Trajectories, cognitive mechanisms, and treatment response of irritability in children and adolescents

Vidal-Ribas Belil, Pablo

Awarding institution:
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

END USER LICENCE AGREEMENT

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International licence. https://creativecommons.org/licenses/by-nc-nd/4.0/

You are free to:
• Share: to copy, distribute and transmit the work

Under the following conditions:
• Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
• Non Commercial: You may not use this work for commercial purposes.
• No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Trajectories, cognitive mechanisms, and treatment response of irritability in children and adolescents

Pablo Vidal-Ribas Belil

Department of Child and Adolescent Psychiatry
Institute of Psychiatry, Psychology & Neuroscience, King’s College London

Thesis submitted for the degree of Doctor of Philosophy
2018
Abstract

Irritability is one of the most common reasons for referral to child and adolescent mental health services and is a strong predictor of current and future functional impairment. Furthermore, irritability is one of the few psychiatric symptoms that is present in both internalising and externalising disorders. In children, irritability is part of the diagnostic criteria of every mood disorder described in the DSM-5, as well as of oppositional defiant disorder (ODD). It is also listed as an associated symptom for conduct disorder (CD), attention-deficit/hyperactivity disorder (ADHD) and autism spectrum disorders (ASD). Recently, irritability as a disorder itself has been introduced in the DSM under the term disruptive mood dysregulation disorder (DMDD). Despite the omnipresence of irritability across distinct psychiatric disorders, we still know little about its predictors, longitudinal outcomes, mechanisms, and, most importantly, effective treatment approaches.

In this thesis, I examine the predictors and longitudinal correlates of irritability in young people. I also test candidate cognitive mechanisms and treatments for irritability driven by results of previous research. Specifically, I examine the potential mechanisms that might explain the specific association between irritability and depression found in longitudinal studies and test a new treatment approach motivated by this association.

First, I present a systematic review and meta-analysis of the longitudinal correlates of chronic severe irritability. I searched for studies in PubMed and Web of Science that investigated DMDD or the irritability dimension of ODD as a predictor of future psychiatric disorders and symptoms. The findings from this study highlight the specific association between chronic severe irritability and future depressive and anxiety disorders.
Second, I examine early predictors of irritability that might explain its specific association with depression. Using a large community-based sample, I tested whether an infant’s sex and variation in physiological stress reactivity (i.e., two known risk factors for depression) interact to predict different patterns of ODD symptom dimensions, including headstrong and irritability, in the pre-school age. We predicted that girls would be at higher risk for irritability whereas boys would be at higher risk for headstrong symptoms. The results show that stress reactivity predicts ODD symptoms in opposite directions in boys and girls. However, this pattern was the same for both ODD dimensions, without specific risk for girls and boys.

Third, using both cross-sectional and longitudinal analyses, I test whether deficits in emotion recognition, commonly seen in youth with irritability and youth with depression, are associated with an increase in the risk of depressive symptoms in children and adolescents with DMDD. The results show that emotion recognition deficits in youth with DMDD are associated with co-occurring depressive symptoms and predict these symptoms at follow-up.

Fourth, I provide the results of the first-ever randomised controlled pharmacological trial (RCT) that has used irritability as a primary outcome in youth with severe mood dysregulation (SMD), the precursor of DMDD. Given the specific association between irritability and depression/anxiety, I examine whether adding citalopram to methylphenidate improves irritability to a greater extent than adding placebo. The results of the RCT show that add-on citalopram improves irritability in youth with SMD. However, improvements of irritability are not accompanied by improvements in functional impairment, depressive symptoms or anxiety symptoms.
Finally, I propose a model to explain the association between irritability and future depressive disorder. To do so, I refer to both findings reported in this thesis and results from previous works. The model also includes aspects that are hypothesised to play a role in this transition but still need to be tested. The hypothesised model is based on a shared factors model, in which factors contributing to irritability are shared with those contributing to depression. I discuss genetic, environmental and early life factors, as well as aberrant responses to reward and emotional stimuli. I conclude this thesis by proposing some directions for future research.
Statement of Work

This thesis contains one meta-analysis (Chapter 2) and three empirical studies (Chapters 3, 4, and 5). The empirical studies employed data from two independent studies.

Chapter 3 used data from the Wirral Child Health and Development Study (WCHADS) led by Professor Jonathan Hill from the University of Reading and Dr Helen Sharp from the University of Liverpool, with developmental and statistical expertise from my supervisor Professor Andrew Pickles at King’s College London. This study was funded by grants from the UK Medical Research Council (G0400577 and G0900654). Chapters 4 and 5 used data from a study conducted at the Emotion and Development Branch, National Institute of Mental Health (NIMH), National Institutes of Health (NIH, Bethesda, Maryland US). This study aims to examine the characterisation and pathophysiology of severe mood and behavioural dysregulation in youth and was supported by the NIMH Intramural Research Program (ZIAMH002786-15, ZIAMH002778-17). The study in Chapter 4 was conducted under the NIH Clinical Study Protocol 02-M-0021 (ClinicalTrials.gov identifier: NCT00025935) and is led by Dr Ellen Leibenluft. The study in Chapter 5 was conducted under the NIH Clinical Study Protocol 09-M-0034 (ClinicalTrials.gov identifier: NCT00794040) and was led by my supervisor Dr Argyris Stringaris.

The research questions for the meta-analysis in Chapter 2 were developed by myself with the guidance of Dr Argyris Stringaris. I was responsible for conducting the literature search, as well as analysing and interpreting the data for this study, and writing up the publication for the study, under the supervision of Dr Stringaris and with feedback from co-authors, Dr Ellen Leibenluft, Dr Melissa A. Brotman, and Ms Isabel Valdivieso.
The data collection or pre-processing of raw data for the studies presented in Chapter 3, 4, and 5 was conducted by members of the WCHADS (Chapter 3) and the Emotion and Development Branch at the NIMH (Chapters 4 and 5) before the start of my PhD. However, I developed the research questions for the study in Chapter 3 under the guidance of Professor Andrew Pickles and Professor Jonathan Hill. I was also responsible for analysing and interpreting the data for the study in Chapter 3, with analytical support from Professor Andrew Pickles. I was responsible for writing up and disseminating the results in peer-reviewed publication, under the supervision of Professor Andrew Pickles, Professor Jonathan Hill, and Dr Helen Sharp, and with feedback from co-authors.

Similarly, for the study in Chapter 4, I developed the research questions and retrieved the relevant data to examine those questions from available electronic databases. I was also responsible for analysing and interpreting the results, and writing up the publication of the study, under the supervision of Dr Argyris Stringaris, and with feedback from co-authors, especially Dr Melissa A. Brotman and Dr Ellen Leibenluft.

Finally, for the randomised controlled trial (RCT) presented in Chapter 5, I was responsible for developing, with the supervision of Professor Andrew Pickles, the Statistical Analysis Plan after the end of data collection, but prior to the group-label unblinding. I also analysed and interpreted the results and wrote up the manuscript to be published in a peer-reviewed journal, under the supervision of Dr Kenneth Towbin and Dr Argyris Stringaris, with feedback from co-authors, especially Dr Melissa A. Brotman, Dr Ellen Leibenluft, Professor Andrew Pickles, and Dr Daniel S. Pine.

All publications resulting from this thesis are a direct product of my own work, achieved with the supervision of Professor Andrew Pickles and Dr Argyris Stringaris and feedback
from co-authors. When referring to the methods and results of individual studies included in this thesis, I use the third person ("we", "our") for consistency with published articles. This thesis represents my own, original work.
List of publications and presentations relevant to this thesis

Publications

Parts of Chapter 1 are adapted from:


Chapter 2 is adapted from:

Chapter 3 is adapted from:


Chapter 4 is adapted from:


Chapter 5 is adapted from:


**Presentations**


*Jointly first author.
Acknowledgements

First and foremost, I would like to thank my supervisors, Professor Andrew Pickles and Dr Argyris Stringaris, for giving me the opportunity to learn from the best in their fields, and for being a source of support and encouragement throughout the years. Thank you to Andrew for his guidance, insightful scientific contribution and statistical expertise; it has been a privilege working and learning first-hand statistical knowledge with him. Argyris’s ability to combine outstanding academic and clinical work, unwavering enthusiasm and passion for science and inquiry, and simultaneous engagement in research/clinical leadership, has been a true inspiration to me. I am so grateful for how Andrew and Argyris have challenged and supported me to reach my potential, and in a very patient way!

I thank the Institute of Psychiatry, Psychology, and Neuroscience (IoPPN) and the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King’s College London for the award that allowed me to undertake the research contained in this thesis. I also thank the Department of Child and Adolescent Psychiatry at the IoPPN for kindly providing me with a workspace and facilities for my research. In addition, I thank the Intramural Research Program at the National Institute of Mental Health (NIMH), National Institutes of Health (NIH), and the Emotion and Development Branch (Bethesda, Maryland, USA) for proving me with a supplemental predoctoral award to undertake part of my graduate training at their facilities.

I would also like to thank everyone who collaborated with me on the studies included in this thesis. First, I thank the team for the Wirral Child Health and Development Study (First Steps), in particular, Professor Jonathan Hill, Dr Helen Sharp, Dr Florin Tibu, Mr Matthew Bluett-Duncan, and Mr Stuart Kehl. Second, I highly appreciate the insight from the team
at the Emotion and Development Branch at NIH, with a special thanks to Dr Kenneth Towbin, Dr Ellen Leibenluft, Dr Melissa A. Brotman, and Dr Daniel S. Pine.

My colleagues and friends at the Mood and Development Lab and IoPPN – Dr Selina Wolke, Dr Lorena Fernández de la Cruz, Dr Laia Villalta, Dr Aleksa Kaurin, Dr Nina Mikita, Dr Simone Pisano, Dr Marta Vila and Ms Hazel Deacon – thank you for the advice and support. Also, to my colleagues and friends at the Mood Brain and Development Unit (NIMH) – Dr Hanna Keren, Dr Georgia O’Callaghan, Dr Narun Pornpattananangkul, Ms Catarina Farinha, Ms Ariela Kaiser, Ms Liana Meffert, Ms Katie Miller, Ms Aria Vitale, and all the remaining MBDU team of research assistants and clinicians – thank you so much for giving me the strength to finish my PhD when I most needed it. My huge thanks go to my fondest friend Steve Lukito, who has helped to proof-read the bulk of this thesis, for all your support through these years, and for being always there for me.

To my parents, Carlos, Montse, Gianni, and my brothers and their families, Carlos, Marta, Manu, Maria, Mariona, Guillem (who was born shortly after this thesis defense), Guille and Paula. Also, to my cousins Carlota and Maria, and my goddaughter Manuela. Your love, powerful encouragement, and unconditional support have always been there with me, irrespective of the thousand miles between us. I would also like to thank Marta for her patience throughout these years.

Finally, the biggest thanks go to my son Martí, whose love has been the engine that made me keep moving forward. You have been in my mind from the very first thought to the very last word of this thesis; I would not have done this without you by my side.
Table of Contents

Abstract ................................................................................................................................. 3
Statement of Work .................................................................................................................. 6
List of publications and presentations relevant to this thesis ........................................... 9
Acknowledgements .............................................................................................................. 12
Table of Contents ................................................................................................................ 14
List of Tables ........................................................................................................................ 24
List of Figures ....................................................................................................................... 25
Abbreviations ....................................................................................................................... 29

CHAPTER 1: Introduction ...................................................................................................... 34

1.1 The importance of studying irritability in children and adolescents ...................... 34
1.2 A conceptual background of irritability ..................................................................... 35
  1.2.1 Working definitions of irritability ........................................................................... 35
    1.2.1.1 An introduction to DMDD ............................................................................... 36
    1.2.1.2 An introduction to the irritability ODD dimension/subtype .............................. 37
  1.2.2 Features of irritability ............................................................................................ 38
    1.2.2.1 Episodic vs chronic irritability ....................................................................... 38
    1.2.2.2 Phasic vs tonic irritability .............................................................................. 40
    1.2.2.3 Normative vs pathological irritability ............................................................ 40
  1.2.3 Irritability and related constructs .......................................................................... 42
1.3 Irritability constructs in the current thesis ................................................................. 45
  1.3.1 Irritability as DMDD .............................................................................................. 46
    1.3.1.1 Phenotypic presentation .................................................................................. 46
    1.3.1.2 The need for a psychiatric diagnosis with irritability as a primary symptom 48
      1.3.1.2.1 The controversy regarding paediatric BD in youth ................................. 48
1.3.1.2.2  Why ODD was not enough

1.3.1.3  Distinguishing chronic vs episodic irritability

1.3.1.4  Epidemiology

1.3.1.5  Psychopathological correlates

1.3.1.6  Differential diagnosis

1.3.2  Irritability as ODD dimension/subtype

1.3.2.1  Phenotypic presentation

1.3.2.2  Epidemiology

1.3.2.3  The multidimensionality of ODD

1.3.2.4  Psychopathological correlates

1.4  Irritability in the context of other psychiatric disorders

1.4.1  Irritability in Bipolar disorders

1.4.2  Irritability in Depressive disorders

1.4.3  Irritability in Anxiety disorders

1.4.4  Irritability in ADHD

1.4.5  Irritability in ASD

1.5  Assessment and measurement of irritability
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.5.1 Reliability in the measurement of irritability</td>
<td>75</td>
</tr>
<tr>
<td>1.5.2 Instruments for assessing irritability</td>
<td>79</td>
</tr>
<tr>
<td>1.6 Pathophysiological mechanisms of irritability</td>
<td>82</td>
</tr>
<tr>
<td>1.6.1 Genetic and family studies</td>
<td>83</td>
</tr>
<tr>
<td>1.6.2 Behavioural and neuroimaging mechanisms</td>
<td>85</td>
</tr>
<tr>
<td>1.6.2.1 Aberrant response to frustrative non-reward</td>
<td>86</td>
</tr>
<tr>
<td>1.6.2.2 Aberrant approach to threat</td>
<td>87</td>
</tr>
<tr>
<td>1.6.3 Environmental factors</td>
<td>88</td>
</tr>
<tr>
<td>1.6.4 Early predictors of irritability in children and adolescents</td>
<td>90</td>
</tr>
<tr>
<td>1.6.5 Physiological correlates of irritability</td>
<td>92</td>
</tr>
<tr>
<td>1.7 Treatment approaches to irritability</td>
<td>93</td>
</tr>
<tr>
<td>1.7.1 Psychological treatments</td>
<td>94</td>
</tr>
<tr>
<td>1.7.1.1 Parent management training (PMT)</td>
<td>95</td>
</tr>
<tr>
<td>1.7.1.2 Cognitive behavioural therapy (CBT)</td>
<td>96</td>
</tr>
<tr>
<td>1.7.1.3 Development of psychological treatments for severe irritability</td>
<td>96</td>
</tr>
<tr>
<td>1.7.2 Pharmacological treatments</td>
<td>98</td>
</tr>
<tr>
<td>1.8 The Aims of this Thesis</td>
<td>102</td>
</tr>
<tr>
<td>1.8.1 Longitudinal outcomes of irritability</td>
<td>103</td>
</tr>
<tr>
<td>1.8.2 Early predictors of irritability</td>
<td>104</td>
</tr>
<tr>
<td>1.8.3 Cognitive mechanism of irritability</td>
<td>106</td>
</tr>
<tr>
<td>1.8.4 Treatment outcomes of irritability</td>
<td>107</td>
</tr>
</tbody>
</table>

