)

	No		A little		A lot	
DSM-5 classification	N	%	N	%	N	%
Specific fears						
Animals	443	46.2	382	39.9	133	13.9
Natural Environments						
Storms, thunder, heights	550	57.4	295	30.8	113	11.8
Dark	503	52.5	315	32.9	140	14.6
Blood, injection, injury						
Blood, injection, injury	525	54.8	308	32.2	125	13
Dentists, doctors	712	74.3	186	19.4	60	6.3
Situational						
Types of transport	894	93.3	49	5.1	15	1.6
Enclosed spaces	861	89.9	73	7.6	24	2.5
Toilets	882	92.1	60	6.3	16	1.7
Others						
Monsters	700	73.1	197	20.6	61	6.4
Vomiting, choking, diseases	774	80.8	145	15.1	39	4.1
Loud noises	800	83.5	120	12.5	38	4
People who look unusual	874	91.2	67	7	17	1.8
Social fears						
Non-restricted to performance						
Meeting new people	913	95.3	39	4.1	6	0.6
Meeting a lot of people	902	94.2	44	4.6	12	1.3
Eating in front of others	900	93.9	49	5.1	9	0.9
Performance only						
Speaking in class	794	82.9	141	14.7	23	2.4
Reading aloud in front of others	730	76.2	181	18.9	47	4.9
Writing in front of others	838	87.5	97	10.1	23	2.4

Note: Fear of ‘other things’ – ‘No’ (n=900; 93.9%), ‘A little’ (n= 31; 3.2%) and ‘A Lot’ (n=27; 2.8%)

Supplemental Table 2- Prevalence of common fears according to the ages in the randomly selected sample (n=958)

DSM-5 classification	Age group diferences								Gender differences				Ordinal regression (OR)	
	06 to 08		09 to 10		11 to 12		13 to 14		Male		Female		Age (continuous,years)	Gender, (Ref: male)
	N	%	n	%	n	%	n	%	n	%	n	%		
Specific fears														
Animals														
A little	133	40.9	137	39.4	82	39.0	30	41.1	169	35.1	213	44.7		
A lot	42	12.9	49	14.1	32	15.2	10	13.7	46	9.6	87	18.2	1.007 (0.944-1.073)	2.099 (1.643-2.681)***
Natural Environments														
Storms, thunder, heights														
A little	110	33.8	109	31.3	57	27.1	19	26.0	146	30.4	149	31.2		
A lot	28	8.6	48	13.8	30	14.3	7	9.6	43	8.9	70	14.7	0.984 (0.922-1.051)	1.377 (1.073-1.766)*
Dark														
A little	125	38.5	114	32.8	62	29.5	14	19.2	146	30.4	169	35.4		

A lot	49	15.1	58	16.7	26	12.4	7	9.6	77	16.0	63	13.2	0.888 (0.832-0.948)***	1.024 (0.803-1.307)
Blood, injection, injury														
Blood, injection, injury														
A little	97	29.8	117	33.6	73	34.8	21	28.8	147	30.6	161	33.8		
A lot	49	15.1	46	13.2	24	11.4	6	8.2	55	11.4	70	14.7	0.969 (0.908-1.034)	1.305 (1.021-1.669)*
Dentists, doctors														
A little	57	17.5	64	18.4	51	24.3	13	17.8	85	17.7	101	21.2		
A lot	15	4.6	26	7.5	15	7.1	4	5.5	27	5.6	33	6.9	1.058 (0.981-1.140)	1.278 (0.958-1.707)
Situational														
Types of transport														
A little	15	4.6	16	4.6	11	5.2	6	8.2	22	4.6	27	5.7		
A lot	5	1.5	6	1.7	3	1.4	1	1.4	8	1.7	7	1.5	1.024 (0.897-1.169)	1.147 (0.690-1.906)
Enclosed spaces														

A little	20	6.2	24	6.9	23	11.0	6	8.2	31	6.4	42	8.8		
A lot	9	2.8	8	2.3	6	2.9	1	1.4	10	2.1	14	2.9	1.045 (0.936-1.166)	1.424 (0.932-2.177)
Toilets														
A little	24	7.4	22	6.3	11	5.2	3	4.1	26	5.4	34	7.1		
A lot	9	2.8	4	1.1	3	1.4	0	0.0	5	1.0	11	2.3	0.876 (0.771-0.995)*	1.534 (0.952-2.472)
Others														
Monsters														
A little	80	24.6	75	21.6	36	17.1	6	8.2	97	20.2	100	21.0		
A lot	22	6.8	26	7.5	10	4.8	3	4.1	24	5.0	37	7.8	0.875 (0.810-0.945)**	1.240 (0.932-1.648)
Vomiting, choking, diseases														
A little	58	17.8	43	12.4	37	17.6	7	9.6	67	13.9	78	16.4		
A lot	10	3.1	18	5.2	7	3.3	4	5.5	13	2.7	26	5.5	0.967 (0.888-1.052)	1.426 (1.033-1.969)*

Loud noises														
A little	52	16.0	40	11.5	21	10.0	6	8.2	56	11.6	64	13.4		
A lot	12	3.7	17	4.9	7	3.3	2	2.7	21	4.4	17	3.6	0.882 (0.805-0.967)**	1.065 (0.757-1.498)
People who look unusual														
A little	26	8.0	26	7.5	10	4.8	5	6.8	29	6.0	38	8.0		
A lot	9	2.8	4	1.1	4	1.9	0	0.0	10	2.1	7	1.5	0.922 (0.818-1.039)	1.175 (0.750-1.842)
Social fears														
Non-restricted to performance														
Meeting new people														
A little	14	4.3	12	3.4	12	5.7	1	1.4	14	2.9	25	5.2		
A lot	2	0.6	1	0.3	1	0.5	2	2.7	5	1.0	1	0.2	1.005 (0.858-1.177)	1.386 (0.757-2.540)
Meeting a lot of people														

A little	13	4.0	15	4.3	14	6.7	2	2.7	18	3.7	26	5.5		
A lot	4	1.2	3	0.9	2	1.0	3	4.1	8	1.7	4	0.8	1.040 (0.903-1.198)	1.160 (0.675-1.992)
Eating in front of others														
A little	15	4.6	14	4.0	19	9.0	1	1.4	24	5.0	25	5.2		
A lot	2	0.6	1	0.3	3	1.4	3	4.1	3	0.6	6	1.3	1.089 (0.947-1,251)	1.169 (0.687-1.992)
Performance only														
Speaking in class														
A little	32	9.8	57	16.4	46	21.9	6	8.2	76	15.8	65	13.6		
A lot	6	1.8	3	0.9	11	5.2	3	4.1	11	2.3	12	2.5	1.143 (1.046-1.249)**	0.868 (0.619-1.217)
Reading aloud in front of others														
A little	49	15.1	64	18.4	52	24.8	16	21.9	90	18.7	91	19.1		
A lot	9	2.8	17	4.9	18	8.6	3	4.1	28	5.8	19	4.0	1.155 (1.069-1.249)***	0.891 (0.662-1.201)

Writing in front of others														
A little	25	7.7	35	10.1	30	14.3	7	9.6	57	11.9	40	8.4		
A lot	4	1.2	7	2.0	10	4.8	2	2.7	14	2.9	9	1.9	1.127 (1.020-1.247)*	0.653 (0.443-0.964)*

Note: Fears were ordered by the DSM-V classification. *p<0.05; **p<0.01; ***p<0.001

Supplemental Table 3 - Fit index of the unidimensional, correlated and bifactor models for specific and social fears

Items	One Factor	2 correlated		Bifactor		
	Fear	Specific	Social	G	Specific	Social
	Λ	λ	λ	λ	λ	Λ
Animals	0.459	0,524		0,221	0,495	
Storms, thunder, Heights	0.567	0,646		0,262	0,62	
Dark	0.537	0.604		0.259	0.566	
Blood. injection. Injury	0.572	0.642		0.34	0.538	
Dentists. doctors	0.56	0.613		0.403	0.444	
Types of transport	0.574	0.635		0.402	0.476	
Enclosed spaces	0.608	0.677		0.37	0.563	
Toilets	0.54	0.599		0.356	0.473	
Monsters etc	0.608	0.68		0.322	0.608	
Vomiting. choking. diseases	0.583	0.645		0.388	0.502	
Loud noises	0.641	0.712		0.344	0.632	
People who look unusual	0.551	0.612		0.351	0.493	
Meeting new people	0.639		0.739	0.783		-0.239
Meeting a lot of people	0.695		0.8	0.89		-0.304
Eating in front of others	0.651		0.753	0.804		-0.065
Speaking in class	0.751		0.831	0.747		0.387
Reading in front of others	0.842		0.918	0.727		0.657
Writing in front of others	0.851		0.899	0.739		0.502
Correlations	-	0.469		0	0	0
Fit indexes						
FP	54	55		72		
X2	3379.9 (df=135; p<0.001)	898.2 (df=134; p<0.001)		627.3 (df=117; p<0.001)		

RMSEA	0.098		0.048		0.042	
CI 90%	0.095-0.101		0.045-0.051		0.038-0.045	
CFI	0.799		0.953		0.968	
TLI	0.772		0.946		0.959	
WRMR	4.242		2.093		1.482	
Reliability indexes						
Omega	0.921	0.889	0.870	0.933	0.889	0.945
Omega h				0.600		
Omega s					0.638	0.036

Note: λ: Factor loadings; G: general fator.