**CHAPTER 2: Longitudinal correlates of chronic severe irritability – a meta-analysis**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Background</td>
<td>109</td>
</tr>
<tr>
<td>2.1.1 Aims and hypotheses</td>
<td>111</td>
</tr>
<tr>
<td>2.2 Methods</td>
<td>111</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>2.2.1 Data sources and search strategies</td>
<td>111</td>
</tr>
<tr>
<td>2.2.2 Inclusion and exclusion criteria</td>
<td>112</td>
</tr>
<tr>
<td>2.2.3 Study selection</td>
<td>112</td>
</tr>
<tr>
<td>2.2.4 Data extraction</td>
<td>113</td>
</tr>
<tr>
<td>2.2.5 Data analysis</td>
<td>115</td>
</tr>
<tr>
<td>2.2.5.1 Estimation of pooled effect sizes</td>
<td>115</td>
</tr>
<tr>
<td>2.2.5.2 Sensitivity analysis</td>
<td>115</td>
</tr>
<tr>
<td>2.2.5.3 Heterogeneity analysis</td>
<td>115</td>
</tr>
<tr>
<td>2.2.5.4 Publication bias</td>
<td>116</td>
</tr>
<tr>
<td>2.3 Results</td>
<td>116</td>
</tr>
<tr>
<td>2.3.1 Characteristics of studies for the meta-analysis</td>
<td>116</td>
</tr>
<tr>
<td>2.3.2 Irritability as a predictor of psychiatric disorders</td>
<td>126</td>
</tr>
<tr>
<td>2.3.2.1 Depressive Disorder</td>
<td>126</td>
</tr>
<tr>
<td>2.3.2.2 Anxiety Disorders</td>
<td>128</td>
</tr>
<tr>
<td>2.3.2.3 Bipolar Disorder</td>
<td>128</td>
</tr>
<tr>
<td>2.3.2.4 Attention Deficit Hyperactivity Disorder</td>
<td>129</td>
</tr>
<tr>
<td>2.3.2.5 Conduct Disorder</td>
<td>130</td>
</tr>
<tr>
<td>2.3.2.6 Oppositional Defiant Disorder</td>
<td>130</td>
</tr>
<tr>
<td>2.3.2.7 Substance Abuse/Dependence</td>
<td>131</td>
</tr>
<tr>
<td>2.3.2.8 Covariates analysis in the prediction of psychiatric disorders</td>
<td>132</td>
</tr>
<tr>
<td>2.3.2.9 Test of publication bias for the prediction of psychiatric disorders</td>
<td>133</td>
</tr>
<tr>
<td>2.3.3 Irritability as a predictor of psychiatric symptoms</td>
<td>136</td>
</tr>
<tr>
<td>2.3.3.1 Depressive Symptoms</td>
<td>136</td>
</tr>
<tr>
<td>2.3.3.2 Anxiety Symptoms</td>
<td>138</td>
</tr>
<tr>
<td>2.3.3.3 Attention Deficit Hyperactivity Disorder Symptoms</td>
<td>138</td>
</tr>
<tr>
<td>2.3.3.4 Conduct Disorder Symptoms</td>
<td>139</td>
</tr>
</tbody>
</table>
2.3.3.5 Oppositional Defiant Disorder Symptoms .................................................. 139
2.3.3.6 Covariates analysis in the prediction of psychiatric symptoms ........... 140
2.3.3.7 Test of publication bias for the prediction of psychiatric symptoms.... 142
2.3.4 Descriptive studies .......................................................................................... 144
2.3.5 Irritability as a predictor of functional impairment ........................................ 144
2.4 Discussion ........................................................................................................... 146
2.4.1 Summary of findings ....................................................................................... 146
2.4.2 Association between irritability and future psychiatric outcomes .......... 146
2.4.3 Variety of psychiatric outcomes tested ............................................................ 147
2.4.4 Evidence for publication bias .......................................................................... 148
2.4.5 Factors contributing to differences between studies .................................... 149
2.4.6 Strengths and limitations .................................................................................. 150
2.5 Conclusion ........................................................................................................... 152

CHAPTER 3: Sex differences in the associations between vagal reactivity and ODD symptoms ........................................................................................................ 153
3.1 Background ......................................................................................................... 153
3.1.1 Respiratory sinus arrhythmia as estimator of vagal reactivity .................... 153
3.1.2 Sex differences in the association between vagal reactivity and psychiatric outcomes .............................................................................................................. 155
3.1.3 Distinct symptom dimensions in ODD ............................................................. 157
3.1.4 Irritability in pre-school years ......................................................................... 158
3.1.5 Aims and hypotheses ....................................................................................... 159
3.2 Methods ............................................................................................................... 160
3.2.1 Participants ..................................................................................................... 160
3.2.2 Measures ........................................................................................................ 161
3.2.2.1 Respiratory Sinus Arrhythmia (RSA) ........................................................... 161
3.2.2.2 Oppositional defiant disorder symptoms .................................................... 164
3.2.2.3 Confounders .............................................................................................................. 165
3.2.3 Statistical analyses ........................................................................................................ 167
3.3 Results ............................................................................................................................... 168
3.3.1 Descriptive Statistics .................................................................................................... 168
3.3.2 Confirmatory factor analyses of ODD symptoms ......................................................... 168
3.3.3 SEM applied to the analysis of RSA predicting ODD symptoms ......................... 171
3.3.4 Vagal tone, vagal reactivity and ODD symptoms ...................................................... 172
3.3.5 Vagal tone, vagal reactivity and ODD irritability and headstrong dimensions ......... 173
3.4 Discussion ......................................................................................................................... 176
3.4.1 Summary of findings .................................................................................................. 176
3.4.2 Sex differences in vagal reactivity as a predictor of ODD symptoms .......... 176
3.4.3 Sex differences in vagal reactivity as a predictor of ODD symptom dimensions ................................................................. 178
3.4.4 Sex differences in baseline vagal tone as a predictor of ODD symptoms and dimensions ................................................................. 181
3.4.5 Strengths and limitations .......................................................................................... 182
3.5 Conclusion ......................................................................................................................... 184

CHAPTER 4: Association between deficits in emotion recognition and depressive symptoms in youth with DMDD ............................................................................... 185
4.1 Background ....................................................................................................................... 185
4.1.1 Emotion processing deficits in irritability and depression ...................................... 185
4.1.2 Other factors contributing to emotion recognition: the case of anxiety, age, and sex 187
4.1.3 Aims and hypotheses ................................................................................................. 189
4.2 Methods ........................................................................................................................... 190
4.2.1 Participants ................................................................................................................. 190
4.2.2 Emotion Recognition Task

4.2.3 Statistical analysis

4.2.3.1 Sample characteristics

4.2.3.2 Categorical analysis

4.2.3.3 Dimensional analysis

4.2.3.4 Effects of comorbid anxiety disorders on emotion recognition

4.2.3.5 Effects of age on emotion recognition

4.2.3.6 Effects of sex on emotion recognition

4.2.3.7 Longitudinal analysis

4.3 Results

4.3.1 Sample characteristics

4.3.2 Categorical analysis

4.3.3 Dimensional analysis

4.3.4 Effects of comorbid anxiety disorders on emotion recognition

4.3.5 Effects of age on emotion recognition

4.3.6 Effect of sex on emotion recognition

4.3.7 Longitudinal analysis

4.4 Discussion

4.4.1 Summary of findings

4.4.2 Deficits in emotion recognition in youth with severe irritability are associated with depressive symptoms

4.4.3 Underlying mechanisms that might explain the association between deficits in emotion processing, irritability and depression

4.4.4 Anxiety is not directly related with a deficit in emotion recognition in youth with DMDD

4.4.5 Effects of age and sex on emotion recognition

4.4.6 Clinical implications
4.4.7 Strengths and limitations ................................................................. 213
4.5 Conclusion ...................................................................................... 215

CHAPTER 5: A double-blind randomised controlled trial of adjunctive citalopram in youth with chronic severe irritability treated with stimulant ............................................. 216
5.1 Background .................................................................................... 216
  5.1.1 Psychological approaches to the treatment of chronic severe irritability .... 217
  5.1.2 Pharmacological approaches to the treatment of chronic severe irritability 218
  5.1.3 Aims and hypotheses .................................................................. 220
5.2 Methods ......................................................................................... 221
  5.2.1 Participants ............................................................................... 221
  5.2.2 Inclusion and exclusion criteria ..................................................... 222
  5.2.3 Enrolment ................................................................................. 224
  5.2.4 Treatment of ADHD symptoms .................................................. 225
  5.2.5 Side Effects ............................................................................... 228
  5.2.6 Assessment ............................................................................... 228
  5.2.7 Statistical analysis ..................................................................... 230
5.3 Results ............................................................................................ 232
  5.3.1 Participant sample derivation and characteristics ................................. 232
  5.3.2 Response to open-label stimulant optimisation .................................. 235
    5.3.2.1 Response to open-label stimulant optimization by SMD symptom domains 238
  5.3.3 Response to citalopram versus placebo ........................................ 240
    5.3.3.1 Response to citalopram versus placebo by SMD symptom domains ... 242
    5.3.3.2 Sensitivity analysis using Last Observation Carried Forward ......... 245
  5.3.4 Tolerability and adverse effects .................................................... 246
5.4 Discussion ....................................................................................... 249
  5.4.1 Summary of findings ................................................................... 249
5.4.2 Effects of adjunctive citalopram on irritability ........................................ 249
5.4.3 Effects of adjunctive citalopram on functional impairment, anxiety and depressive symptoms ........................................................................ 251
5.4.4 Tolerability and side effects .................................................................. 252
5.4.5 Strengths and limitations ..................................................................... 252
5.5 Conclusion ............................................................................................. 254

CHAPTER 6: General Discussion ..................................................................... 255
6.1 Summary of findings by study ................................................................. 255
   6.1.1 Longitudinal correlates of irritability ................................................ 255
   6.1.2 Early stress and irritability ................................................................. 257
   6.1.3 Cognitive mechanisms in irritability ............................................... 258
   6.1.4 Treatment response of irritability .................................................... 259
6.2 Methodological commonalities and disagreements ................................. 260
6.3 Limitations of the studies ...................................................................... 265
6.4 Towards a model for the relation between irritability and future depression 269
   6.4.1 Three potential explanations for the relation between irritability and future depression ................................................................. 270
      6.4.1.1 Model of irritability as an early manifestation of depression .... 271
      6.4.1.2 Model of irritability as cause of depression .............................. 272
      6.4.1.3 Model of shared risk factors ..................................................... 275
   6.4.2 Proposed model for the relationship between irritability and future depressive disorder 278
      6.4.2.1 Genetic and environmental risk factors ...................................... 279
      6.4.2.2 Aberrant response to emotional stimuli .................................. 281
      6.4.2.3 Aberrant response to reward .................................................. 282
6.5 Research implications and future research directions .............................. 288
6.6 Final remarks .......................................................................................... 296
List of Tables

Table 1.1  Comparison of diagnostic criteria between Disruptive Mood Dysregulation Disorder (DMDD) and Severe Mood Dysregulation (SMD)  47

Table 2.1  Characteristics of studies included in the meta-analyses  119

Table 2.2  Characteristics of studies predicting psychiatric disorders  120

Table 2.3  Characteristics of studies predicting psychiatric symptoms  123

Table 3.1  Summary of measures in boys and girls  170

Table 3.2  Model fit of ODD symptoms in Confirmatory Factor Analyses  171

Table 4.1  Demographic and clinical characteristics  196

Table 5.1  Sample demographic and clinical characteristics.  234

Table 5.2  Summary of primary and secondary outcomes of the randomised controlled trial with adjunctive citalopram vs. placebo  237