Supplemental Table 4 - Age and gender Multigroup Confirmatory Factor Analysis investigating measurement invariance for the fear latent trait

Parameters	Gender	Age
Model Fit Information for the Configural Model		
Chi-Square	724.370	873.712
Degrees of Freedom	234	468
RMSEA	0.041	0.037
CFI	0.970	0.973
TLI	0.960	0.965
WRMR	1.670	1.784
Model Fit Information for the Scalar Model		
Chi-Square	761.391	1.020.536
Degrees of Freedom	282	612
RMSEA	0.037	0.033
CFI	0.970	0.973
TLI	0.968	0.973
WRMR	1.892	2.235
ΔCFI Configural – Scalar	0.000	0.000

Note: Chi-square values for each group of gender for the configural model: male (X2=337.631). female (X2=386.739). Scalar model: male (X2=361.560). female (X2=399.831). Chi-square values for each group of age for the configural model: 6 to 8 (X2=264.169). 9 to 10 (X2=270.357). 11 to 12 (X2=218.960). 13 to 14 (X2=120.226). Scalar model: 6 to 8 (X2=272.705). 9 to 10 (X2=256.302). 11 to 12 (X2=270.287). 13 to 14 (X2=221.241).

Appendix #2

Published in the Journal of Academy of Child and Adolescent Psychiatry

J Am Acad Child Adolesc Psychiatry 2016; 55(1):47–53

Positive attributes "buffers" the negative associations between low intelligence and high psychopathology with educational outcomes

Running title: Positive attributes and education.

Mauricio Scopel Hoffmann MD, MSc, Ellen Leibenluft MD, Argyris Stringaris MD, PhD, MRCPsych, Paola Paganella Laporte MD, Pedro Mario Pan MD, PhD, Ary Gadelha MD, PhD, Gisele Gus Manfro MD, PhD, Eurípedes Constantino Miguel MD, PhD, Luis Augusto Rohde MD, PhD, Giovanni Abrahão Salum MD, PhD.

Drs. Hoffmann and Laporte, are with Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil (HCPA) and Department of Psychiatry, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil (UFRGS). Dr. Leibenluft is with Section on Bipolar Spectrum Disorders, Intramural Research Program, National Institute of Mental Health, National Institutes of Health, Department of Health and Human Services, USA. Dr. Stringaris is with Department of Child and Adolescent Psychiatry, King's College London, Institute of Psychiatry, London, UK. Drs. Pan and Gadelha are with the Department of Psychiatry, Universidade Federal de São Paulo, São Paulo, Brazil (UNIFESP) and the National Institute of Developmental Psychiatry for Children and Adolescents, São Paulo, Brazil (INCT-CNPq). Dr. Manfro is with HCPA, UFRGS and INCT-CNPq. Dr. Miguel is with UNIFESP, INCT-CNPq and the Department & Institute of Psychiatry, Universidade de São Paulo, São Paulo, Brazil (USP). Dr. Rohde is with HCPA, UFRGS, INCT-CNPq and USP. Dr. Salum is with UFRGS and INCT-CNPq.

Correspondence: Mauricio Scopel Hoffmann, HCPA, UFRGS, Rua Ramiro Barcelos 2350 – room 2202, Porto Alegre, 90035-003, Brazil. Telephone/Fax (+55) 51 3359 8094. E-mail: mauriciodireito@yahoo.com.br

Number of words: Abstract, 216. Text, 5626.

Number of Figures: 2

Number of Tables: 2

Number of Supplementary Materials: 1

Key words: Noncognitive skills; youth strengths inventory; interaction; school.

Acknowledgments: This work is supported by the National Institute of Developmental Psychiatry for Children and Adolescents, a science and technology institute funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; National Council for Scientific and Technological Development; grant number 573974/2008-0) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP; Research Support Foundation of the State of São Paulo; grant number 2008/57896- 8). The authors thank the children and families for their participation, which made this research possible.

Summary

Objectives: This study examines the extent to which children's positive attributes are distinct from psychopathology. We also investigate whether positive attributes change or 'buffer' the impact of low intelligence and high psychopathology on negative educational outcomes.

Methods: In a community sample of 2,240 children (6-14 years of age), we investigated associations among positive attributes, psychopathology, intelligence, and negative educational outcomes. Negative educational outcomes were operationalized as learning problems and poor academic performance. We tested the discriminant validity of psychopathology vs. positive attributes using Confirmatory Factor Analysis (CFA) and Propensity Score Matching Analysis (PSM) and used generalized estimating equations (GEE) models to test main effects and interactions among predictors of educational outcomes.

Results: According to both CFA and PSM, positive attributes and psychiatric symptoms were distinct constructs. Positive attributes were associated with lower levels of negative educational outcomes, independent of intelligence and psychopathology. Positive attributes buffer the negative effects of lower intelligence on learning problems, and higher psychopathology on poor academic performance.

Conclusion: Children's positive attributes are associated with lower levels of negative school outcomes. Positive attributes act both independently and by modifying the negative effects of low intelligence and high psychiatric symptoms on educational outcomes. Subsequent research should test interventions designed to foster the development of positive attributes in children at high risk for educational problems.

INTRODUCTION

Educational attainment in childhood is a powerful predictor of economic success, health, and well-being later in life.[1–3] Both intelligence[4] and psychiatric symptoms[5,6] influence an individual's performance in educational settings. However, recent econometric studies also highlight the impact of positive attributes – such as being keen to learn, affectionate and caring – on educational attainment.[7–10] Whereas research has begun to examine the role of positive attributes on determining education outcomes[11,12] , major questions remain.

First, it is important to determine whether positive attributes are a distinct construct, separable from the absence of psychiatric symptoms.[11] Economic studies cannot answer this question because they do not include measures of psychopathology. The few available studies in psychiatry[11,12] support the independent contributions of positive attributes and psychiatric symptoms in predicting the subsequent development of psychiatric illness. However, the distinction between positive attributes and psychiatric symptoms has not been examined psychometrically.

Second, if positive attributes are indeed distinct from the absence of psychiatric symptoms, it is important to investigate interactions between these two constructs and intelligence in predicting educational outcomes. Consistent with economic theories of human development, evidence suggests that positive attributes and intelligence may interact in predicting educational outcomes, such as school graduation by age 30.[1,13] However, no studies investigate interactive effects between positive attributes and psychopathology on educational outcomes. Specifically, it is important to ascertain if positive attributes buffer the negative impact of low intelligence and high psychiatric symptoms on educational outcomes. If positive attributes have such buffering properties, then facilitating their emergence might improve outcomes in children who are at risk for adverse educational outcomes because of psychiatric symptoms or low intelligence.

Here we aim to investigate: (1) the discriminant validity of the constructs of positive attributes and psychiatric symptomatology in children; and (2) whether positive attributes are independently associated with educational outcomes and/or if they buffer associations between low intelligence and

negative educational outcomes, and between high psychiatric symptoms and negative educational outcomes. First, we predict that positive attributes are empirically discriminable from psychiatric symptoms. Second, we predict that positive attributes are associated with lower levels of negative educational outcomes independent of intelligence and psychopathology, and through interactions with low intelligence and high levels of psychiatric symptoms that buffer the impact of these two variables on negative educational outcomes.

METHODS

Participants

We used data from a large school-based community study that obtained psychological, genetic and neuroimaging data and was designed to investigate typical and atypical trajectories of psychopathology and cognition over development.[14] The ethics committee of the University of São Paulo approved the study. Written consent was obtained from parents of all research participants and verbal assent was obtained from the children.

The study included screening and assessment phases. The screening phase of the study included children from 57 public schools in São Paulo and Porto Alegre. In Brazil, on specified registration days, at least one caregiver is required to register each child for compulsory school attendance. All parents and children who presented at the selected schools were invited to participate. Families were eligible for the study if the children: (1) were registered by a biological parent capable of providing consent and information about the children's behavior; (2) were between 6-12 years of age; and (3) remained in the same school during the study period.

We screened 9,937 parents using the Family History Survey (FHS).[15] From this pool, we recruited two subgroups - one randomly selected (n=958) and one high-risk (n=1,524). Selection of the high-risk sample involved a risk-prioritization procedure designed to identify individuals with current symptoms and/or a family history of specific disorders.[14]

The assessment phase was performed in multiple visits, in the following order: home interview with parents (one visit), child assessment with a psychologist (one or two visits), child assessment with a speech therapist (one or two visits), and one hospital visit for imaging and blood collection.

From the total sample (N=2,512), missing data for intelligence and learning problems was handled using listwise deletion. Hence, a subset of 2,240 research participants (862 randomly selected and 1,378 high-risk) with complete intelligence measurements[16] were included in the present analysis. In this subsample, 1,987 research participants (783 randomly selected and 1,204 high-risk) had complete measurements of learning problems.[17] Subjects with missing intelligence data had lower mean age (9.53 vs. 10.37 [$F_{(1,2510)}$=81.28, p<0.001]) than included subjects, but did not differ on gender, socioeconomic status or psychiatric symptoms. Parent informants were mother (91.6%), father (4.4%) or both (4%).

Positive Attributes Measurement

To measure positive attributes in children and adolescents, we used the Youth Strength Inventory (YSI), a subscale of the Development and Well-Being Assessment (DAWBA).[11] The YSI is a 24-item scale, divided into two blocks of questions addressed to the caregiver. One block focuses on child characteristics, such as if he/she is “lively”, “easy going”, “grateful”, “responsible”, and has a “good sense of humour”. The other block addresses the child’s actions that please others, such as “helps around the home”, “well behaved”, “keeps bedroom tidy”, “does homework without reminding”. Each question is answered, “No”, “A little”, or “A lot”. A confirmatory factor analysis (CFA) of YSI yielded a one-factor solution with adequate goodness-of-fit indices (i.e., Root Mean Square Error of Approximation (RMSEA) 0.057 (90% CI 0.055-0.059), Comparative Fit Index (CFI) 0.957, Tucker Lewis Index (TLI) 0.950, Chi-Square Test of model fit 2201.316 (p<0.001)). Composite YSI scores were derived from saved factor scores from the CFA model (Table S1, available online).

Intelligence Evaluation

For intelligence, we estimated IQ using the vocabulary and block design subtests of the Weschler Intelligence Scale for Children, 3rd edition – WISC-III,[18] using the Tellegen and Briggs method[19] and Brazilian norms.[16,20]

Psychiatric Evaluation

Psychiatric symptoms were evaluated as a continuous variable, using the Strengths and Difficulties Questionnaire (SDQ).[21] SDQ is a 25-item questionnaire which provides five scores of behavioral and emotional symptoms. For the purposes of this study, we excluded "peer relationships problems" from the SDQ total because of the conceptual overlap among this variable, psychiatric symptoms, and positive attributes. The resulting measure, the SDQ composite (SDQc), includes "emotional symptoms", "inattention/hyperactivity" and "conduct problems".

Psychiatric diagnosis was assessed using the Brazilian Portuguese version[23] of the Development and Well-Being Assessment (DAWBA).[22] This structured interview was administered to biological parents by trained lay interviewers and scored by trained psychiatrists who were supervised by a senior child psychiatrist[14]. For the purposes of the propensity score matching (PSM) analysis we used the DAWBA broad category of 'Any Psychiatric Diagnosis'.

There were low Pearson's correlations between YSI and IQ ($r=0.105$; $p<0.001$) and between SDQ and IQ ($r=-0.146$; $p=<0.001$). There was a moderate correlation between YSI and SDQc ($r=-0.560$; $p=<0.001$).

Educational Evaluations

Educational evaluations consisted of direct measurement of learning problems in children and by the caregiver's report of the child's performance in academic subjects.

Specifically, learning problems were measured by participants' scores on the School Performance Test ("Teste de Desempenho Escolar" - TDE).[17] The TDE is comprised of two subtests, decoding (recognition of words isolated from context) and writing (isolated words in dictation). A previous TDE study from our group used Latent Class Analysis (LCA) to identify a cluster of children (18.5% of the sample) with poor decoding and writing skills.[24] Here, we used membership in this cluster to identify children with learning problems.

Academic performance was measured using Child Behavior Checklist for ages 6-18 (CBCL-school),[25] completed by the caregiver. The academic subjects assessed were Portuguese or literature, history or social studies, English or Spanish, mathematics, biology, sciences, geography, and computer studies. Each subject was scored as failing, below average, average, and above average. The CFA of CBCL-school using one-factor solution resulted in adequate goodness-of-fit indices (i.e., RMSEA 0.056 (90% CI 0.048-0.065), CFI 0.997, TLI 0.996, Chi-Square Test of model fit 49787.4
($p<0.001$)). The composite CBCL-school (academic performance) scores were derived from saved factor scores from the CFA model (Table S2, available online).

Statistical Analysis

We performed a stepwise analysis. We used two analytic methods to test the first hypothesis. First, we performed a CFA to investigate if YSI and SDQc items load onto one or two latent factors. Specifically, we fitted a one factor, two factors, second order and bifactor models. (For CFA methods and results, see Supplementary Material, available online). Second, we used a LCA to identify groups differing on level of positive attributes. We then used propensity score matching (PSM) to test if children differing only in positive attributes (and not on psychiatric diagnosis, symptoms, medication, IQ, age, gender, siblings, socioeconomic status or parents' psychiatric diagnosis) differ on school outcomes. Specifically, after propensity score matching, generalized estimating equations (GEE) models were used to test between-group differences in school outcomes. Since school outcomes might vary among the 57 schools, we controlled for cluster effects (random-effects) in all statistical tests. The LCA and PSM methods and results are detailed in Supplementary Material, available online.

We tested the second hypothesis using univariate models that included one independent variable at a time (i.e., YSI, IQ, SDQc); followed by bivariate models that included YSI and IQ or SDQc in the same model without the interaction term and finally a full model that included the main effects of YSI and IQ or SDQc and the interaction term (i.e., YSI*IQ and YSI*SDQc). To facilitate interpretation, IQ, positive attributes and psychiatric symptom scores were transformed into standardized units (z- scores), regressing out the effects of age and gender (using Studentized residuals). Again, study hypotheses were tested using GEE models in SPSS 17 (SPSS Inc, Chicago, Illinois, USA). We used binary logistic and linear regression models for learning problems and poor academic performance respectively. Therefore, model estimates (OR and β) reflect the outcome additive increase for changing one standardized unit of the predictors. Interactions were represented graphically using regression surfaces implemented in R (plot3D package[26]). We used marginal effects implemented in Stata version 13 (StataCorp, College Station, Texas, USA) to test the significance of the continuous interactions. Marginal effects represent the change in linear prediction (linear regression) and probability (logistic regression) of an outcome for a one IQ or SDQc standardized unit change when YSI is held constant at different values (-3.5 to 3.5, with 0.5 unit increases). For logistic regression, results were transformed from chances into probabilities to facilitate interpretation. For marginal effects analysis, we used the inverse levels of IQ (IQ * (-1)). For post-hoc power analyses of the main models, see Supplementary Material.

RESULTS

Hypothesis 1: Positive attributes are empirically discriminable from psychiatric symptoms.

CFA indicated that the model with two correlated factors showed the best fit indices over the other models (one factor, second order and bifactor models). The model with two correlated factors ('psychiatric symptoms' and 'positive attributes') showed acceptable goodness-of-fit across indices: RMSEA 0.061 (90% CI 0.059-0.062), CFI 0.903, TLI 0.895, Chi-Square Test of model fit 66086.108
($p<0.001$) as the model with one factor provided an unacceptable fit to the data according to two out of three fit indexes: RMSEA 0.077 (CI90% 0.076 – 0.079), CFI 0.842, TLI 0.830, Chi-Square Test of model fit 11012.799, df=689, $p<0.001$. Chi-Square Test for Difference Testing one-dimensional vs. correlated two factor models showed advantages of the two-factor correlated model over the one- factor model ($\chi^2=667.338$, df=1, $p<0.0001$). Second-order and bifactor models did not converge.