Table 5.3  Adverse effect rates of youth randomly allocated to citalopram or placebo  247
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>Distinct time courses of irritability.</td>
<td>39</td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>The position of irritability within commonly used terminology.</td>
<td>44</td>
</tr>
<tr>
<td>Figure 1.3</td>
<td>Changes in US rates for groups of psychiatric disorders coded as primary diagnoses in acute US inpatient units from 1996-2004 (Data from Blader &amp; Carlson, 2007)</td>
<td>50</td>
</tr>
<tr>
<td>Figure 1.4</td>
<td>US national trends in bipolar disorder visit rates in outpatient services from 1994-2003 (Data from Moreno et al. 2007).</td>
<td>50</td>
</tr>
<tr>
<td>Figure 1.5</td>
<td>Bars with standard errors show the percentage of patients with either severe mood dysregulation (SMD) or bipolar disorder (BD) who developed a (hypo-)manic or mixed episode during the follow-up period. Adapted from Stringaris et al. (2010a).</td>
<td>53</td>
</tr>
<tr>
<td>Figure 1.6</td>
<td>Oppositional defiant disorder dimension models most frequently used and compared in the literature</td>
<td>64</td>
</tr>
<tr>
<td>Figure 1.7</td>
<td>Mechanism contributing to the development and maintenance of irritability in youth.</td>
<td>82</td>
</tr>
<tr>
<td>Figure 1.8</td>
<td>Brain regions involved in the aberrant processing of reward and threat (Adapted from Brotman et al., 2017).</td>
<td>85</td>
</tr>
<tr>
<td>Figure 1.9</td>
<td>Dysfunctional parent-child interactions contribute to the maintenance of irritability.</td>
<td>91</td>
</tr>
<tr>
<td>Figure 1.10</td>
<td>Treatment algorithm for irritability.</td>
<td>94</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>Review and selection of articles for meta-analysis</td>
<td>114</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>Forest plot of irritability as a predictor of future psychiatric disorders</td>
<td>127</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>2.3</td>
<td>Funnel plot with pseudo 95% confidence limits, using data from 11 studies predicting depression (A) and 8 of these studies after excluding outlier estimates (B).</td>
<td></td>
</tr>
<tr>
<td>2.4</td>
<td>Funnel plot with pseudo 95% confidence limits, using data from 12 studies predicting anxiety disorder.</td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>Forest plot of irritability as a predictor of future psychiatric symptoms.</td>
<td></td>
</tr>
<tr>
<td>2.6</td>
<td>Funnel plot with pseudo 95% confidence limits, using data from 14 studies predicting depressive symptoms.</td>
<td></td>
</tr>
<tr>
<td>2.7</td>
<td>Funnel plot with pseudo 95% confidence limits, using data from 11 studies predicting anxiety symptoms.</td>
<td></td>
</tr>
<tr>
<td>3.1</td>
<td>Participant flow diagram.</td>
<td></td>
</tr>
<tr>
<td>3.2</td>
<td>Summary of structural equation modelling analysis of baseline vagal tone and vagal reactivity predicting overall ODD symptoms.</td>
<td></td>
</tr>
<tr>
<td>3.3</td>
<td>Summary of structural equation modelling analysis of baseline vagal tone and vagal reactivity predicting ODD dimensions of irritability and headstrong symptoms.</td>
<td></td>
</tr>
<tr>
<td>3.4</td>
<td>Interplay between sex and respiratory sinus arrhythmia (RSA) reactivity in the prediction of irritability and headstrong ODD symptom dimensions.</td>
<td></td>
</tr>
<tr>
<td>4.1</td>
<td>Number of identification errors by emotion and modality across groups.</td>
<td></td>
</tr>
<tr>
<td>4.2</td>
<td>Misidentification of happy stimuli as other emotions across groups.</td>
<td></td>
</tr>
<tr>
<td>4.3</td>
<td>Misidentification of angry stimuli as other emotions across groups.</td>
<td></td>
</tr>
<tr>
<td>5.1</td>
<td>Diagram of the trial design</td>
<td></td>
</tr>
<tr>
<td>Figure 5.2</td>
<td>CONSORT diagram of the trial.</td>
<td>233</td>
</tr>
<tr>
<td>Figure 5.3</td>
<td>Change in irritability severity before and after randomisation in the sample included in the intent-to-treat analysis.</td>
<td>236</td>
</tr>
<tr>
<td>Figure 5.4</td>
<td>Change in temper outbursts severity before (depicted as the black dotted line) and after randomisation in the sample included in the intent-to-treat analysis.</td>
<td>238</td>
</tr>
<tr>
<td>Figure 5.5</td>
<td>Change in mood between outbursts severity before (depicted as the black dotted line) and after randomisation in the sample included in the intent-to-treat analysis.</td>
<td>239</td>
</tr>
<tr>
<td>Figure 5.6</td>
<td>Change in hyperarousal symptoms severity before (depicted as the black dotted line) and after randomisation in the sample included in the intent-to-treat analysis.</td>
<td>239</td>
</tr>
<tr>
<td>Figure 5.7</td>
<td>Proportion of treatment response by week and treatment group.</td>
<td>240</td>
</tr>
<tr>
<td>Figure 5.8</td>
<td>Proportion of treatment response by week and treatment group for temper outbursts.</td>
<td>243</td>
</tr>
<tr>
<td>Figure 5.9</td>
<td>Proportion of treatment response by week and treatment group for mood between outbursts.</td>
<td>243</td>
</tr>
<tr>
<td>Figure 5.10</td>
<td>Proportion of treatment response by week and treatment group for hyperarousal symptoms.</td>
<td>244</td>
</tr>
<tr>
<td>Figure 6.1</td>
<td>Model of irritability as an early manifestation of depression.</td>
<td>271</td>
</tr>
<tr>
<td>Figure 6.2</td>
<td>Model of irritability as cause of depression.</td>
<td>273</td>
</tr>
<tr>
<td>Figure 6.3</td>
<td>Model of shared risk factors between irritability and depression.</td>
<td>275</td>
</tr>
<tr>
<td>Figure 6.4</td>
<td>Hypothesised model of the relationship between irritability and the development of depressive disorder.</td>
<td>279</td>
</tr>
<tr>
<td>Figure 6.5</td>
<td>Rolls’ theory of emotions</td>
<td>287</td>
</tr>
<tr>
<td>Figure 6.6</td>
<td>Transition from irritability to depression.</td>
<td>292</td>
</tr>
</tbody>
</table>
Abbreviations

ABC Aberrant Behavior Checklist

ACC Anterior cingulate cortex

ADHD Attention deficit hyperactivity disorder

APA American Psychiatric Association

ARI Affective Reactivity Index

ASD Autism spectrum disorder

BCAMHS British Child and Adolescent Mental Health Survey

BD Bipolar disorder

BD-NOS Bipolar disorder – not otherwise specified

CBCL Child Behavior Checklist

CBT Cognitive behavioural therapy

CD Conduct disorder

CDI Children’s Depression Inventory

CFA Confirmatory factor analysis

CGAS Children’s Global Assessment Scale
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CGI-I</td>
<td>Clinical Global Impression – Change and Improvement Scale</td>
</tr>
<tr>
<td>CGI-S</td>
<td>Clinical Global Impression – Severity Scale</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence interval</td>
</tr>
<tr>
<td>COBY</td>
<td>Course and Outcome of Bipolar Illness in Youth study</td>
</tr>
<tr>
<td>CTP</td>
<td>Citalopram</td>
</tr>
<tr>
<td>DBT</td>
<td>Dialectical Behavior Therapy</td>
</tr>
<tr>
<td>DMDD</td>
<td>Disruptive mood dysregulation disorder</td>
</tr>
<tr>
<td>DP</td>
<td>Dysregulation profile</td>
</tr>
<tr>
<td>DSM</td>
<td>Diagnostic And Statistical Manual Of Mental Disorders</td>
</tr>
<tr>
<td>EFA</td>
<td>Exploratory factor analysis</td>
</tr>
<tr>
<td>ERP</td>
<td>Event-related brain potential</td>
</tr>
<tr>
<td>ES</td>
<td>Effect size</td>
</tr>
<tr>
<td>fMRI</td>
<td>Functional magnetic resonance imaging</td>
</tr>
<tr>
<td>GAD</td>
<td>Generalized anxiety disorder</td>
</tr>
<tr>
<td>GSMS</td>
<td>Great Smoky Mountains Study</td>
</tr>
<tr>
<td>HC</td>
<td>Healthy controls</td>
</tr>
<tr>
<td>HV</td>
<td>Healthy volunteers</td>
</tr>
</tbody>
</table>
IBT  Interpretation Bias Training

ICD  International Classification of Diseases

IPT-MBD  Interpersonal Psychotherapy for mood and behaviour dysregulation

K-SADS-PL  Kiddie Schedule for Affective Disorders Present and Lifetime Version

LAMS  Longitudinal Assessment of Manic Symptoms study

LCA  Latent class analysis

MAP-DB  Multidimensional Assessment of Preschool Disruptive Behavior

MDD  Major depression disorder

MPH  Methylphenidate

MTA  Multimodal treatment study of children with ADHD

NICE  National Institute for Health and Care Excellence

NIH  National Institutes of Health

NIMH  National Institute of Mental Health

OCD  Obsessive-compulsive disorder

ODD  Oppositional defiant disorder

OFC  Orbitofrontal cortex

PBO  Placebo
PMT  Parent management training

PTSD  Posttraumatic stress disorder

RCT  Randomized controlled trial

R-MOAS  Retrospective-Modified Overt Aggression Scale

SAD  Social anxiety disorder

SAD  Separation anxiety disorder

SDQ  Strengths and Difficulties Questionnaire

SMD  Severe mood dysregulation

SP  Specific phobia

SRI  Serotonin reuptake inhibitor

SSR  Sympathetic skin response

TAU  Treatment as usual

TOSCA  Treatment of Severe Childhood Aggression study

UK  United Kingdom

US  United States

vmPFC  Ventromedial prefrontal cortex

VS  Ventral striatum
CHAPTER 1: Introduction

1.1 The importance of studying irritability in children and adolescents

Irritability is one of the most common reasons for referral to child and adolescent mental health services (Mikita & Stringaris, 2013; Peterson, Zhang, Santa Lucia, King, & Lewis, 1996; Stringaris, Vidal-Ribas, Brotman, & Leibenluft, 2018) and is a strong predictor of current and future functional impairment in young people (Althoff, Verhulst, Rettew, Hudziak, & van der Ende, 2010; Stringaris, Cohen, Pine, & Leibenluft, 2009; Vidal-Ribas, Brotman, Valdivieso, Leibenluft, & Stringaris, 2016), including suicide (Orri, Perret, Turecki, & Geoffroy, 2018b; Pickles et al., 2010). Furthermore, it is one of the few psychiatric symptoms to cut across internalising and externalising disorders (APA, 2013).

For instance, in children, irritability is part of the diagnostic criteria of every mood disorder described in the DSM-5, including all types of bipolar disorder (BD) and major depressive disorder (MDD). It is included in the diagnostic criteria for generalised anxiety disorder (GAD) and posttraumatic stress disorder (PTSD) and is also an associated feature of social anxiety disorder (SAD), separation anxiety disorder (SAD), specific phobia (SP), and selective mutism. In addition, irritability is a diagnostic criterion for oppositional defiant disorder (ODD) and is listed as an associated symptom for conduct disorder (CD), attention-deficit/hyperactivity disorder (ADHD), and autism spectrum disorders (ASD).

Recently, irritability as a disorder itself has been introduced in the DSM under the term disruptive mood dysregulation disorder (DMDD), which I will introduce in section 1.2.1.1, and describe in detail in section 1.3.1.
Surprisingly, despite the omnipresence of irritability across distinct psychiatric disorders, we still know little about its predictors, longitudinal outcomes, mechanisms, and, most importantly, effective treatment approaches. In this thesis, I will provide some answers to these questions with a series of quantitative analyses and empirical studies.

Before moving towards filling the gaps in the literature, I will describe what we know so far about irritability in children and adolescents. First, I will provide an initial working definition of irritability for the current thesis, along with a description of distinct irritability features, and other irritability-related constructs. Second, I will describe in detail the two conceptualisations of irritability that will be used in the empirical studies of this thesis; these are 1) irritability as a categorical diagnosis, namely DMDD, and 2) irritability as a dimension/subtype of ODD. This will be followed by a summary of the phenotypic presentation and correlates of irritability in other common psychiatric disorders in children and adolescents.

The last section of this chapter provides information on the assessment of irritability, its pathophysiological mechanisms, and the current knowledge of treatment approaches to irritability in children and adolescents.

1.2 A conceptual background of irritability

1.2.1 Working definitions of irritability

Irritability can be defined as an increased proneness to anger relative to peers at the same developmental level (Brotman, Kircanski, Stringaris, Pine, & Leibenluft, 2017). Proneness to anger is a dimensional trait that can be found in the population (Copeland, Brotman, & Costello, 2015) and shows to be stable over time (Caprara, Paciello, Gerbino, & Cugini,
2007). However, as I will discuss in section 1.2.2.3, the dimension of proneness to anger ranges from normality to pathology. This is in line with the increased recognition that common psychiatric symptoms are on a continuum (Plomin, Haworth, & Davis, 2009). The studies that form this thesis focus on the extreme end of this continuum; that is, on severe manifestations of irritability in youth. Specifically, Chapters 2, 4 and 5 consider severe irritability as the new DSM-5 diagnosis of DMDD, or its precursor, severe mood dysregulation (SMD), and Chapters 2 and 3 consider severe irritability as an ODD dimension/subtype. Although these are described in detail in sections 1.3.1 and 1.3.2, respectively, I am providing a definition of each of these constructs now to better understand the following sections.

1.2.1.1 An introduction to DMDD

Recently, the American Psychiatric Association introduced DMDD as a new diagnostic category in the DSM-5 under the section of Depressive Disorders (APA, 2013). Chronic, severe irritability is the hallmark of DMDD, which manifests through persistent irritable mood, and frequent temper outbursts. As I describe in section 1.3.1.1.1, the introduction of DMDD by the APA came as a response to the increased diagnostic rates of BD in children, with the consequent increase in prescriptions of antipsychotic medications. This increase in rates of BD diagnoses in children was the result of considering chronic, non-episodic irritability – in contrast to the classical episodic form of mania in adults - as a cardinal symptom of paediatric BD (Wozniak et al., 1995). This controversy motivated the definition of the severe mood dysregulation (SMD) criteria by Leibenluft et al. (2003) with the aim of facilitating empirical comparisons of pathophysiological mechanisms between youths with chronic irritability and those with classic episodic BD. As in DMDD, SMD
criteria also include persistent irritable mood and frequent temper outbursts; however, the criteria also include hyperarousal symptoms like those found in children with ADHD. Both DMDD and its precursor SMD, are characterised by chronic and severe irritability that has a tonic component (i.e., persistent irritable mood) and a phasic component (i.e., temper outbursts). Distinctions between chronic and episodic irritability, as well as between tonic and phasic irritability are described in sections 1.2.2.1 and 1.2.2.2 respectively.

1.2.1.2 An introduction to the irritability ODD dimension/subtype

The second conceptualisation of severe irritability in this thesis is irritability as an ODD dimension/subtype. According to the DSM-5, ODD is characterised by persistent irritable mood, and argumentative and defiant behaviours against authority figures such as parents or teachers (APA, 2013). In previous versions of the DSM, if ODD was accompanied by serious antisocial behaviours and violence, then the diagnosis of CD superseded that of ODD. However, research has shown that ODD is not only a precursor of CD (Burke, Loeber, Lahey, & Rathouz, 2005; Costello, Mustillo, Erkanli, Keeler, & Angold, 2003), and other disruptive behaviour disorders such as ADHD (Angold, Costello, & Erkanli, 1999), but also of emotional disorders such as depression and anxiety (Boylan, Vaillancourt, Boyle, & Szatmari, 2007; Burke et al., 2005; Maughan, Rowe, Messer, Goodman, & Meltzer, 2004). This wide range of associations led some researchers to consider ODD as a multidimensional disorder. In fact, nearly 10 years ago, Stringaris and Goodman (2009c) showed that irritability in ODD – defined as touchiness, easy annoyance and anger – was shown to have differential predictions compared to other symptoms such as defiance or vindictiveness. Since then, several studies have shown the independence of an irritability dimension in ODD through exploratory factor analyses (Stringaris, Zavos,
Leibenluft, Maughan, & Eley, 2012b), confirmatory factor analyses (Aebi, Plattner, Metzke, Bessler, & Steinhausen, 2013; Burke et al., 2014; Herzhoff & Tackett; Krieger et al., 2013; Stringaris et al., 2012b), and latent class analyses (Althoff, Kuny-Slock, Verhulst, Hudziak, & van der Ende, 2014; Burke, 2012; Herzhoff & Tackett; Kuny et al., 2013). In fact, this distinction amongst ODD has indeed been taken up by the DSM-5, which groups ODD symptoms into three categories, named Angry/Irritable Mood, Argumentative/Defiant Behaviour, and Vindictiveness. As in DMDD and SMD, irritability in ODD is considered chronic and severe.

1.2.2 Features of irritability
In this section, I briefly define features that are related to the course of irritability over time, and also discuss whether irritability might be considered normative or pathological.

1.2.2.1 Episodic vs chronic irritability

Based on its time course over weeks/months, current psychiatric nosology makes a distinction between chronic irritability and episodic irritability (Leibenluft, Cohen, Gorrindo, Brook, & Pine, 2006).

Chronic irritability is severe and persistent over time and it does not constitute a change from the child’s baseline mood. Chronic irritability is what characterises DMDD, and its precursor SMD, in which clinically significant irritability must have been present for at least 12 months. Chronic irritability is also seen in ODD, in which symptoms must have been present for at least 6 months (APA, 2013).