An item-level inspection of information curves from the CFA of the two-factor correlated model showed that YSI and SDQc provide information in different areas of a common metric (*i.e.*, YSI is better at discriminating among typically developing children, while SDQc is better at discriminating among atypically developing children). Specifically, the mean threshold of SDQc items was -0.19, whereas the mean threshold of YSI items was 0.83 (Figure S1, available online).

LCA indicated that the sample is divided into high (63.2%) and low (36.8%) positive attributes classes (Figure S2, available online). PSM procedures were able to generate two groups differing only in positive attributes levels (Figure S3, available online). As predicted, compared to the low YSI group, the high YSI group had lower means on the scale measuring poor academic performance ($\beta=0.72$; 95% CI [0.65-0.79]; $p<0.001$). Contrary to our predictions, YSI was not associated with a lower chance of having learning problems (OR=0.98; 95% CI [0.73-1.30], p=0.88).

Hypothesis 2: Positive attributes are associated with lower levels of negative educational

outcomes independent of intelligence and psychopathology, and through interactions with low intelligence and high levels of psychiatric symptoms that buffer the impact of these two variables on negative educational outcomes.

Positive attributes and intelligence

First we analyzed the associations of IQ and YSI on each outcome variable (Table 1). In both univariate and bivariate models, higher YSI and IQ were associated with lower chances of learning problems and lower levels of poor academic performance. For poor academic performance, the associations with IQ and YSI were independent of each other (Table 1, Model 3). For learning problems, there was a significant interaction between YSI and IQ, such that the association of intelligence on learning problems is moderated by children's positive attributes (Table 1, Model 3 and Figure 1A). Marginal effect analysis revealed that decreasing levels of IQ were significantly associated with higher probabilities of learning problems for individuals with YSI lower than 1.5 z-score, but not for those with YSI equal or higher than 1.5 z-score (Figure 1B). The strength of the association between levels of intelligence and learning problems decreases as a function of increasing levels of positive attributes. For example, at a YSI of -3.5 z score, the probability of learning problems increases 17.90% (95%CI 10.46% to 25.33%, $p<0.001$) for each IQ standardized unit decrease. At a YSI of 1 z-score, the probability of learning problems increases 4.21% (95%CI 1.50 to 6.93, $p=0.002$) for each IQ standardized unit decrease (Figure 1B). Importantly, when the YSI is ≥ 1.5 z-score, the associations between IQ and learning problems are non-significant (Figure 1B), suggesting that high levels of positive attributes buffer the negative impact of low intelligence on learning problems.

TABLE 1

FIGURE 1

Positive attributes and psychiatric symptoms

Lastly, we investigated the effect of psychiatric symptoms (SDQc) on school outcomes, again in univariate and bivariate models with child positive attributes (YSI) (Table 2). In the univariate model, higher SDQc were associated with higher levels of negative educational outcomes (Table 2, Model 1). In the bivariate models, both YSI and SDQc were significantly associated with learning problems and academic performance (Table 2, Model 2). For learning problems, associations with SDQc and YSI were independent (Table 2, Model 3). However, for poor academic performance, there was a significant interaction between YSI and SDQc, revealing that the association of psychiatric symptoms on performance in academic subjects is moderated by children's positive attributes (Table 2, Model 3 and Figure 2A). Marginal effect analysis revealed that increasing levels of psychiatric symptoms was significantly associated with poorer academic performance, for children and adolescents with YSI lower than 1.5 z-score, but not for those with YSI equal or higher than 1.5 z-score (Figure 2B). The strength of the association between levels of psychiatric symptoms and poor academic performance decreases as a function of increasing levels of positive attributes. For example, at a YSI of -3.5 z score, linear prediction of poor academic performance increases 0.403 z-score (95%CI 0.272 to 0.534, $p<0.001$) for each SDQc standardized unit increase. At a YSI of -1 z score, linear prediction of poor academic performance increases 0.115 z-score (95%CI 0.033 to 0.197, $p=0.007$) for each SDQc standardized unit increase (Figure 2B). At YSI $\geq$ 1.5 z score, the association between SDQc and poor academic performance is non-significant, suggesting that high levels of positive attributes buffer the negative impact of psychiatric symptoms on academic performance (Figure 2B).

TABLE 2

FIGURE 2

As a post-hoc analysis, we ran a second CFA for YSI, excluding items that could overlap with school outcomes ("keen to learn", "good at school work", "does homework without needing to be reminded"). A good model fit remained (RMSEA 0.057, 90% CI 0.055-0.060; CFI 0.961; TLI 0.955; Chi-Square Test of model fit 1681.197, $p<0.001$). We re-ran all the regressions using YSI scores

without school items and found the same main effects and interactions described above. Also, for each model, three-way interactive models among YSI, SDQc and IQ were non-significant, as were interactions with gender.

DISCUSSION

In this school-based community sample, we first used two analytic approaches to investigate the validity of the children's positive attributes construct. In particular, we were interested in ascertaining the extent to which positive attributes and psychiatric symptoms are distinct constructs. First, confirmatory factor analysis showed that a model with two correlated factors (positive attributes and psychiatric symptoms) fit better than a unidimensional model. Second, propensity score analysis showed that, even after matching participants for psychiatric symptoms, psychiatric disorders, intelligence, and other potential confounders, children with low positive attributes had worse performance in academic subjects than those with high positive attributes. Finally, we found that positive attributes are associated with better educational outcomes both independent of intelligence and psychiatric symptoms, and by buffering associations among low intelligence, high levels of psychiatric symptoms, and negative educational outcomes.

Consistent with other studies,[11,12] our results suggests that positive attributes in children are not merely the absence of psychopathology. Whereas the measurement of psychiatric symptoms might characterize developmental disruptions in children with high levels of psychopathology, the measurement of positive attributes might improve the characterization of behavioral and emotional variability within the normal range, adding incremental health risk prediction.[11,27] This may explain why positive attributes can predict the risk for later psychiatric disorders in healthy children, beyond predictions based on baseline psychiatric symptoms.[11]Additionally, our PSM results revealed that, in groups matched on other relevant characteristics, children high in positive attributes have better academic performance than those low in positive attributes . This is consistent with Krapohl and colleagues,[28] who found that academic performance was predicted not only by intelligence, but also by

personality traits and well-being. Hence, the CFA and PSM analyses supported the validity of the positive attributes construct by improving behavioral characterization and prediction of academic performance.

Most studies examine the predictive value of one variable alone, either positive attributes,[11,12,29,30] intelligence[4,31] or psychiatric symptoms,[32,33] without investigating interactions. In agreement with previous studies, we found that intelligence, psychiatric symptoms and positive attributes did, indeed, have independent associations with educational outcome. However, our study indicates that these variables also interact. Previous studies suggest that early interventions designed to improve noncognitive abilities in disadvantaged children impact on IQ briefly, but have longer- lasting effects on school attainment and employment.[33] Our results suggest that these lasting effects may result from the impact of noncognitive abilities (*i.e.*, positive attributes) on learning. Specifically, based on our findings, it is reasonable to hypothesize that children with low IQ would show particularly marked benefit from early interventions that increase positive attributes, since the impact of low IQ on learning problems is buffered by positive attributes. Also, an association between high positive attributes and lower psychiatric symptoms has been reported,[11] and interventions that improve such noncognitive skills in childhood appear to be associated with decreased psychiatric symptoms later in life.[33,35] While our results are consistent with these previous studies, our study also reveals that, with respect to academic performance, the positive effects of noncognitive abilities might be particularly important in highly symptomatic children, as well as in those with low intelligence. This is especially important given that mental health in adolescence predicts later educational and occupational attainment, rather than background economic and educational status[36].