In contrast, episodic irritability, whilst still severe compared to others, constitutes a change from a child’s baseline mood. This is the case of depression or BD in children, in which
Irritability takes an episodic form, and is commonly accompanied by depressed and elated mood, respectively, with very few cases presenting only irritable mood (Hunt et al., 2009; Stringaris, Maughan, Copeland, Costello, & Angold, 2013).

**Figure 1.1** shows the pattern of chronic (blue) and episodic (red) irritability over time. Whereas both might be similar when measured at a single time point, the time course differs. And, as we will see in section 1.3.1.3, both chronic and episodic irritability also differ in their longitudinal correlates, family aggregation and pathophysiological mechanisms.

---

**Figure 1.1.** Distinct time courses of irritability. Episodic irritability (red) involves a change from baseline mood, whereas chronic irritability (blue) does not. Also, whereas tonic and phasic irritability have been discussed mostly in the context of chronic irritability, these can also take place during a mood episode with irritability.
1.2.2.2 Phasic vs tonic irritability

Irritability is conceptualised as having two components: *tonic* irritability and *phasic* irritability (Avenevoli, Blader, & Leibenluft, 2015). Phasic irritability refers to behavioural outbursts of intense anger, whereas tonic irritability refers to persistently angry, grumpy, or grouchy mood seen between these temper outbursts. Note, these two components are the core symptoms of DMDD: severe recurrent temper outbursts (i.e., phasic irritability) and irritable or angry mood between outbursts (i.e., tonic irritability). However, they can also be seen during epidosic irritability (see Figure 1.1).

Unlike what we know about the distinction between chronic and episodic irritability (see section 1.3.1.3), it is unclear whether tonic and phasic irritability are distinct constructs in terms of their longitudinal correlates, treatment response, pathophysiology, or family aggregation. In a recent study, Copeland et al. (2015) examined the course of tonic and phasic irritability using data from the prospective, community-based, Great Smoky Mountains Study (GSMS). The authors found that tonic and phasic irritability overlapped substantially in the community, and that tonic irritability (i.e., persistent irritable mood) in the absence of phasic irritability (i.e., temper outbursts) was rare (Copeland et al., 2015). They concluded that the distinction between tonic and phasic irritability might not be useful in community samples and needs to be tested in clinical data using instruments specifically designed to capture each component.

1.2.2.3 Normative vs pathological irritability

Being irritable or angry can be a normal and adaptive reaction (Carver & Harmon-Jones, 2009; Stringaris & Taylor, 2015). However, whether irritability is normative or pathological will be determined by the intensity of angry responses, their frequency, the
duration of irritable mood, and the context and its consequences, including resulting impairment.

Expressions of anger are more common early in development during the pre-school period and then decline with a slight increase in adolescence (Leibenluft & Stoddard, 2013). Thus, deciding where to set the threshold above which irritability would be considered pathological can be difficult, especially because such threshold can change dramatically during development. One study on a large sample (N=1,490) of preschool children found that nearly 84% of them had temper outbursts sometimes over the preceding month (Wakschlag et al., 2012) and close to 9% had daily tantrums. Although normative irritability peaks during the preschool years, a retrospective analysis of data from the GSMS found high rates of normative temper outburst (51%) and persistent irritable mood (28%) at any point during childhood and adolescence (Copeland et al., 2015). The same study found that normative irritability decreased with age regardless of sex. However, an increased number of tantrums was associated with an increased risk for future psychosocial impairment, with those scoring in the 90th percentile being at the highest risk for future impairment (Copeland et al., 2015).

Difficulties in setting the threshold for normative/pathological irritability is what led to APA to specify that, for DMDD criteria to be met, temper outbursts should be inconsistent with the developmental level of the child, and the diagnosis should not be assigned before the age of 6. In any case, clinical thresholds are often arbitrary and may impact upon the reliability (see section 1.5.1) and prevalence of severe irritability in the population.
For example, the lifetime prevalence of SMD is about 3.3% in children aged 9-19 (Brotman et al., 2006). The rates of DMDD were similar (3.3%) in a sample of pre-school children, and decreased to 1% in two samples of older youth (Copeland, Angold, Costello, & Egger, 2013). However, studies examining broader constructs such as mood dysregulation and mood lability, have reported higher prevalence estimates. In community studies, 5-6% of youth aged 8-19 had a lot of mood lability (Stringaris & Goodman, 2009b), and 3.8% of youth aged 4-16 years presented the CBCL Dysregulation phenotype (Althoff et al., 2010), which is linked to irritability. Finally, about 20% of adolescents participating in the Isle of Wight study (Pickles et al., 2010) were rated as displaying significant irritability in terms of frequency, severity and duration.

1.2.3 Irritability and related constructs

Irritability is referred by several names, not only by the lay public but also in the scientific literature and even in the DSM-5 itself. It is not the aim of this section to provide an exhaustive list of these names - for an in-depth review, I refer the reader to Toohey and DiGiuseppe (2017) – but I will discuss the most common of these nomenclatures. Indeed, some of these interrelated constructs are included in distinct irritability definitions. The most common are anger, frustration, and aggression (Buss & Durkee, 1957; Caprara et al., 1985; Dollard, Miller, Doob, Mowrer, & Sears, 1939; Leibenluft & Stoddard, 2013). Another common term is mood dysregulation, which captures irritability, but also other forms of mood liability.

Anger: Anger is the emotion that characterises irritability and is mentioned in our working definition as “increased proneness to anger”. As emotion, anger does not need to enter conscious awareness and people might be angry without being aware of it. Emotions are
described as action tendencies, and the effects of ‘unaware’ anger as emotion have been demonstrated through the subliminal presentation of stimuli and assessment of the participants’ response biases (Aarts et al., 2010). However, anger can also be a feeling, in that individuals are consciously aware of a set of thoughts and bodily sensations which they describe as anger. Similarly, an individual may sense a proneness to anger as feeling “on edge” or feeling “touchy”, which is another description for feeling irritable.

**Frustration:** Irritability and anger are often elicited by the frustration that arises when a goal is blocked (Berkowitz, 1989). Indeed, irritability is included in the National Institute of Mental Health (NIMH) Research Domain Criteria (RDoC) framework (Insel et al., 2010) as a negatively valenced construct named *frustrative non-reward*, which is defined as the reaction to blocked goal attainment. This construct will be discussed in detail in section 1.6.2.1 under pathophysiological mechanisms. However, it is important to understand that conceptualising irritability as a response to blocked goal attainment allows for research across species. Also, this involves that anger serves an adaptive function because is associated with an increased effort toward goal attainment (Carver & Harmon-Jones, 2009; Leibenluft & Stoddard, 2013).

Irritability and anger are negatively valenced and unpleasant, as are fear or sadness. However, while people typically avoid fear-inducing stimuli, people will approach anger-inducing stimuli in experimental paradigms (Aarts et al., 2010). This unique position of irritability among emotional valence and approach/avoidance tendencies, as described by Stringaris and Taylor (2015) is depicted in Figure 1.2. The behavioural manifestation of what experimental psychologists describe as an approach tendency may be the propensity to fight in order to obtain a blocked goal, which links irritability to aggression.
Figure 1.2. The position of irritability within commonly used terminology. Irritability is characterised by an underlying negative valence, like anxiety and depression, but elicits approach behaviour and is therefore linked to elation in mania. Adapted from Stringaris and Taylor (2015).

**Aggression:** The most dramatic consequence of irritability is aggression. Verbal or physical aggression are often the triggers for referring irritable youth to mental health services (Connor, 2002). However, aggression is one of the many consequences of irritability and by no means, it is the same thing. Data from the 2004 British Child and Adolescent Mental Health Survey (BCAMHS) suggest that aggression occurs in a minority of those with irritability; only 12% of general population subjects who report temper outbursts also report aggressive behaviour (Stringaris & Goodman, 2009b). Shouting and fighting, which are different types of aggression, can arise out of increased anger; however, irritable children may simply be grumpy, huffing and puffing, rather than being aggressive. Therefore,
separating irritability from aggression is important because, first, recognizing irritability might prevent the manifestation of aggression, and second, because irritability may require clinical attention independently of the presence of aggression.

**Mood dysregulation:** Mood dysregulation is a supraordinate term that can include irritability but also many other dysregulated emotions (Figure 1.2), such as dysregulated happiness in mania, dysregulated sadness in depression, or dysregulated fear in anxiety disorders. In research, the presence of mood dysregulation has typically been operationalised as having scores two or more standard deviations above the mean on the Child Behavior Checklist (CBCL) subscales of anxiety/depression, aggression, and attention problems (i.e., CBCL Dysregulated Profile or CBCL-DP) (Althoff et al., 2010; Deutz et al., 2018b). The CBCL-DP is associated with a wide range of internalising and externalising problems in follow up studies (Althoff et al., 2010) including personality pathology and suicidality (Deutz, Geeraerts, van Baar, Dekovic, & Prinzie, 2016; Deutz et al., 2018b). Another way of operationalising mood dysregulation has been by using the scores of the emotional, hyperactivity and conduct problems subscales on the Strengths and Difficulties Questionnaire (SDQ) (Goodman, 1997). The SDQ-DP has been associated with future antisocial behaviour and disciplinary measures in youth (Deutz et al., 2018a).

1.3 Irritability constructs in the current thesis

In section 1.2.1 I have introduced the two constructs of severe irritability that will be used in the empirical chapters of this thesis. Specifically, **Chapters 2, 4 and 5** consider severe irritability as the new DSM-5 diagnosis of DMDD, or its precursor, severe mood dysregulation (SMD); and **Chapters 2 and 3** consider severe irritability as an ODD dimension/subtype. In this section, I will describe in detail both conceptualisations of
irritability, including their phenotypic presentations, epidemiology, and clinical correlates. For DMDD, I will also provide the reasons and historical context that led to its introduction in the DSM, as well as the results from research focused on differentiating chronic (i.e., SMD/DMDD) from episodic (i.e., BD) irritability. In addition, since DMDD is the clinical category for irritability itself, I also provide information on its differential diagnosis. Finally, I will further summarise the different factorial models for ODD that have been suggested in the literature.

1.3.1 Irritability as DMDD

1.3.1.1 Phenotypic presentation

DMDD is a new category in the DSM-5 classified under the section of Depressive Disorders. DMDD is characterised by persistently irritable mood, and severe (i.e. out of proportion in intensity or duration) and frequent (i.e., three or more times per week) temper outbursts. These features should have been present for at least one year and begun before age 10, although the diagnosis should not be made before age 6 or after age 18. The irritability must be present in at least two settings, e.g., home, school, peers, and severely impairing in at least one of them. Most of our knowledge about severe chronic irritability as a category comes from research on the precursor to DMDD i.e., SMD (Leibenluft, 2011; Leibenluft et al., 2003). SMD and DMDD overlap substantially with two main differences. First, SMD criteria require persistent negative mood, which may be either irritability or sadness, whereas DMDD criteria only require irritability. Second, the definition of SMD includes a hyperarousal criterion which is omitted in the DMDD criteria. Table 1.1 highlights the differences between the diagnostic criteria of DMDD and SMD (Leibenluft et al., 2003).
Table 1.1. Comparison of diagnostic criteria between DMDD and SMD

<table>
<thead>
<tr>
<th>Criteria</th>
<th>DMDD</th>
<th>SMD</th>
</tr>
</thead>
<tbody>
<tr>
<td>**Age at diagnosis *</td>
<td>6-17 years</td>
<td>7-17 years</td>
</tr>
<tr>
<td>**Age of onset *</td>
<td>Before 10 years</td>
<td>Before 12 years</td>
</tr>
<tr>
<td><strong>Temper outbursts</strong></td>
<td>Severe and recurrent temper outbursts, inappropriate for age and/or precipitating event, manifested verbally and/or behaviourally (e.g. verbal rages, and/or aggression toward people or property.)</td>
<td>Severe and recurrent temper outbursts, inappropriate for age and/or precipitating event, manifested verbally and/or behaviourally (e.g. verbal rages, and/or aggression toward people or property.)</td>
</tr>
<tr>
<td><strong>Frequency of temper outbursts</strong></td>
<td>On average, at least three times a week</td>
<td>On average, at least three times a week</td>
</tr>
<tr>
<td>**Mood between outburst *</td>
<td>Irritable or angry mood present at least half of the day most days, and is noticeable by others (e.g., parents, teachers, peers).</td>
<td>Abnormal mood (specifically anger or sadness), present at least half of the day most days, and is noticeable by others (e.g., parents, teachers, peers).</td>
</tr>
<tr>
<td><strong>None</strong></td>
<td></td>
<td>Presence of at least three of the following symptoms: insomnia, agitation, distractibility, racing thoughts or flight of ideas, pressured speech, intrusiveness</td>
</tr>
<tr>
<td>**Hyperarousal symptoms *</td>
<td>Symptoms should have been present for at least 12 months without any symptom-free periods exceeding 3 months.</td>
<td>Symptoms should have been present for at least 12 months without any symptom-free periods exceeding 2 months.</td>
</tr>
<tr>
<td>**Duration *</td>
<td>The symptoms are present in at least two setting (i.e., at home, school, or with peers) and are severe in at least one of these.</td>
<td>The symptoms are present in at least two setting (i.e., at home, school, or with peers) and are severe in at least one of these.</td>
</tr>
</tbody>
</table>

* Criterion differs between DMDD and SMD; differences are highlighted in bold.
1.3.1.2  The need for a psychiatric diagnosis with irritability as a primary symptom

There were two main reasons for the American Psychiatric Association to introduce DMDD in the DSM. The first reason was the increasing rates of BD diagnoses in children in the United States in the last decades, and the second was the need to diagnostically accommodate youth who have severe irritability as a primary problem. Both reasons are described below.

1.3.1.2.1  The controversy regarding paediatric BD in youth

In the past decades, the rates of BD diagnoses in youth have increased dramatically in the US. The rates increased more than 400% in inpatient units (Blader & Carlson, 2007) between 1996 and 2004 (Figure 1.3), and 40-fold in outpatient services (Moreno et al., 2007) between 1994 and 2003 (Figure 1.4). The most plausible explanation for this increase seems to be the changes in how BD was diagnosed in children (Leibenluft, 2011; Mikita & Stringaris, 2013), for which irritability, along with hyperarousal symptoms, played a central role. This was because some researchers thought that mania may present differently in children, prompting them to suggest that paediatric BD was characterised by chronic, non-episodic irritability, as opposed to the classical episodic form of mania in adults (Wozniak et al., 1995). Of note, the DSM criteria for mania require a “distinct period of abnormally and persistently elevated, expansive or irritable mood” which I discuss in section 1.4.3. It seems likely that ignoring the requirement of “distinct period” led to many children with chronic irritability being misdiagnosed with BD. The increase in rates of BD diagnoses also coincided with a rise in prescription rates of antipsychotic drugs (Olfson, Blanco, Liu, Moreno, & Laje, 2006). For example, a recent survey from prescriptions in the US state of Kentucky showed that over 10,000 children of six years of age or less (over 4%
of all children of that age enrolled in Medicaid) received a prescription for antipsychotic medication, and that 75% of those prescriptions were for a diagnosis of BD (Lohr, Chowning, Stevenson, & Williams, 2015).