The interactions that we observed among positive attributes, intelligence and psychiatric symptoms are consistent with developmental theories that focus on adaptive human characteristics.[37] In particular, Heckman's theory of human skills formation[1,7,38] is well-suited to explain the present findings, since it predicts interactions among cognitive skills, noncognitive skills and health.[38] As we observed, positive attributes interact with intelligence and psychiatric symptoms to impact on school learning and performance in children and adolescents, suggesting mechanisms by which these variables can affect on adult outcomes, including educational attainment, employment, crime and

health.[1] The interactions found in our study further suggest that remediation of single domain deficits in a developing child could be important not only for that specific domain, but to potentiate other facets of behavioral function. Considering Vidal-Ribas[11] work and ours, it is plausible to suggest a "noncognitive reserve mechanism" through which positive attributes decrease the odds of developing psychopathology and educational impairments, similar to the "cognitive reserve hypothesis" which proposes that cognitive function acts as a buffer against the development of psychopathology.[31]

Some limitations need to be considered in order to interpret our findings properly. *First*, since this is a cross-sectional study, the possibility of reverse causality (*i.e.,* school factors influencing positive attributes, intelligence and symptoms) cannot be ruled out. However, a previous longitudinal study on positive attributes[11] reported larger effects for positive attributes on psychopathology than those reported here. *Second*, although propensity score matching minimizes the role of potential confounding factors, unobserved variables might introduce residual confounding effects on the associations between YSI and school outcomes and decrease the effect size of positive attributes on reported associations. *Third*, apart from learning problems, which were measured by a standardized test, other child characteristics and outcomes were assessed by parental report, which may have led to effect overestimation. Further studies should include other sources of information such as school reports, test scores, and teacher reports. *Fourth*, this study was carried in a community sample of a single country and the results may not generalize to other cultures.

Taken together, our study provides further validity for the positive attributes construct and suggests that positive attributes may interact with intelligence to predict learning problems, and with psychiatric symptoms to predict academic performance. Importantly, the deleterious associations of psychiatric symptoms and low intelligence are buffered by children's positive attributes. Further studies should focus on understanding the mechanisms mediating these interactions, and on testing mechanistically-informed interventions designed to increase positive attributes, particularly in children with psychiatric symptoms and/or low intelligence.

REFERENCES

1. Heckman JJ, Stixrud J, Urzua S. The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *J Labor Econ*. 2006;24(3):411-482.

2. Heyman GM, Dunn BJ, Mignone J. Disentangling the correlates of drug use in a clinic and community sample: a regression analysis of the associations between drug use, years-of-school, impulsivity, IQ, working memory, and psychiatric symptoms. *Addict Disord Behav Dyscontrol*. 2014;5:70.

3. Kirkcaldy B, Furnham A, Siefen G. The Relationship Between Health Efficacy, Educational Attainment, and Well-Being Among 30 Nations. *Eur Psychol*. 2004;9(2):107-119.

4. Plomin R, Deary IJ. Genetics and intelligence differences: five special findings. *Mol Psychiatry*. 2015;20(1):98-108.

5. Kessler RC, Foster CL, Saunders WB, Stang PE. Social consequences of psychiatric disorders, I: Educational attainment. *Am J Psychiatry*. 1995;152(7):1026-1032.

6. Lee S, Tsang A, Breslau J, et al. Mental disorders and termination of education in high- income and low- and middle-income countries: epidemiological study. *Br J Psychiatry J Ment Sci*. 2009;194(5):411-417.

7. Cunha F, Heckman JJ, Schennach SM. Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica*. 2010;78(3):883-931.

8. Fogel A, King BJ, Shanker SG. *Human Development in the Twenty-First Century*. 1st ed. Cambridge University Press; 2011.

9. Gottlieb G, Wahlsten D, Lickliter R. The Significance of Biology for Human Development: A Developmental Psychobiological Systems View. In: *Handbook of Child Psychology*. John Wiley and Sons, Inc.; 2007.

10. Lerner RM. Developmental Science, Developmental Systems, and Contemporary Theories of Human Development. In: *Handbook of Child Psychology*. John Wiley and Sons, Inc.; 2007.

11. Vidal-Ribas P, Goodman R, Stringaris A. Positive attributes in children and reduced risk of future psychopathology. *Br J Psychiatry J Ment Sci*. 2015;206(1):17-25.

12. Bromley E, Johnson JG, Cohen P. Personality strengths in adolescence anddecreased risk of developing mental health problems in early adulthood. *Compr Psychiatry*. 2006;47(4):315-324.

13. Heckman JJ. Skill Formation and the Economics of Investing in Disadvantaged Children. *Science*. 2006;312(5782):1900-1902.

14. Salum GA, Gadelha A, Pan PM, et al. High risk cohort study for psychiatric disorders in childhood: rationale, design, methods and preliminary results. *Int J Methods Psychiatr Res*. 2015;24(1):58-73.

15. Weissman MM, Wickramaratne P, Adams P, Wolk S, Verdeli H, Olfson M. Brief screening for family psychiatric history: the family history screen. *Arch Gen Psychiatry*. 2000;57(7):675-682.

16. Figueiredo VLM. Uma adaptação brasileira do teste de inteligência WISC-III. . Brasília, DF: Curso de Pós-Graduação em Psicologia. 2001.

17. Stein LM. *TDE Teste de Desempenho Escolar*. São Paulo: Casa do Psicólogo; 1998.

18. Wechsler D. *WISC-III: Escala de Inteligência Wechsler Para Crianças*. 3rd ed. São Paulo: Casa do Psicólogo; 2002.

19. Tellegen A, Briggs PF. Old wine in new skins: grouping Wechsler subtests into new scales. *J Consult Psychol*. 1967;31(5):499-506.

20. Nascimento E do, Figueiredo VLM de. WISC-III and WAIS-III: alterations in the current american original versions of the adaptations for use in Brazil. *Psicol Reflex E Crítica*. 2002;15(3):603-612.

21. Goodman R, Ford T, Simmons H, Gatward R, Meltzer H. Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *Br J Psychiatry J Ment Sci*. 2000;177:534-539.

22. Goodman R, Ford T, Richards H, Gatward R, Meltzer H. The Development and Well- Being Assessment: description and initial validation of an integrated assessment of child and adolescent psychopathology. *J Child Psychol Psychiatry*. 2000;41(5):645-655.

23. Fleitlich-Bilyk B, Goodman R. Prevalence of child and adolescent psychiatric disorders in southeast Brazil. *J Am Acad Child Adolesc Psychiatry*. 2004;43(6):727-734.

24. Cogo-Moreira H, Carvalho CAF, de Souza Batista Kida A, et al. Latent class analysis of reading, decoding, and writing performance using the Academic Performance Test: concurrent and discriminating validity. *Neuropsychiatr Dis Treat*. 2013;9:1175-1185.

25. Ivanova MY, Achenbach TM, Dumenci L, et al. Testing the 8-Syndrome Structure of the Child Behavior Checklist in 30 Societies. *J Clin Child Adolesc Psychol*. 2007;36(3):405- 417.

26. Karline Soetaert. *plot3D: Plotting Multi-Dimensional Data*. New Zeland; 2014. http://cran.r-project.org/web/packages/plot3D/index.html.

27. Israel S, Moffitt TE. Assessing conscientious personality in primary care: an opportunity for prevention and health promotion. *Dev Psychol*. 2014;50(5):1475-1477.

28. Krapohl E, Rimfeld K, Shakeshaft NG, et al. The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. *Proc Natl Acad Sci*. 2014;111(42):15273-15278.

29. Radigan M, Wang R. Relationships between youth and caregiver strengths and mental health outcomes in community based public mental health services. *Community Ment Health J*. 2013;49(5):499-506.

30. Tackett JL. Evaluating models of the personality–psychopathology relationship in children and adolescents. *Clin Psychol Rev*. 2006;26(5):584-599.

31. Koenen KC, Moffitt TE, Roberts AL, et al. Childhood IQ and Adult Mental Disorders: A Test of the Cognitive Reserve Hypothesis. *Am J Psychiatry*. 2009;166(1):50-57.

32. Caspi A, Houts RM, Belsky DW, et al. The p Factor One General Psychopathology Factor in the Structure of Psychiatric Disorders? *Clin Psychol Sci*. 2014;2(2):119-137.

33. Dodge KA, Bierman KL, Coie JD, et al. Impact of early intervention on psychopathology, crime, and well-being at age 25. *Am J Psychiatry*. 2015;172(1):59-70.

35. Frenkel TI, Fox NA, Pine DS, Walker OL, Degnan KA, Chronis-Tuscano A. Early childhood behavioral inhibition, adult psychopathology and the buffering effects of adolescent social networks: a twenty-year prospective study [published online ahead of print February 2015]. *J Child Psychol Psychiatry*. DOI:10.1111/jcpp.12390.

36. Slominski L, Sameroff A, Rosenblum K, Kasser T. Longitudinal predictors of adult socioeconomic attainment: the roles of socioeconomic status, academic competence, and mental health. *Dev Psychopathol*. 2011;23(1):315-324.

37. Rider C k S and EA. *Life-Span Human Development 7th Edition*. 7th edition. Australia ; Belmont, CA: Wadsworth Cengage Learning; 2009.

38. Heckman JJ. The economics, technology, and neuroscience of human capability formation. *Proc Natl Acad Sci*. 2007;104(33):13250-13255.

Table 1. Univariate, bivariate and interactive models of Positive Attributes and Intelligence on school outcomes

		Learning Problems[a]	**Poor Academic Performance**[a]
	z-score[b]	OR (LB – UB)	β (LB – UB)
Model 1 (Univariate)	YSI	0.78 *** (0.70 to 0.87)	-0.31*** (-0.34 to -0.27)
	IQ	0.60*** (0.52 to 0.68)	-0.22*** (-0.26 to -0.18)
Model 2 (Bivariate)	YSI	0.81*** (0.73 to 0.91)	-0.29*** (-0.32 to -0.25)
	IQ	0.61*** (0.53 to 0.70)	-0.19*** (-0.23 to -0.15)
Model 3 (Interactive)	YSI	0.86* (0.76 to 0.97)	-0.28*** (-0.32 to -0.25)
	IQ	0.62*** (0.55 to 0.71)	-0.19*** (-0.22 to -0.15)
	YSI*IQ	1.16* (1.02 to 1.32)	0.02 (-0.02 to 0.06)

Note: YSI = Youth Strengths Inventory; IQ = estimated intelligence quotient (defined in the text); OR = odds ratio; β = regression coefficient β; UB = upper bound; LB = lower bound. *p-value≤0.05; **p-value≤0.01; ***p-value≤0.001.

a. Outcomes defined in the text.

b. The 1st z-score was used as a reference for each independent variable. Estimates reflect the additive OR or β increase associated with changing one z-score.

Table 2. Univariate, bivariate and interactive models of Positive Attributes and Psychiatric Symptoms on school outcomes

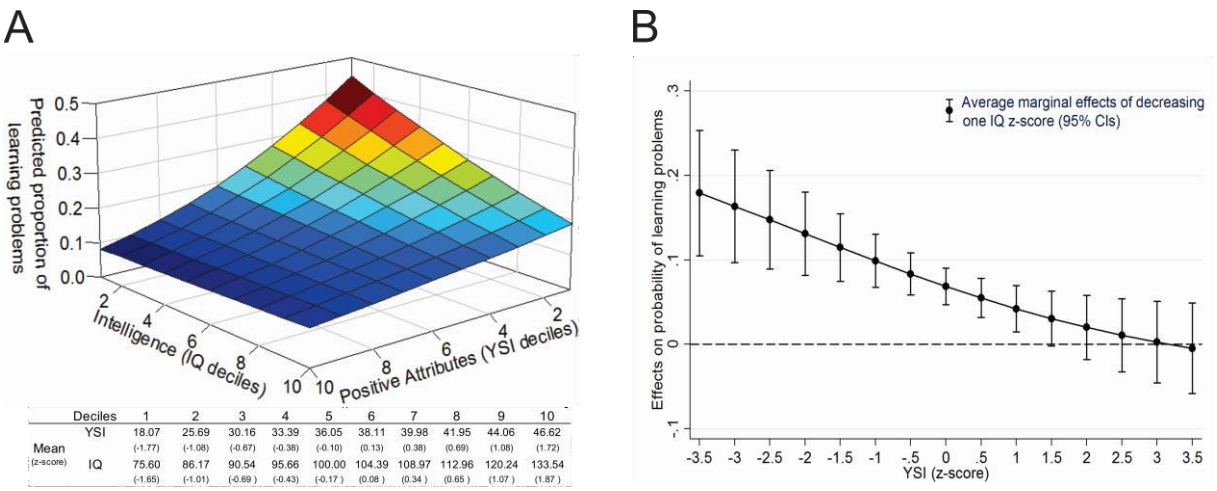
		Learning Problems[a]	**Poor Academic Performance**[a]
	z-score[b]	OR (LB – UB)	β (LB – UB)
Model 1 (Univariate)	YSI	0.78 *** (0.70 to 0.87)	-0.31*** (-0.34 to -0.27)
	SDQc	1.27*** (1.14 to 1.42)	0.30*** (0.26 to 0.34)
Model 2 (Bivariate)	YSI	0.84* (0.73 to 0.96)	-0.20*** (-0.25 to -0.16)
	SDQc	1.15* (1.00 to 1.32)	0.19*** (0.14 to 0.23)
Model 3 (Interactive)	YSI	0.83** (0.72 to 0.95)	-0.20*** (-0.25 to -0.16)
	SDQc	1.18* (1.02 to 1.35)	0.18*** (0.14 to 0.22)
	YSI*SDQc	1.10 (0.98 to 1.24)	-0.06*** (-0.10 to -0.03)

Note: YSI = Youth Strengths Inventory; SDQc = composite of Strengths and Difficulties Questionnaire (defined in the text); OR = odds ratio; β = regression coefficient β; UB = upper bound; LB = lower bound. *p-value≤0.05; **p-value≤0.01; ***p-value≤0.001.

a. Outcomes were defined in the text.

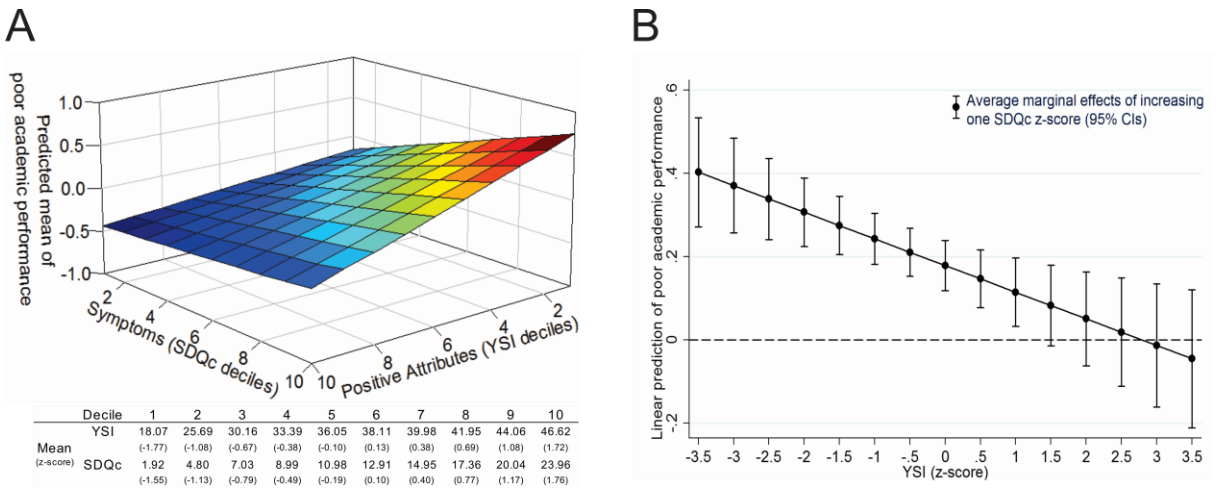
b. The 1st z-score was used as a reference for each independent variable. Estimates reflect the additive OR or β increase associated with changing one z-score.

Figure 1 – Interaction and Marginal Effects of Intelligence and Positive Attributes on Learning Problems



Note: (A) The y-axis represents the probability of learning problems by deciles of intelligence (x-axis) and positive attributes (z-axis). (B) The y-axis represents the probability of learning problems (defined in the text), quantified by the average marginal effect of decreasing one IQ z-scores (black dots with CIs) at each YSI z-scores (x-axis). CIs = Confidence Intervals; YSI = Youth Strengths Inventory; IQ = estimated Intelligence Quotient (defined in the text).

Figure 2 – Interaction and Marginal Effects of Psychiatric Symptoms and Positive Attributes on Poor Academic Performance



Mean (z-score)	Decile	1	2	3	4	5	6	7	8	9	10
	YSI	18.07	25.69	30.16	33.39	36.05	38.11	39.98	41.95	44.06	46.62
		(-1.77)	(-1.08)	(-0.67)	(-0.38)	(-0.10)	(0.13)	(0.38)	(0.69)	(1.08)	(1.72)
	SDQc	1.92	4.80	7.03	8.99	10.98	12.91	14.95	17.36	20.04	23.96
		(-1.55)	(-1.13)	(-0.79)	(-0.49)	(-0.19)	(0.10)	(0.40)	(0.77)	(1.17)	(1.76)

Note: (A) The y-axis represents the mean of poor academic performance by deciles of psychiatric symptoms (x-axis) and positive attributes (z-axis), (B) The y-axis represents the linear prediction of poor academic performance (defined in the text), quantified by the average marginal effect of increasing one SDQc z-score (black dots with CIs) at each YSI z-scores (x-axis). CIs = Confidence Intervals; YSI = Youth Strengths Inventory; SDQc = composite of Strengths and Difficulties Questionnaire (defined in the text).