The controversy about whether chronic irritability was a hallmark of paediatric bipolar, motivated the definition of the SMD criteria by Leibenluft et al. (2003), which is defined by persistent negative mood and hyperarousal symptoms. The purpose of defining SMD was to facilitate the empirical comparison of youths with chronic irritability and those with classic episodic BD by using longitudinal designs, looking at family aggregation, and examining pathophysiological mechanisms. The studies that followed provided evidence against the notion that severe chronic irritability is an early-life form of BD (Leibenluft, 2011) (for a summary of these studies see section 1.3.1.3). Thus, to tackle the dramatic increase of bipolar diagnoses and the rise of antipsychotic prescriptions in youth, the APA introduced DMDD in the DSM-5 (APA, 2013).
Figure 1.3. Changes in US rates for groups of psychiatric disorders coded as primary diagnoses in acute US inpatient units from 1996-2004 (Data from Blader & Carlson, 2007)

Figure 1.4. US national trends in bipolar disorder visit rates in outpatient services from 1994-2003 (Data from Moreno et al. 2007). Whereas BD visit rates increased slightly in adults, these raised dramatically in youth within a 10-year period.
1.3.1.2.2 Why ODD was not enough

The other reason for the APA to create DMDD was to diagnostically accommodate youth presenting with severe irritability as a primary symptom, which was not codable under the DSM-IV unless if subsumed under ODD.

Several researchers and clinicians have argued that it is hard to differentiate DMDD from ODD and that adding an irritability specifier for ODD in the new DSM-5 would have been preferable to the new diagnosis of DMDD (Runions et al., 2016). These views were supported by the high overlap seen between rates of DMDD and ODD (Axelson et al., 2012; Freeman, Youngstrom, Youngstrom, & Findling, 2016; Mayes, Waxmonskey, Calhoun, & Bixler, 2016) and by the ample empirical evidence demonstrating that symptoms of ODD can be broken down into three dimensions, i.e., irritable, headstrong and hurtful (Burke et al., 2014; Krieger et al., 2013; Stringaris & Goodman, 2009a, 2009c; Vidal-Ribas et al., 2016). This distinction is reflected in the DSM-5, which groups ODD symptoms in three categories, namely Angry/Irritable mood, Argumentative/Defiant Behaviour, and Vindictiveness.

Nevertheless, evidence shows that youth with DMDD are extreme cases of the most irritable ODD youth (APA, 2013), who have more functional impairment and are in need for substantially more services (Conner, Meldrum, Wieczorek, Duberstein, & Welte, 2004; Nock, Kazdin, Hiripi, & Kessler, 2007; Peterson et al., 1996; Pickles et al., 2010; Stringaris et al., 2009). Moreover, diagnostic specifiers are often not used by clinicians. Thus, simply adding a specifier may not adequately address the pressing public health need. Furthermore, whereas most youth with DMDD will meet criteria for ODD, the opposite is not the case, since DMDD is much rarer. Importantly, the multidimensionality of ODD suggests it is a
heterogeneous construct that encompasses behaviours ranging from temper outbursts to vindictiveness. While it could be advantageous to have a broad-sweep diagnostic category clinically, the approach might undesirably group together symptoms with distinct aetiologies. Indeed, evidence suggests that oppositionality and irritability, which are both components of ODD, have differential longitudinal and genetic associations (Stringaris et al., 2012b). In this context, having a specific irritability category may be advantageous in terms of external validity, and possibly, treatment specificity. Finally, given the significant mood component of DMDD and the longitudinal, genetic, and cross-sectional associations among irritability, anxiety, and depression (Althoff et al., 2014; Copeland, Shanahan, Egger, Angold, & Costello, 2014; Stringaris et al., 2012b), DMDD is appropriately placed in the mood disorders section, whereas ODD is in the Disruptive Behaviour Disorders section of the DSM-5 (APA, 2013).

1.3.1.3 Distinguishing chronic vs episodic irritability

Although there are some studies on DMDD, most of the research which were aimed at differentiating chronic irritability from paediatric BD have been done using the SMD criteria. The outcomes of such research provide compelling evidence against the notion that severe chronic irritability is an early-life form of BD. I summarise some of these findings below.

First, longitudinal studies in community samples show no associations between dimensional measures of irritability and later BD (Brotman et al., 2006; Leibenluft et al., 2006; Stringaris et al., 2009). The results are similar in referred samples. A longitudinal clinical study (median follow-up time 28.7 months) demonstrated a clear difference in the rate of manic symptoms between youth with SMD and those with classic episodic BD. As
shown in Figure 1.5, only one of 84 SMD subjects (1.2%) experienced a (hypo-) manic episode during the study, whereas the frequency of such episodes was more than 50 times higher in those with narrowly defined BD (58/93, 62.4%) (Stringaris et al., 2010a). Findings were similar in other studies (Axelson et al., 2012; Deveney et al., 2015).

Figure 1.5. Bars with standard errors show the percentage of patients with either severe mood dysregulation (SMD) or bipolar disorder (BD) who developed a (hypo-) manic or mixed episode during the follow-up period. Adapted from Stringaris et al. (2010a).

Second, individuals with SMD and BD also differ in family history of psychiatric disorders (Axelson et al., 2012; Brotman et al., 2007). Parents of youth with narrowly-defined BD were significantly more likely to have BD (14/42, 33.3%) than parents of youth with SMD (1/37, 2.7%) (Brotman et al., 2006). Indeed, a recent study found higher rates of DMDD
symptoms in offspring of parents with depression than in offspring of parents with BD, and DMDD diagnoses were only present in the former (Propper et al., 2017).

Third, behavioural and functional MRI studies have found that while both youths with BD and those with SMD have impairments in labelling facial emotions (Guyer et al., 2007; Rich et al., 2008), the neural correlates of this deficit differ between the two groups (Brotman et al., 2010; Thomas et al., 2012; Wiggins et al., 2016). A similar pattern was found in their response to frustration; that is, although both SMD and BD youth displayed significantly more negative affect than healthy controls in response to negative feedback, patients with SMD differed from those with BD in their event-related potentials (Rich et al., 2007) and brain activation patterns (Rich et al., 2011). In addition, recent evidence suggests that the pathophysiological correlates of trait irritability itself differ between BD and DMDD (Wiggins et al., 2016). Section 1.6 provides more details on genetics, family studies, brain imaging and pathophysiology of irritability.

### 1.3.1.4 Epidemiology

One of the main concerns about the introduction of DMDD has been the pathologisation of normal childhood behaviours, such as temper outbursts (Althoff et al., 2016; Axelson et al., 2012; Freeman et al., 2016; Mayes et al., 2016). However, the low prevalence of DMDD in epidemiological studies suggests this was an exaggerated concern. As mentioned when I discussed the boundaries between normative and pathological irritability in section 1.2.2.3, SMD, which is closely related to DMDD, has been estimated to occur in about 3% of the population of 9- to 16-year-olds (Brotman et al., 2006). More recently, a study by Copeland et al (2013) using data from two different datasets showed that prevalence estimates of DMDD varied between 0.8% and 1.1 % for children between 9 and 17 years of age.
Moreover, studies using the most stringent criteria in adolescents have found rates as low as 0.12% (Althoff et al., 2016). These rates are relatively low compared to other conditions such as ADHD or ODD. In fact, the prevalence of DMDD is low (3.3%) even in youth at risk for mood disorders (Propper et al., 2017).

The temporal stability of the diagnosis and main symptoms of DMDD is low (Mayes et al., 2015a), at least in community samples. For example, only 19% of children initially diagnosed with DMDD in the Longitudinal Assessment of Manic Symptoms study maintained the disorder across 12- and 24-month follow-ups (Axelson et al., 2012). This is in contrast to the moderate stability of the dimensional measure of irritability, with longitudinal correlations ranging between 0.29 and 0.88 (Leadbeater & Homel, 2015; Roberson-Nay et al., 2015; Stringaris et al., 2012a; Whelan, Stringaris, Maughan, & Barker, 2013). The reasons for the low continuity are unclear but likely to be related to the arbitrarily chosen thresholds for the duration of symptoms and maximum period allowed without symptoms. Indeed, most of the cases with DMDD who do not meet criteria at follow up have significant ongoing subthreshold irritability and severe impairment (Deveney et al., 2015).

It is important to note that all the epidemiological data on DMDD have been collected with instruments that were not specifically designed to capture the disorder (since the condition had not been defined yet during the design and data collection of these studies). In addition, the diagnoses for DMDD in these studies have been assigned retrospectively, for most of the cases using items from ODD and depression interview sections. Therefore, it is still unclear whether the prevalence rates and stability of the DMDD diagnosis would differ in
community-based samples if the information was collected prospectively and with specifically designed instruments.

1.3.1.5 Psychopathological correlates

DMDD co-occurs with other emotional or behavioural disorder in 65% to 90% of cases; mainly with depression (~27%), anxiety (~26%), ADHD (~28%), CD (~26%) and ODD (~60%) (Althoff et al., 2016; Axelson et al., 2012; Brotman et al., 2006; Copeland et al., 2013; Dougherty et al., 2014). However, the rates of comorbidity differ across studies, possibly due to the type of sample examined and the age of the participants. For example, the rates of comorbidity are higher in clinically-referred samples and in older participants, where mood and behavioural disorders might co-occur with DMDD in ~40% and ~70% of cases, respectively (Althoff et al., 2016; Axelson et al., 2012). These numbers decrease to under ~20% and ~50%, respectively, in community or younger samples (Copeland et al., 2013; Dougherty et al., 2014). Of note, some of the overlap with behavioural disorders is artificial due to item overlap (e.g., ODD) in post-hoc analyses, as mentioned in the previous section. For this reason, ODD should not be coded if DMDD criteria are met (APA, 2013). In any case, it will be important to establish whether DMDD shows more comorbidity that is characteristic of other established psychiatric disorders, for example ADHD, after excluding the artificial overlap due to employing the same items. In addition to higher rates of comorbidities, the diagnosis of DMDD is associated with high levels of social impairment, service use, school suspensions, and family poverty (Axelson et al., 2012; Copeland et al., 2013).

Regardless of the several associations between irritability and other disorders in cross-sectional analyses, longitudinal studies have demonstrated that irritability is a specific
predictor of future depression and anxiety (Brotman et al., 2006; Copeland et al., 2014; Leibenluft et al., 2006; Stringaris et al., 2009). The results of these studies are systematically reviewed and summarised in a meta-analysis in **Chapter 2** of this thesis. Evidence suggests that this longitudinal association between irritability and depression/anxiety is explained by shared genetic variance rather than shared environmental risks (Savage et al., 2015; Stringaris et al., 2012b). The relation between irritability and depression is also evident in family studies; for example, maternal depression predicts irritability in young children (Wiggins, Mitchell, Stringaris, & Leibenluft, 2014); and irritability, alongside anxiety, is a major pathway to depression for children and adolescents at high risk of depression (Rice et al., 2017; Whelan, Leibenluft, Stringaris, & Barker, 2015). Finally, DMDD and severe chronic irritability have also been associated with future functional impairment in longitudinal studies (Copeland et al., 2014; Dougherty et al., 2016), including suicidality independently of other psychopathology (Pickles et al., 2010).

### 1.3.1.6 Differential diagnosis

Before diagnosing DMDD, one must exclude other conditions that may lead to tantrums and grumpiness, including medical conditions (particularly in hospital or other medical settings), drug-induced conditions, and other psychiatric problems that may be treatable in their own right.

**Bipolar disorders (BD):** Episodes are the single most reliable way to differentiate DMDD and BD-I or BD-II. Patients with DMDD do not have a course of illness with episodes that last for several days or weeks; instead, their symptoms are chronic, lasting for at least a year. In contrast, patients with BD and their parents report more-or-less clearly defined
periods of either mania or depression. As discussed later in section 1.4.1, mania is commonly characterised by elation, although irritability alone can be the predominant mood in few cases. Such irritability should be a noticeable change from the patients’ baseline mood and should be accompanied by the onset of other manic symptoms. The diagnosis of DMDD should not be assigned if the patient has ever had a manic or hypomanic episode, or if elevated or expansive mood is present.

**Oppositional defiant disorder (ODD):** DSM-5 criteria for DMDD specify that ODD may not be diagnosed when both criteria are met. The intensity and frequency of temper outbursts, and the chronicity of disrupted mood between outbursts, is more severe in youth with DMDD than in those with ODD. For that reason, whereas most youth with DMDD will meet criteria for ODD (65%-70%) the opposite is not the case; for example, rates of DMDD in youth with ODD were approximately 30% in a large epidemiological study (Copeland et al., 2013).

**Attention Deficit Hyperactivity Disorder (ADHD):** A child meeting criteria for DMDD can also be assigned a diagnosis of ADHD if the criteria of the latter are met. Rates of comorbidity of ADHD in children with DMDD are not different from those of DMDD in children with ADHD in community-based samples, both are approximately 20% or lower (Copeland et al., 2013; Mulraney et al., 2016). However, rates of ADHD in youth with DMDD are close to 85% in clinically-referred samples (Deveney et al., 2015). One study in children with ADHD found that those who also met criteria for DMDD had higher rates of comorbidity, especially ODD and anxiety disorders, and poorer self-control and social functioning, compared to those without DMDD (Mulraney et al., 2016). For the differential diagnosis, a careful history of the timing of irritability is important. In youth with ADHD,
the severity of irritability commonly fluctuates as a consequence of external stimuli. Environmental events such as the beginning of a new school year or those that place greater demands on concentration and the need to sit still, often coincide with worsening of irritability in those with ADHD.

*Autism Spectrum Disorders (ASD):* The DSM-5 criteria for DMDD stipulate that if the patient’s presentation is better explained by ASD, the diagnosis of DMDD should not be assigned. Temper outburst in children with ASD are common and might occur as a response to change in routines or due to sensory sensitivity.

*Depressive and Anxiety Disorders:* Youth with depression or anxiety can be diagnosed with DMDD if the criteria for both disorders are met. However, children whose irritability is only present in the context of a depressive episode or an anxiety disorder, as discussed in detail in the sections 1.4.2 and 1.4.3, should not receive an additional diagnosis of DMDD. Of note, depression and anxiety in youth with DMDD is usually undiagnosed because disruptive symptoms over-shadow those of low mood or anxiety. Given the evidence presented above about the high likelihood for irritable youth to develop these disorders as they grow up, it is therefore important to assess depressive and anxiety symptoms in children with DMDD.

### 1.3.2 Irritability as ODD dimension/subtype

#### 1.3.2.1 Phenotypic presentation

DSM-5 criteria define ODD as a pattern of angry/irritable mood, argumentative/defiant behaviour, or vindictiveness lasting at least 6 months evidenced by at least 4 symptoms, from a list of 8, divided into the following categories:
Angry/Irritable Mood

- Often loses temper.
- Is often touchy or easily annoyed.
- Is often angry and resentful.

Argumentative/Defiant Behaviour

- Often argues with authority figures or, for children and adolescents, with adults.
- Often actively defies or refuses to comply with requests from authority figures or with rules.
- Often deliberately annoys others.
- Often blames others for his or her mistakes or misbehaviour.

Vindictiveness

- Has been spiteful or vindictive at least twice within the past 6 months.

This pattern of behaviour should be seen in the children during interactions with individuals other than siblings. DSM-5 also adds a note related to the boundaries between normative and pathological behaviour. Specifically, for children under 5 years, these behaviours should occur most of the days, whereas in children who are 5 years or older these behaviours should occur at least once a week. As in DMDD, symptomatic behaviours are linked to what is expected based on each individual’s developmental level in terms of persistence, frequency and intensity. In contrast with DMDD, in which impairment must be pervasive in at least two settings, ODD symptoms must be present in at least one setting for the diagnosis to be given. However, it is common for ODD to cause impairment in more
than one setting, and pervasiveness across settings is associated with greater functional impairment (Frick & Nigg, 2012).

ODD is therefore described as a combination of both behavioural problems and negative mood. Irritability is represented as having a low threshold for temper outburst, being frequently angry, and easily becoming annoyed. Thus, since diagnostic criteria are met if four out of the possible eight symptoms are present, most children with ODD will show irritability. Also, as mentioned earlier in section 1.2.1.2, irritability in ODD is considered chronic, as opposed to episodic.

1.3.2.2 Epidemiology

A systematic review found that the prevalence of ODD is around 3.3% across cultures (Canino, Polanczyk, Bauermeister, Rohde, & Frick, 2010). The prevalence of ODD increases in preschool children (9-12%) (Lahey et al., 2000; Lavigne et al., 2001) and decreases in school-aged children and adolescents (range 3-6%) (Boylan et al., 2007; Merikangas et al., 2010). However, reported prevalence rates vary substantially across studies (1-16%) (Canino et al., 2010; Loeber, Burke, Lahey, Winters, & Zera, 2000). The differences in prevalence estimates might be associated with where the threshold between normative and pathological behaviour is defined. Defiant, oppositional behaviour and irritability might be part of the normal development in early childhood. For that reason, DSM-5 now provides guidance on when these behaviours are a departure from normal development; it states that temper outbursts for preschool-aged children are common and specifies that different frequencies and intensities of symptoms should be taken into account before and after 5 years of age when assigning ODD diagnosis to a child (APA, 2013).
Findings from studies examining sex differences in rates of ODD are mixed, but overall ODD seems to be slightly more prevalent in boys than girls before adolescence (Loeber et al., 2000); however, these differences disappear when ODD behaviours are reported by parents (Lahey et al., 2000; Maughan et al., 2004).

1.3.2.3 The multidimensionality of ODD

Before examining the psychopathological correlates of ODD, it is important to consider the multidimensionality of this disorder. As I will describe in the next section, nearly 10 years ago, Stringaris and Goodman showed that irritability had differential predictions compared to other symptoms of ODD, such as defiance or vindictiveness (Stringaris & Goodman, 2009c). This finding prompted a series of exploratory and confirmatory factor analyses that demonstrated that irritability is a distinguishable dimension within ODD (Aebi et al., 2013; Burke et al., 2014; Herzhoff & Tackett; Krieger et al., 2013; Stringaris et al., 2012b).

Several models have been proposed (Burke, 2012; Burke, Hipwell, & Loeber, 2010; Rowe, Costello, Angold, Copeland, & Maughan, 2010; Scott & O'Connor, 2012; Stringaris & Goodman, 2009c), including 2-factor and 3-factor models with correlated factors, and also models that add a general ODD (bi)factor. All these models have in common that consider at least one irritability dimension and one oppositional behaviour dimension. Despite their small differences, these models have been compared across different samples (Aebi et al., 2013; Burke et al., 2014; Ezpeleta, Granero, de la Osa, Penelo, & Domenech, 2012; Herzhoff & Tackett, 2016; Lavigne, Bryant, Hopkins, & Gouze, 2015). Possibly the most comprehensive model comparison analysis to date was carried out by Burke et al. (2014) in five large community-based samples; they concluded that an irritable dimension defined as anger, temper outburst and easy annoyance fit the data best, along with a headstrong/hurtful
dimension as proposed by Stringaris and Goodman (2009c). In fact, this is the conceptualisation adopted by the DSM-5 (APA, 2013). **Figure 1.6** shows the most employed and/or compared models in the literature. Note that 2-factor and 3-factor models could also have a second-order ODD factor or be bifactor models, in which individual items load directly on both, ODD dimensions, and a general ODD factor. The latter is probably what best accounts for the relationship between ODD symptoms (Burke et al., 2014).

The independence of irritability as a construct has also been confirmed using latent class analysis (Althoff et al., 2014; Burke, 2012; Herzhoff & Tackett; Kuny et al., 2013). Unlike factor analysis, which examines dimensions within the data, this method aims to identify groups of people based on their response to a questionnaire.

### 1.3.2.4 Psychopathological correlates

The validity of irritability as a distinct ODD dimension is supported by its distinct associations with other psychiatric disorders. In a cross-sectional study using data from the BCAMHS, Stringaris and Goodman (2009c) found that the irritable dimension was associated with emotional disorders, including depression and GAD, whereas the headstrong dimension was associated with ADHD and non-aggressive CD, and the hurtful dimension was associated with aggressive conduct problems and callous/unemotional traits. The differential cross-sectional associations of the three dimensions of ODD not only have been replicated in other community-based samples, including preschool children (Ezpeleta et al., 2012) and in other countries (Aebi et al., 2013; Krieger et al., 2013) but also in clinical samples of children with ASD (Mandy, Roughan, & Skuse, 2014) or chronic tic disorder (Thériault et al., 2014). Moreover, when splitting groups of youth based in the
presence of irritability and defiant behaviours, those with increased levels of irritability are characterised by a higher rate of emotional problems (Drabick & Gadow, 2012; Wesselhoeft et al., 2018) and suicidality (Aebi et al., 2016).

Figure 1.6. Oppositional defiant disorder dimension models most frequently used and compared in the literature. Model 2 is from Burke (2012). Model 3 is from Rowe at al. (2010). Model 4 is from Stringaris and Goodman (2009c). Model 5 is from Burke et al. (2010). Model 6 is from Aebi et al. (2010). These same models have been also tested adding a general ODD factor either as a second-order level or in a bifactor model, in which individual items load directly on this general factor. ODD: Oppositional defiant disorder. Irritab.: Irritability. Neg. Affect: Negative affect. Opp. Behav.: Oppositional behaviour. Heads.: Headstrong. Antag. Behav.: Antagonistic behaviour.
As it happens with the chronic irritability seen in DMDD (see section 1.3.1.5), the irritable dimension of ODD is predictive of depression and anxiety in longitudinal studies (Stringaris & Goodman, 2009a), whereas the headstrong and hurtful dimensions are more predictive of ADHD and CD. These distinct longitudinal associations across ODD dimensions have also been replicated in other samples (Burke, 2012; Burke et al., 2010; Lavigne, Gouze, Bryant, & Hopkins, 2014; Rowe et al., 2010; Stringaris et al., 2012b; Whelan et al., 2013), including children with chronic tic disorder (Theriault et al., 2018). Furthermore, similar results have been found when comparing groups of children generated by latent class analysis (LCA); that is, those children characterised by high levels of irritability are more likely to present emotional problems in the future than those children with low levels of irritability (Althoff et al., 2014; Herzhoff & Tackett, 2016; Kuny et al., 2013).

1.4 Irritability in the context of other psychiatric disorders

As mentioned in section 1.1, irritability is present in many other psychiatric disorders in children and adolescents. However, the phenomenology of irritability differs across these disorders, either in its timing (e.g., episodic as opposed to chronic irritability) or the underlying mechanisms. In this section, I describe the phenomenology and correlates of irritability in the context of psychiatric disorders in which irritability, despite not being a defining symptom, is a common manifestation. I especially focus on BD, depression and anxiety, which are the most relevant for this thesis. However, I also provide some information on irritability in the context of ADHD and ASD.
1.4.1 Irritability in Bipolar disorders

1.4.1.1 Phenotypic presentation

BD is defined by the presence of at least one manic (BD-I) / hypomanic (BD-II) episode that may be preceded (must be in the case of hypomanic episodes) or followed by major depressive episodes. The DSM-5 Criterion A for manic/hypomanic episodes requires a “distinct period of abnormally and persistently elevated, expansive or irritable mood”. This period of abnormal mood - either elated, irritable or both - is associated with the onset or worsening of other symptoms (Criterion B) including inflated self-esteem or grandiosity, decreased need for sleep, increased talkativeness, pressured speech, flight of ideas, racing thoughts, distractibility, increased goal-directed activity, psychomotor agitation, and/or excessive involvement in activities that might have undesirable consequences (APA, 2013).

Two things should be noted from the definition set in Criterion A. First, the type of irritability seen in BD is episodic – that is, the irritability manifests as a change from the child’s baseline mood, and parents describe it as something that is out of character. Therefore, in children and adolescents showing severe irritability, a diagnosis of BD should only be applied to those who have shown a “distinct period” during which the irritability clearly differed from the child’s usual mood state and was accompanied by the onset of other manic symptoms listed in Criterion B.

Second, irritability in BD is considered a cardinal symptom alongside elated mood, in both adults and children. As described in section 1.4.2, this contrasts with the criteria for major depressive episodes, in which irritability is only considered a cardinal symptom in children and adolescents. To note, aspects of the DSM guidelines might differ from guidelines applied in other countries, such as the UK, where the National Institute for Health and Care
Excellence advises that irritability without elation or grandiosity should not allow a diagnosis of mania in children (NICE, 2014). This might partly explain the 72-fold difference in the discharge rates for paediatric BD between the US and UK, being higher in the US, from the year 2000 to 2010 (James et al., 2014). Nevertheless, the importance of elation over irritability in BD seems to be implicitly recognised by the DSM-5 through the requirement of four additional manic symptoms from Criterion B if the dominant mood is irritable, but only three symptoms if the dominant mood is elated or expansive.

Episodic irritability in BD is strongly influenced by the patient’s interactions with the environment. As symptoms of disinhibition become more severe and the patient faces increasing restrictions by their family or mental health professionals, increasingly more behavioural goals are blocked which will lead to frustration. Indeed, it is typical in such instances to observe extreme euphoria turns into very severe irritability, which increases the risk of aggression.

1.4.1.2 Epidemiology

The prevalence of paediatric BD is estimated to be 1.8% (Van Meter, Moreira, & Youngstrom, 2011). Yet, this rate might be inflated by the inclusion of broad definitions of BD in some studies; those employing narrow definitions have found rates as low as 0.1% (Costello et al., 1996; Stringaris, Santosh, Leibenluft, & Goodman, 2010c).

Epidemiological studies that include bipolar disorder-not otherwise specified (BD-NOS) report higher rates of BD (range 2.4-6.7%) (Van Meter et al., 2011). BD-NOS is a heterogeneous and broadly defined phenotype characterised by episodic bipolar symptoms that are too brief to meet the DSM-5 duration criteria. Many children with severe chronic irritability and short temper tantrums are erroneously diagnosed with BD-NOS even though
evidence shows differences with SMD in clinical presentation and longitudinal course (Towbin, Axelson, Leibenluft, & Birmaher, 2013). Moreover, recent data show that irritability as a symptom in BD-NOS is less common (46%) than in BD-I or BD-II (Van Meter, Burke, Kowatch, Findling, & Youngstrom, 2016), which further emphasises the distinction between BD-NOS and SMD/DMDD.

Episodic irritability is one of the most common symptoms among youth with BD, with a prevalence of 77%, along with increased energy (79%) and mood lability (76%) (Van Meter et al., 2016). However, irritability in the absence of elation is very rare; one study found that about 10% of youth with BD had only irritability as a cardinal symptom (Hunt et al., 2009); and most of them eventually also experienced elation in the course of four years (Hunt et al., 2013).

1.4.1.3 Psychopathological correlates

Few studies have examined the correlates of episodic irritability within the context of BD. Reports from the Course and Outcome of Bipolar Illness in Youth (COBY) study, a large prospective study examining youth with BD, showed that those who only displayed irritable mood as a cardinal symptom (10%) were younger than those with elation or those with both elation and irritability (Hunt et al., 2009). In addition, they had more second-degree relatives with depressive disorders and higher risk for depressive episodes (Hunt et al., 2013). However, they did not differ from elated BD youth in their subtypes, rates of psychiatric comorbidities, severity of illness, duration of illness, family history of mania, duration of mood episodes, risk for suicidal attempts or functional impairment. Data from the adult literature show that irritability in BD is associated with lifetime anxiety disorders.
as well as higher recurrence of, and slower recovery from depressive episodes (Yuen et al., 2016a; Yuen et al., 2016b).

1.4.2 Irritability in Depressive disorders

1.4.2.1 Phenotypic presentation

A major depressive episode is defined in the DSM-5 by a period of at least 2 weeks during which there is either depressed mood or loss of interest or pleasure in most activities (Criterion A). In addition, for children and adolescents, the DSM-5 further notes that the mood may be irritable rather than sad.

In contrast to the criteria for BD, irritability as a cardinal symptom in depression is only allowed in young people, but not in adults. This is the only difference between the diagnosis of depression in children and adults. However, similar to the irritability seen in paediatric mania, irritability in children with depression is also episodic; that is, it should be a change from the patient’s baseline mood noticeable by parents, teachers or friends.

1.4.2.2 Epidemiology

The lifetime prevalence estimate for depression in children and adolescent in the US is 10-15% (Avenevoli, Swendsen, He, Burstein, & Merikangas, 2015; Merikangas et al., 2010). Reports from the adult literature in large community-based samples suggest that episodic irritability is a common symptom in depression, with a prevalence around 30-55% (Fava et al., 2010; Judd, Schettler, Coryell, Akiskal, & Fiedorowicz, 2013). However, episodic irritability in the absence of depressed mood or anhedonia is extremely rare.

Evidence for irritability as a major criterion of depression in young people suggests the same profile. Using data from the GSMS, Stringaris et al. (2013) divided 9–16-year-olds
participants (n=1420) who met criteria for depression into three groups: those with depressed mood and no irritability, those with irritability and no depressed mood, and those with both depressed and irritable mood. In this study, the most common cardinal symptom in young people with depression was depressed mood (58.7%), followed by the co-occurrence of depressed and irritable mood (35.6%), while irritable mood alone was rare (5.7%). This suggests that specifying irritability as a cardinal symptom in its own right in paediatric depression does not have a significant impact in terms of identifying new cases. In this same investigation, Stringaris et al. (2013) found that boys were more likely to have depressed mood with irritability than girls, who presented more often depressed mood only. Interestingly, there were no differences in age and pubertal stage between those with depressed and irritable mood and those with only depressed mood, suggesting that early-onset depression is not associated with higher likelihood of presenting with irritability.