Supplementary Material

Supplementary methods, analysis and results

Post-hoc power analysis

Post-hoc power analyses were conducted for our main outcomes. For our linear outcomes (academic performance), the observed power for the main effects of Youth Strengths Inventory (YSI) and Strengths and Difficulties Questionnaire composite (SDQc), and for their interaction, were >0.99,
>0.99 and >0.95 respectively. For our binary outcome (learning problems), observed power for the main effects of YSI and SDQc, and for their interaction, were all >0.99.

Factor analysis from YSI and CBCL school items

For all confirmatory factor analysis (CFA), we used delta parameterization and weighted least square using a diagonal weight matrix with standard errors and mean- and variance-adjusted chi- square test statistics (WLSMV) estimators, using MPLUS 7.1 software (Muthén & Muthén, Los Angeles, California, USA). Model fit parameters were Chi Square Test of model fit, root mean square error of approximation (RMSEA), Comparative Fit Index (CFI) and Tucker Lewis Index (TLI). Values of RMSEA near or below 0.08 represent acceptable model fit, and values lower than 0.06 represent good-to-excellent model fit.[1] CFI and TLI values near or above 0.90 represent acceptable model fit, while values higher than 0.95 represent a good-to-excellent model fit. Nested models were tested using Chi-Square for Differences using the DIFFTEST option.
YSI

The YSI is a 24-item scale, divided into two blocks of questions addressed to the caregiver. One block focuses on characteristics of the child, such as if he/she is “lively”, “easy going”, “grateful”, “responsible”, and has a “good sense of humour”. The other block addresses the child’s actions that please others, such as “helps around the home”, “well behaved”, “keeps bedroom tidy”, “does homework without reminding” and others. All questions have three possible answers: “No”, “A little”, “A lot”. The CFA of YSI using a one-factor solution resulted in adequate goodness-of-fit indexes in our sample, converging to a single factor denominated “positive attributes” (see main text). The composite YSI scores were derived from saved factor scores from the CFA model (Table S1).

CBCL-school items

For academic performance, the CFA of CBCL-school using one-factor solution resulted in adequate goodness-of-fit indexes in our sample (see main text). The composite CBCL-school (academic performance) scores were derived from saved factor scores from the CFA model (Table S2).

Testing if YSI and SDQc are overlapping constructs

CFA models including the YSI and SDQc was used to test whether the two scales assess the same underlying latent construct. The category threshold indicates the expected value of the latent factor at which there is a $\geq$ 50% probability of endorsing a given category. The mean threshold for each item was computed as the item location on the severity continuum in order to inform the location of the latent trait in which items were more informative.

CFA models were run to test whether the two scales assess the same underlying latent construct. We fitted a one-factor model (all items loading into a general component), a correlated two- factor model with SDQc items loading onto a 'psychiatric symptoms' dimension and YSI items loading onto a 'positive attributes' dimension; a second-order model, with 'psychiatric symptoms' and 'positive attributes' loading onto one higher order factor; and a bifactor model, with all items loading into a general factor and residuals loading onto two specific factors – 'psychiatric symptoms' and 'positive attributes'. The model with one factor provided an unacceptable fit to the data according to two out of three fit indexes (see main text) and the model with two correlated factors ('psychiatric symptoms' and 'positive attributes') showed acceptable goodness-of-fit in practically all indices (see main text). Chi- Square Test for Difference Testing one-dimensional vs. correlated two factor models showed advantages of the two-factor correlated model over the one-factor model (χ^2=667.338, df=1, p<0.0001). Second-order and bifactor models were not identified.

An item-level inspection of information curves from CFA of the two-factor correlated model showed that YSI and SDQc provide information in different areas of a common metric (*i.e.*, YSI is better at discriminating among typically developing children, while SDQc is better at discriminating among atypically developing children). Specifically, the mean threshold of SDQc items was -0.19, whereas the mean threshold of YSI items was 0.83 (Figure S1).

Propensity Score Matching Methods

As a stringent test of discriminant validity, we used propensity score matching[2] to verify whether associations between a child's positive attributes and school outcomes are independent of intelligence, psychopathology, and other potential confounders. The analyses were conducted in R, using the PSM[3] and MatchIt[4] packages from R-project.

Before the propensity score matching (PSM) procedure, a latent class analysis (LCA) was performed to create empirically-derived groups with different levels of positive attributes (YSI score). This analysis was conducted in MPLUS 7.1 (Muthén & Muthén, Los Angeles, California, USA). A solution with two classes (FP=97, Loglikelihood=-44513.83, AIC=89221.66, IC=89787.02, ssaBIC=89478.82) showed a high entropy =0.925 and divided the sample into high positive attribute (63.2%) and low positive attribute (36.8%) classes (Figure S2). A solution with three classes showed an intermediate group with moderate level of positive attributes, while one with four classes showed

overlapping classes with no discrimination. A two-class solution was selected to maximize sample size and because of the higher entropy level.

We used the nearest neighbour method for the PSM analysis, with a caliper of 0.25, i.e., the largest allowable difference in propensity score for matched participants was 25%. Before and after matching, we used a measure of standardized bias to assess the balance of the covariates. Standardized differences of means <0.20 are acceptable and differences <0.10 are considered negligible.

The PSM procedure selected a total of 671 children with low positive attributes who were matched 1:1 with children with high positive attributes, as described in Methods. By this method, we were able to successfully reduce the magnitude of differences (standardized bias) between children with high and low positive attributes. The mean standardized bias for all covariates is shown in Figure S3.

References:

1. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Model Multidiscip J*. 1999;6(1):1-55.
2. Heckman JJ, Ichimura H, Todd P. Matching As An Econometric Evaluation Estimator. *Rev Econ Stud*. 1998;65(2):261-294.
3. Stig Bousgaard Mortensen, Søren Klim. *PSM: Non-Linear Mixed-Effects Modelling Using Stochastic Differential Equations.*; http://www.imm.dtu.dk/psm. Published September 10, 2013. Accessed August 1, 2014.
4. Daniel Ho, Kosuke Imai, Gary King, Elizabeth A. Stuart. MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *J Stat Softw*. 2011;42:1-28.

Supplementary Figure S1

Figure S1

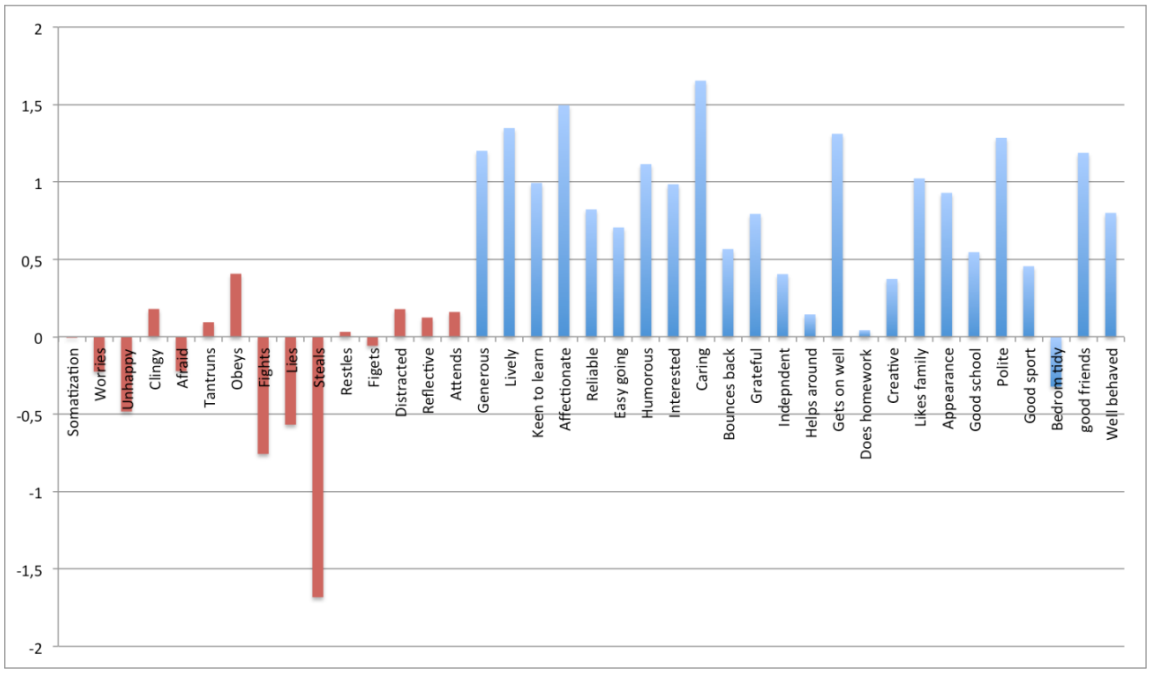


Figure S1: Standardized average thresholds of each item of the Strengths and Difficulties items (SDQc in red) and Youth Strengths Inventory items (YSI in blue).