1.4.2.3 Psychopathological correlates

In terms of psychiatric correlates, Stringaris et al. (2013) found that nearly 50% of youth with depression and irritability had a comorbid ODD or CD, more than double the rates of these comorbidities in the depressed-only group. In adults, depression with irritability is also associated with higher rates of comorbidity as well as a more severe and longer course of illness and functional impairment (Judd et al., 2013). Since most cases of adult depression have an onset in adolescence (Kessler et al., 2005), it is essential to assess irritability symptoms in young people with depression, regardless of the predominant mood.
1.4.3 Irritability in Anxiety disorders

1.4.3.1 Phenotypic presentation

Irritable mood and temper outbursts appear in a host of anxiety disorders in children (Stoddard et al., 2014). For instance, irritability is included in the DSM-5 as an associated symptom of GAD for both children and adults. Yet, for children, it might be the only symptom needed alongside excessive worry to meet the criteria for diagnosis, whereas three additional symptoms are needed for adults. Temper outbursts are also listed in DSM-5 as associated features for selective mutism and separation anxiety. Furthermore, in children, the criteria for specific and social phobia emphasise that the excessive fear and anxiety might be expressed through tantrums (APA, 2013).

Irritability is also an associated symptom of posttraumatic stress disorder (PTSD) and acute stress disorder. Notably, this type of irritability that occurs after trauma is one of the most impairing symptoms of the so-called complex trauma, which is a condition characterised by mood dysregulation resulting from prolonged exposure to severe stressors in developmentally vulnerable times such as early childhood or adolescence. A recent systematic review and meta-analysis found a pooled effect size of 0.37 for the association between mood dysregulation and PTSD (Villalta, Smith, Hickin, & Stringaris, 2018), suggesting that they are robustly related.

Although not listed in the DSM-5 as an associated feature, it is also common in children with obsessive-compulsive disorder (OCD) to experience severe tantrums when their rituals are obstructed, especially in those who also suffer from depression (Krebs et al., 2013).
In anxiety disorders, the expression of irritability is associated with the stimuli or situation that causes distress and fear, usually as a way to avoid exposure to such stimuli. For example, a child with separation anxiety might show anger toward the person who forces the separation. And an adolescent with severe social anxiety might have a tantrum in order to not attend school. When the stimulus or situation that causes distress disappears or is avoided, irritability also tends to wane. In GAD, irritability is associated with excessive worry about daily life issues. This context-dependent pattern is what differentiates the irritability seen in anxiety disorders from the more severe chronic irritable mood characteristic of DMDD.

1.4.3.2  Epidemiology

Anxiety disorders are the most common type of disorder in children and adolescents, with an estimated prevalence of 30%, with rates for individual disorders ranging from 2% for GAD to 19% for specific phobia (Merikangas et al., 2010). Irritability as a symptom is also very common in anxiety disorders. A study in 239 treatment-seeking youth with GAD showed that impairing irritability was present in more than 90% of cases (Comer, Pincus, & Hofmann, 2012). Moreover, the authors found that irritability was the stronger predictor of GAD, increasing the odds of having GAD in children by 12, even after controlling for the presence of depression. However, irritability is also significant in other anxiety disorders. In a more recent study, Stoddard et al. (2014) found that irritability as a symptom was a problem for 50% of youth with anxiety disorders. When comparing parent- and self-reported levels of irritability across four groups: youth with anxiety disorders (including GAD, separation anxiety and social phobia), youth with SMD, youth with BD, and a healthy comparison group, the authors found that irritability was higher in those with
anxiety than the healthy group regardless of the informants. Furthermore, parent-reported irritability uniquely predicted social phobia and separation anxiety disorder whereas child-reported irritability was predictive of GAD.

1.4.3.3 Psychopathological correlates

A recent study using a latent construct approach showed that the association between anxiety and irritability remained large even after accounting for other comorbid disorders such as depression and ODD. Moreover, the magnitude of that association was not moderated by sex, age, or the presence of an anxiety disorder (Cornacchio, Crum, Coxe, Pincus, & Comer, 2016). In the study by Stoddard et al. (2014), parent-reported irritability was higher in children with anxiety and comorbid SMD or BD than those in the anxious-only group. Self-reported irritability was not different among the three patient groups.

Child anxiety disorders are not only associated with irritability, but also have high comorbidity rates with disorders in which irritability is a cardinal component, such as depression (Cummings, Caporino, & Kendall, 2014), ODD (Fraire & Ollendick, 2013), and DMDD (Copeland et al., 2013). In community samples, irritability is prospectively associated with GAD and specific phobias (Leibenluft et al., 2006; Stringaris et al., 2009). Similarly, as discussed earlier DMDD has also been found to be associated with future anxiety disorders (Copeland et al., 2014).

1.4.4 Irritability in ADHD

Irritability in children with ADHD, which is often referred to as emotional or mood lability or dysregulation in the ADHD literature, is common both in epidemiological and clinical samples (Shaw, Stringaris, Nigg, & Leibenluft, 2014). More than 35% of children with
ADHD in a population-based study showed marked irritability (Stringaris & Goodman, 2009b), and the presence of irritability in those with ADHD was associated with increased impairment, independent of the severity of ADHD symptoms and other psychiatric comorbidities. In a clinical sample of 696 children with ADHD, a recent study found that 91% of children presented with at least one irritability symptom and the rate of comorbid DMDD was 31%. Moreover, irritability in ADHD was associated with higher rates of anxiety, depression and family history of depression (Eyre et al., 2017).

A twin study in the general population (Merwood et al., 2014) found that the overlap between irritability and ADHD is mostly accounted for by shared genetic risks, rather than environmental factors. In line with these results, a recent study has shown that ADHD and early irritability share a genetic liability (Riglin et al., 2017). However, it is also plausible to consider that irritability might be a consequence of “evocative transactions” in ADHD (Rutter et al., 1997), whereby hyperactivity or impulsivity in the child evoke a negative response from the environment. Although this hypothesis has not been directly tested, the fact that irritability improves upon treatment of ADHD (Fernandez de la Cruz et al., 2015) argues in favour of this model. In addition, as mentioned above, the manifestations of irritability in children with ADHD fluctuate according to external stimuli and environmental changes.

1.4.5 Irritability in ASD

Irritability is a common problem in youth with ASD and is usually one of the reasons for referral to psychiatric services for these children. Data from community samples suggest that DMDD criteria is met by 20% of youth with ASD; and that the severity of irritability is not associated with the severity of ASD symptoms, but rather with higher rates of
comorbidity with anxiety and ADHD, as well as significant psychosocial impairment, particularly at schools (Copeland, Simonoff, & Stringaris, 2016). In clinical samples, up to 45% of children with ASD might present DMDD symptoms (Mayes et al., 2015b).

Youth with ASD and severe irritability have more relatives who also suffer from affective disorders than youths with ASD only (Simonoff et al., 2012), suggesting a shared risk. In addition, the irritability dimension of ODD assessed in children with ASD was associated with internalising as opposed to externalising symptoms (Mandy et al., 2014), mirroring the findings from community-based samples (Stringaris & Goodman, 2009c). Cognitive rigidity, as assessed in a card sorting task, has been also associated with irritability in youth with ASD (Simonoff et al., 2012), which could suggest that the temper outburst is usually triggered by changes in routines or schedules. Other factors that might increase the risk of irritability, and need to be carefully assessed in children with ASD, are communication impairments that compromise their ability to express internal states or coexisting physical symptoms such as pain (Malone, Delaney, Luebbert, Cater, & Campbell, 2000). Youth with ASD and irritability have also been shown to have blunted cortisol responses and decreased heart rate reactivity to a stress test, compared to those without irritability (Mikita et al., 2015).

1.5 Assessment and measurement of irritability

1.5.1 Reliability in the measurement of irritability

The term reliability refers to whether a measurement yields consistent results across different conditions or time points (Cronbach & Meehl, 1955). A reliable measurement of a psychological construct, such as irritability, is a pre-requisite for testing its validity (i.e.
testing whether the measure captures what it is purporting to be measuring). The reliability of measures of irritability has been assessed in the following ways:

**Internal consistency**: This assesses how well items of a scale correlate with each other. An implicit assumption underlying the assessment of irritability is that it is a unitary construct. This assumption is supported by current evidence showing the unidimensional structure and high internal consistency of instruments measuring irritability. The average internal consistency of irritability scales is high ($\alpha=0.75$) (Dougherty et al., 2013; Ezpeleta & Penelo, 2015; Herzhoff & Tackett; Leadbeater & Homel, 2015; Stringaris et al., 2009; Stringaris et al., 2012a; Stringaris et al., 2012b; Whelan et al., 2013) ranging from 0.49 in ad-hoc created scales (Stringaris et al., 2009) to 0.92 for specifically-developed irritability scales (Stringaris et al., 2012a). The unidimensional structure and good reliability of assessment of irritability has been found in different psychiatric conditions, including youth with high-functioning autism (Mikita et al., 2015), youth with SMD/DMDD, youth with BD, youth at-risk of BD and healthy youth (Meffert, Vidal-Ribas, Leibenluft, Brotman, & Stringaris, 2017). Therefore, irritability seems to be a transdiagnostic feature that can be reliably measured regardless of its psychopathological context.

**Test-retest reliability**: This assesses how stable a measured construct is over time. In the case of irritability, the goodness of test-retest reliability differs according to whether irritability is measured continuously or categorically. Studies employing dimensional approaches show that irritability is moderately stable over time, with longitudinal correlations ranging from 0.29 to 0.88 (Leadbeater & Homel, 2015; Stringaris, 2011; Stringaris et al., 2012a; Whelan et al., 2013). For example, a recent study of 2,620 children aged 8-9 years followed for more than 10 years found that both parent- and self-reports of
irritability were moderately correlated over time (r parent, 0.32–0.49; r self, 0.31–0.45) (Roberson-Nay et al., 2015).

However, compared to other psychiatric disorders, the current categorical definition of irritability shows low stability over time. DSM-5 field trials has shown poor test-retest reliability for DMDD (kappa=0.25; 95%CI 0.15-0.36) (Regier et al., 2013). In a clinical sample of children with DMDD aged 6-12 years, only 19% met criteria at 1- and 2-years follow-ups (Axelson et al., 2012). Similarly, findings across four time points in the Great Smoky Mountain study showed that most youth with SMD (82.5%) met SMD criteria in one wave but only 1.4% met criteria in all four waves of assessment (Brotman et al., 2006). However, findings in the same sample showed that youth with either persistent anger or temper outburst had a 75% likelihood of having either of these problems 1 year later (Copeland et al., 2015).

Thus, it appears that irritability defined stringently and categorically remains chronic only in a small proportion of children. As mentioned in section 1.2.2.3, categorically-defined irritability probably lacks stability because an arbitrarily set threshold, such as the thresholds for the frequency of temper outbursts, the duration of irritable mood, or the maximum time allowed without symptoms, is not met. Yet, a high proportion of children who do not meet current DMDD criteria still present with impairing chronic irritability (Deveney et al., 2015). Recently, a study evaluated only the main symptoms of DMDD (i.e., irritable-angry mood and temper outburst) in a community sample of children followed over 8 years (Mayes et al., 2015a). The authors found that whereas rates of symptom remission were high (71%), the prevalence of new cases was also considerable.
Moreover, 29% of the participants with frequent DMDD symptoms at baseline also displayed these symptoms at follow-ups (Mayes et al., 2015a).

**Inter-rater reliability:** This assesses the level of agreement between reports from different informants (e.g., parents and child). The agreement between parent and child ratings of irritability on a questionnaire is about $r=0.30-0.40$ (Aebi et al., 2013; Roberson-Nay et al., 2015; Stoddard et al., 2014; Stringaris et al., 2012a), which is in the upper range of correlation coefficients reported in meta-analyses of other scales (Achenbach, McConaughy, & Howell, 1987). For example, in a German sample of 1,031 children aged 10-18 (Aebi et al., 2013) the correlation between parent- and child-reports of irritability was $r=0.32$, very similar to the results found in a recent longitudinal study ($r=0.23$ to $0.36$) with 2,620 Swedish children (Roberson-Nay et al., 2015). These studies, though, employed ad-hoc measures of irritability. A study using a measure specifically designed to capture irritability found that parent and self-report scales were strongly correlated in both a US sample ($r = 0.58; 95\%CI 0.47–0.66$) and a UK sample ($r = 0.73; 95\%CI 0.56–0.85$) (Stringaris et al., 2012a).

In summary, irritability shows good internal consistency across diagnostics, and substantial longitudinal stability, at least when measured continuously. Test-retest reliability of the categorical construct of DMDD is poor, probably because children may fall just below threshold, but still show functional impairment due to irritability (Deveney et al., 2015). Inter-rater reliability of irritability between parent- and self-reports is consistent with the reliability of other behavioural and emotional constructs (Achenbach et al., 1987) or even larger when measured with instruments specifically designed to capture irritability.
1.5.2 Instruments for assessing irritability

In this section, I describe briefly the most common tools that have been used in the literature to ascertain irritability, some of which can be also employed in clinical settings. I will indicate which of these measures are used in the empirical chapters of this thesis.

Questionnaires:

In 2012, Stringaris and colleagues developed the Affective Reactivity Index (ARI), a concise scale with six items that allows for self-, parent-, and teacher report on a young person’s irritability. The items of the ARI inquire about the frequency, duration and threshold for irritability. It also has an item that asks about the impact of irritability on the child’s life. The ARI has good reliability and validity for both parent- and self-reported scales (Stringaris et al., 2012a). It is a short instrument designed to easily capture changes in irritability in clinical settings, but also for use in large studies where participants are often asked to complete many questionnaires. The ARI is one of the instruments employed in Chapter 4 of this thesis to capture dimensional ratings of irritability.

A comprehensive alternative is the Multidimensional Assessment of Preschool Disruptive Behavior (MAP-DB) Questionnaire, which is a valid instrument to assess irritability features in preschool children (Wakschlag et al., 2010). The MAP-DB Temper Loss subscale contains 20 items that assess features of tantrums and anger regulation with excellent reliability and covering a broad range of behaviours.

More specific consequences of irritability, such as aggressive behaviours, can be assessed with the Retrospective-Modified Overt Aggression Scale (R-MOAS), an adaptation of the Modified Overt Aggression Scale (MOAS). This is a parent-rated instrument that covers
the frequency of 16 aggressive behaviours during the preceding week grouped in four areas: verbal aggression, physical toward others, physical toward oneself and toward property. It has a good internal consistency (Cronbach’s alpha=0.82) and has been employed in several trials (Blader, Schooler, Jensen, Pliszka, & Kafantaris, 2009; Donovan et al., 2000).