Supplementary Figure S2

Figure S2

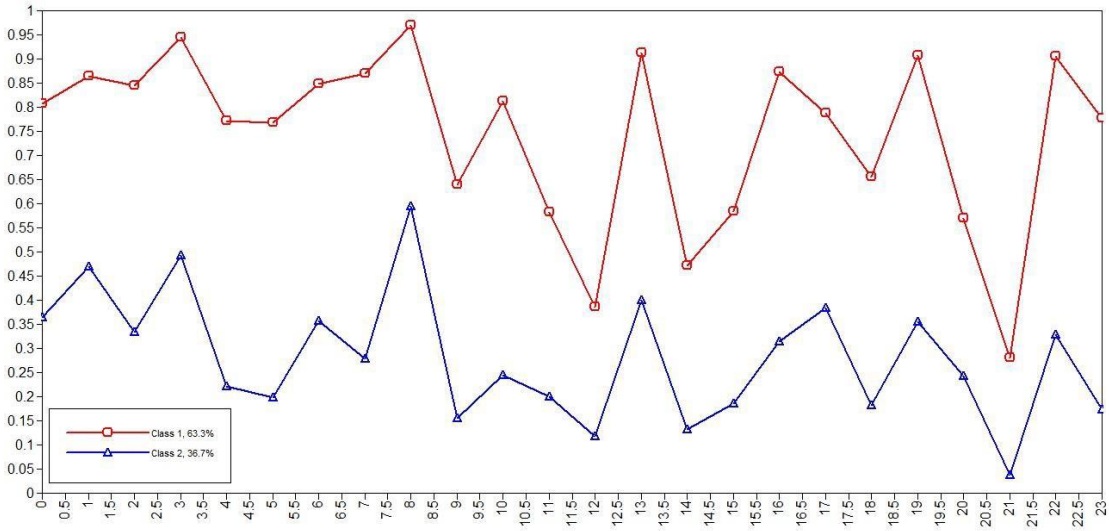


Figure S2: In red, Higher YSI score class, in blue, Lower YSI score class. Graph represents the chance of endorsement (Y axis) of each item of the YSI (X axis).

Supplementary Figure S3

Figure S3

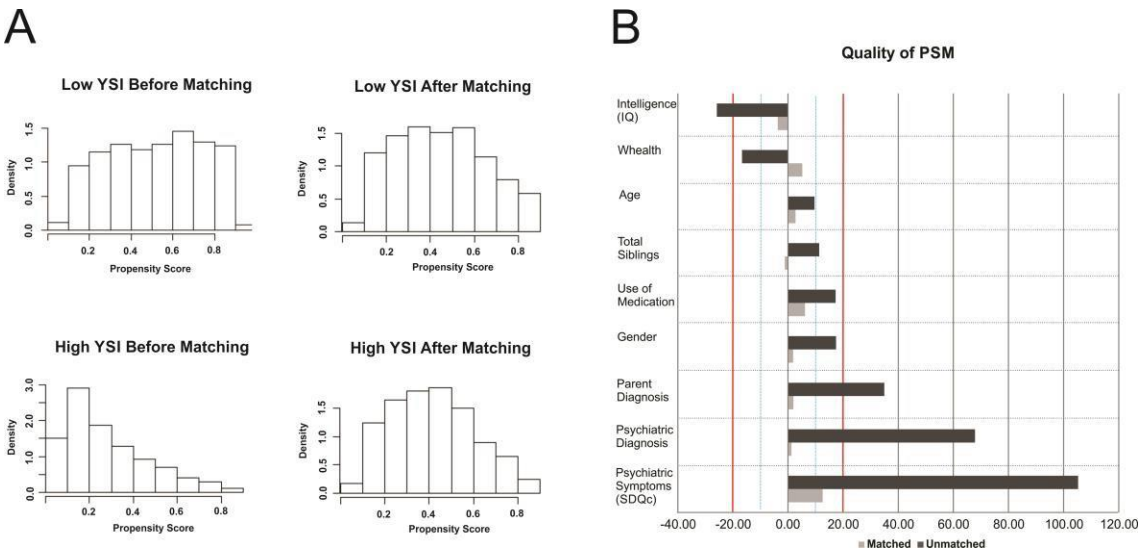


Figure S3: (A) Histograms of propensity score matching (PSM) of High YSI and Low YSI before and after matching and (B) standardized bias (%) of covariates before and after matching. Blue line represents 10% standardized bias limit; below the blue line was considered negligible. Red line represents 20% limit of standardized bias; below the red line was considered acceptable.

Supplementary Table S1

Table S1. Confirmatory Factor Analysis of Youth Strengths Inventory

	Factor Loadings	SE	Thresholds	
			B1	B2
Generous	0.598	0.019	-2.031	-0.371
Lively	0.620	0.019	-2.117	-0.579
Keen to learn	0.677	0.016	-1.581	-0.406
Affectionate	0.753	0.016	-2.222	-0.770
Reliable and responsible	0.740	0.013	-1.469	-0.176
Easy going	0.746	0.012	-1.264	-0.148
Good fun, good sense of humour	0.698	0.015	-1.796	-0.434
Interested in many things	0.759	0.013	-1.577	-0.393
Caring, kind-hearted	0.777	0.018	-2.343	-0.965
Bounces back quickly after setbacks	0.654	0.015	-1.229	0.096
Grateful, appreciative of what he gets	0.761	0.012	-1.324	-0.263
Independent	0.535	0.017	-0.954	0.145
Helps around the home	0.438	0.020	-0.852	0.563
Gets on well with the rest of the family	0.762	0.015	-2.024	-0.597
Does homework without needing to be reminded	0.514	0.018	-0.478	0.393
Creative activities: art, acting, music, making things	0.571	0.017	-0.903	0.156
Likes to be involved in family activities	0.740	0.014	-1.609	-0.436
Takes care of his appearance	0.577	0.019	-1.502	-0.357
Good at school work	0.618	0.016	-1.139	0.046
Polite	0.779	0.014	-2.031	-0.539
Good at sport	0.458	0.02	-1.036	0.125
Keep his bedroom tidy	0.53	0.018	-0.23	0.874
Good with friends	0.773	0.013	-1.871	-0.505
Well behaved	0.763	0.012	-1.46	-0.140

Note: Errors of the following item were correlated in the model: *Good at School* with *Keen to Learn* (r=0.278), *Does homework without need to be reminded* (r=0.399) and *Creative activities* (r=0.212). *Good fun/humour* with *Lively* (r=0.353). *Interested in many things* with *Keen to learn* (r=0.251). *Caring/Kind-hearted* with *Affectionate* (r=0.215) and *Generous* (r=0.204). *Keep his/her bedroom tidy* with *Helps around*(r=0.272) and *Does homework without need to be reminded* (r=0.208). *Well behaved* with *Polite* (r=0.178). *Affectionate* with *Generous* (r=0.223). *Creative activities* with *Does homework without need to be reminded* (r=0.249).

Supplementary Table S2

Table S2. Confirmatory Factor Analysis of Performance in Academic Subjects from Child Behaviour Checklist

	Factor Loadings	SE	Thresholds		
			B1	B2	B3
Portuguese/Literature	0.876	0.006	-1.421	-0.779	0.898
History/Social Studies	0.904	0.005	-1.563	-0.978	0.999
Mathematics	0.690	0.012	-1.484	-0.732	0.721
Science	0.887	0.005	-1.610	-1.023	0.978
Geography	0.928	0.004	-1.591	-1.034	1.034
English/Spanish	0.735	0.015	-1.484	-0.940	0.957
Computer course	0.662	0.024	-1.844	-1.429	0.696
Biology	0.888	0.015	-1.259	-0.891	1.091

Note: Errors of the following item were correlated in the model: *English/Spanish with Biology* (0.198), *Computer course with Biology* (0.170), *English/Spanish with Computer course* (0.202).