Finally, several studies have employed specific items in the Child Behaviour Checklist (CBCL) (Achenbach, 1991) to capture irritability as well as other ODD dimensions in young people, although this measure is strictly a general psychopathology screening questionnaire (Aebi et al., 2013; Althoff et al., 2014; Stringaris et al., 2012b). Specifically, irritability items are “have a hot temper”, “mood/feelings change suddenly”, and “stubborn, sullen or irritable”. The CBCL is used in Chapter 3 of this thesis to assess irritability and other ODD dimensions.

**Semi-structured interviews:**

The most widely used clinical interview has been a version of the Kiddie Schedule for Affective Disorders Present and Lifetime Version (K-SADS-PL) (Kaufman et al., 1997) with an additional supplement for SMD, developed by Leibenluft and colleagues at NIMH in collaboration with Dr Joan Kaufman, PhD. The K-SADS-PL has been recently modified to fit the DMDD criteria. After the publication of DSM-5, the K-SADS-PL DMDD module was shortened for ease of use in clinical practice. The K-SADS-PL SMD/DMDD module is employed in Chapters 4 and 5 of this thesis to ascertain DMDD diagnoses.
Structured interviews:

The DMDD module of the Development and Wellbeing Assessment (DAWBA) is a structured interview for the ascertainment of DMDD that assesses the full range of DMDD symptoms and impairment. It is set up for convenient online completion and provides space for open-ended comments by the respondent. Although data with the DMDD module of the DAWBA have been collected in nearly ~9,000 children and adolescents from a community sample, the instrument itself has not yet been systematically validated. However, the validation is underway, taking the K-SADS DMDD module as a gold standard. Moreover, the data from this community sample will be the first data on DMDD collected with an instrument specifically designed to ascertain DMDD diagnosis. Previous epidemiological studies on DMDD have relied on ad-hoc measures from ODD or MDD items from other instruments (Althoff et al., 2016; Axelson et al., 2012; Copeland et al., 2013).

Change measures:

The Clinical Global Impression (CGI) scale for irritability is a clinician-rated instrument to assess the severity of irritability. It works analogously to the CGI for depression or bipolar disorder. It has two sections, the CGI severity scale (CGI-S) and the CGI improvement and change scale (CGI-I). The CGI is the primary outcome measure in the pharmacological trial reported in Chapter 5 of this thesis (ClinicalTrials.gov identifier NCT00794040) and in other psychological treatment trials of irritability that are currently underway (ClinicalTrials.gov identifier NCT02531893).

The Clinician ARI is another clinician-rated instrument that focuses on a number of irritability domains that include aggression. It is one of the first clinician-rated instruments
to assess specific behaviours in depth, requesting examples and enquiring about the context in which these problems occur. This instrument is currently being improved by the addition of other irritability dimensions that might include rumination or somatic symptoms.

1.6 Pathophysiological mechanisms of irritability

The causes of irritability are unknown. However, evidence from genetic and family studies, as well as behavioural and neuroimaging studies, can give us some ideas of the factors that contribute to the development and maintenance of irritability. The findings from these studies, summarised in Figure 1.7, are described in this section.

![Figure 1.7](image)

**Figure 1.7.** Mechanism contributing to the development and maintenance of irritability in youth. Genetic factors increase the risk of irritability. Environmental factors, the most studied being negative and inconsistent parenting behaviours, increase the risk of aberrant responses to rewards and threats. Similarly, deficits in instrumental learning contribute to both aberrant reward and threat processing in youth with irritability. Both types of aberrant responses interact with one another and increase the likelihood of irritability manifestations such as anger reactions, frustration and aggression. Adapted from Brotman et al (2017).
1.6.1 Genetic and family studies

The heritability of a trait (e.g., irritability) is estimated by comparing the phenotypic similarity of monozygotic twins, who are genetically identical (i.e., share 100% of genes), with that of dizygotic twins, who share on average 50% of their genes. Twin studies allow us to decompose the observed variance of a trait into its genetic, shared environmental and non-shared environmental components. Shared environmental factors are those environmental influences that make siblings in the same family to be similar to one another (e.g., growing up in the same neighbourhood, usually going to the same school and having the same education opportunities), whereas non-shared environment factors are environmental influences that make siblings in the same family to be different from one another (e.g., having different friends, experiencing different life events) (Pike, Reiss, Hetherington, & Plomin, 1996).

In the case of irritability, genetic factors contribute to 30%–40% of its variance in both adults (Coccaro, Bergeman, Kavoussi, & Seroczynski, 1997) and adolescents (Stringaris et al., 2012b), which is similar to the heritability estimates found for depression and anxiety (Eley, 1999). In addition, some evidence suggests that the irritability variance explained by genetic factors increases slightly over time in males and decreases in females (Roberson-Nay et al., 2015). Moreover, unique environmental factors, as opposed to shared, explain most of the remaining variance.

Genetic studies also provide information on the factors (i.e. genetic and environmental) that contribute to the covariance between two or more phenotypes (e.g., irritability and depression). In the case of ODD dimensions, the genetic correlation between irritability and headstrong/hurtful dimension is moderate to high (r=0.46-0.63) with the remaining
variance explained by non-shared environmental factors (Mikolajewski, Taylor, & Iacono, 2017; Stringaris et al., 2012b). However, twin studies show that the longitudinal association between irritability – as opposed to headstrong - and depression/anxiety is substantially explained by shared genetic factors (Mikolajewski et al., 2017; Stringaris et al., 2012b). These studies also showed that any differences between both phenotypes were explained by unique environmental factors. In addition, another study found that the genetic covariance between irritability and anxiety/depression was highest in early adolescence (74%) and that the impact of irritability on future internalising symptoms was higher than the impact of internalising symptoms on future irritability (Savage et al., 2015). Irritability in children also seems to be genetically associated with developmental disorders such as ADHD (Riglin et al., 2017). This is also in line with the significant genetic overlap between the broader construct of emotional dysregulation and ADHD symptoms (Merwood et al., 2014).

In addition to twin studies, family studies can be used to examine how phenotypes aggregate across generations which can be suggestive of genetic overlap between distinct phenotypes. Studies examining parents and children suggest that a family history of depression is associated with irritability in the offspring (Krieger et al., 2013; Wiggins et al., 2014). In addition, the relation between maternal history of depression and adolescent depression is partly mediated by the presence of irritability in childhood (Whelan et al., 2015). These findings are supported by studies examining categorical diagnoses of irritability (DMDD/SMD) in offspring. In one study, parents of youth with narrow BD were more likely to be diagnosed with BD than parents of youth with SMD (Brotman et al., 2007), results that have been replicated also in youth with DMDD (Fristad et al., 2016).
Youth with DMDD, instead, have more parents with a history of depression (Propper et al., 2017). However, one study showed no differences in parental history of psychiatric disorders between youth with and without DMDD (Axelson et al., 2012).

1.6.2 Behavioural and neuroimaging mechanisms

Taking a neuroscientific approach, irritability can be conceptualised as (1) aberrant responding to frustrating non-reward and (2) aberrant responding to threat (Brotman et al., 2017). Both types of responses have been described across species (i.e., not only humans); and in both cases, instrumental learning plays a pivotal role. Instrumental learning can be defined as the process through which an individual learns and adapts his/her behaviour in order to obtain rewards or avoid punishments. As I will describe below, youth with severe irritability have altered approaches to rewards and punishments due to difficulties in either learning or adapting their behaviours as a response to environmental stimuli. And these aberrant responses involve several brain regions, as shown in Figure 1.8.

![Figure 1.8](image)

**Figure 1.8.** Brain regions involved in the aberrant processing of reward and threat.
1.6.2.1 Aberrant response to frustrative non-reward

The concept of *frustrative non-reward*, introduced in section 1.2.3, refers to the normative, adaptive response, and the emotional state that is induced when an expected reward is not delivered. This concept was initially described in rodents by Amsel (1958), who noticed that activity and aggression increased when expected food was omitted. Increased activity when rewards are omitted has also been shown in humans (Dollard et al., 1939). The concept of blocked goal attainment is easily manipulable in an experimental setup and has been employed previously to study irritability in adults (Abler, Walter, & Erk, 2005; Siegrist et al., 2005) youth (Adleman et al., 2011; Deveney et al., 2013; Rich et al., 2007), young children (Perlman et al., 2015), and babies (Lewis & Ramsay, 2005; Lewis, Sullivan, & Kim, 2015). Aberrant responses to frustrative non-reward in irritable youth are evidenced by *deficits in reward learning* and *altered sensitivity to reward receipt and omission*.

Compared with healthy volunteers (HV), youth with SMD have deficits in learning reward contingencies and adapting their behaviours when these contingencies change (Adleman et al., 2011; Dickstein et al., 2007). The neural correlates of such deficits are altered responses in the cingulate, striatum and inferior frontal gyrus (IFG) (Adleman et al., 2011). Moreover, when rewards are omitted, youth with severe irritability show decreased striatal activity (Deveney et al., 2013) and decreased activation of frontal regions (Perlman et al., 2015) compared to controls, and their self-reported levels of frustration in such situations are correlated with irritability (Deveney et al., 2013; Rich et al., 2011; Tseng et al., 2017). In contrast, when rewards are delivered, irritable children show hyperactivation in frontal regions and report more positive mood than controls (Perlman et al., 2015). Frustrative non-reward is also associated with attentional dysfunction in irritable youth compared to
healthy volunteers following reward omission. Such attentional deficits correlate with
dysfunctional activations in the amygdala, striatum, cingulate and parietal cortex (Deveney
et al., 2013; Tseng et al., 2018).

1.6.2.2 Aberrant approach to threat

On the other hand, threats are situations or objects that signal potential harm for an
organism. Under perceived threat, an organism will response either by avoiding it as
commonly seen in anxious children, or fighting it, which is an approach response
(Fanselow, 1994; LeDoux & Pine, 2016; Rolls, 2014). The latter type of responses (i.e.,
angry fighting reactions) are more likely to happen when threats are imminent and
inescapable (Blair, 2012; Panksepp, 1990). The neural circuit associated with threat
responses include the prefrontal cortex, the amygdala, the hypothalamus and the
periaqueductal grey. Severe irritability has been associated with an increased approach
towards threatening stimuli, as well as dysfunction in these brain regions.

For example, compared with healthy volunteers, youth with severe irritability are more
likely to pay attention to threatening angry faces than neutral faces (Hommer et al., 2014;
Salum et al., 2017). Moreover, irritable youth also tend to interpret ambiguous or neutral
faces as more threatening (Brotman et al., 2010; Stoddard et al., 2016). The bias towards
threatening faces is not unique to youth with severe irritability but is also present in people
with depression and anxiety (Armstrong & Olatunji, 2012; Bar-Haim, Lamy, Pergamin,
Bakermans-Kranenburg, & van, 2007).

Irritable youth also have a generalised deficit in emotion recognition. That is, compared
with HV, youth with severe irritability make more errors when asked to recognize emotions
in faces (Guyer et al., 2007; Rich et al., 2008) and voices (Deveney, Brotman, Decker, Pine, & Leibenluft, 2012). And, as shown in Chapter 4 of this thesis, deficits in emotion recognition in irritable youth might partially explain the relationship between irritability and depression.

Deficits in facial emotion processing are associated with amygdala dysfunction in youth with severe irritability. Evidence suggests that amygdala hypoactivity may be present during explicit processing (e.g. rating fear or hostility) of face emotions (Brotman et al., 2010) while amygdala hyperactivity is evident during implicit processing (e.g., reporting nose width or gender) (Brotman et al., 2010; Thomas et al., 2013). Additionally, inappropriate modulation of amygdala activity has been seen in youth with severe irritability compared with HV as a response to increasing facial anger and happiness (Thomas et al., 2012; Wiggins et al., 2016).

1.6.3 Environmental factors

Twin studies show that most of the variance in irritability is explained by non-shared environmental factors (see section 1.6.1). Most of the existing studies examining the impact of these factors have focused on the effects that parenting behaviours have on the development and maintenance of irritability.

Discordant monozygotic (MZ) twin designs assume that any observed phenotypic differences between members of a MZ-twin pair (e.g. differences in irritability) may be attributable to non-shared environment because MZ-twins share 100% of their genetic background and grow up in, largely, the same environment (Vitaro, Brendgen, & Arseneault, 2009). These designs are therefore very suitable to study the effects of non-
shared environmental factors. To this end, Oliver (2015) employed a longitudinal monozygotic (MZ) twins difference design to examine the non-shared environmental links between negative parenting behaviour, conduct problems and irritability. Two main findings can be extracted from this study. First, the cross-sectional association between negative parenting and irritability explained by non-shared environment in early and late childhood was larger than the association between negative parenting and conduct problems. Second, while irritability in early childhood predicted negative parenting in middle childhood, the latter predicted higher levels of irritability in late childhood.

These findings highlight the bidirectional associations between parenting behaviours and irritability (Kiff, Lengua, & Zalewski, 2011b). In line with previous models, it might be that irritability in early life (e.g. in the form of temperament) shapes parenting behaviours negatively, and negative parenting might impact on the development of irritability over time (Belsky, 1984). For example, it has been shown that irritable temperament predicted inconsistent discipline, which in turn was predictive of greater irritability (Lengua & Kovacs, 2005). Early findings from a longitudinal study of temperament’s effect on social development has shown that children with persistently high levels of irritably in childhood (from 2 to 7 years of age) were more likely to show higher levels of irritability in early adolescence if they were exposed to negative parenting behaviours (Schneider, Filippi, Pine, & Leibenluft, 2018).

In summary, it seems that irritable youth are exposed to inconsistent parenting environments which may unintentionally reinforce angry responses (Becht, Prinzie, Dekovic, van den Akker, & Shiner, 2016). However, given the bidirectional associations mentioned above, it is plausible to think that environmental influences interact with genetic
factors, giving rise to dysfunctional parent-child interactions (O'Connor, Deater-Deckard, Fulker, Rutter, & Plomin, 1998). For example, the increased anger that arises from aberrant responses to frustrative non-reward and threat, which may have a genetic component, will have an impact on the relationships within the family (Figure 1.9). Parents might respond with negative affect to the child’s anger, often also in the form of anger or aggression, leading to a self-reinforcing pattern that is “mutually coercive” (Patterson, 1982). In addition, parents will eventually give in and acquiesce to a child’s request and thus reinforce angry behaviours. Dysfunctional parent-child interaction cycles have been extensively described in relation to the aetiology and maintenance of ODD (Barkley, 2013a). Thus, it seems plausible to think that similar mechanisms operate also for irritability. And, as I describe in the next section 1.7.1.1, parenting interventions that aim to increase consistent reinforcement of child’s behaviours are within the most effective treatments for irritability and disruptive behaviours.

1.6.4 Early predictors of irritability in children and adolescents

Very few studies have examined predictors of chronic severe irritability – either defined as DMDD or ODD dimension – in young people.

To my knowledge, only two studies have examined predictors of DMDD. In a preschool sample of children aged 3 years, Dougherty et al. (2014) examined predictors of having DMDD at six years of age. The authors found that DMDD at age six was independently predicted by having ADHD, ODD and high CBCL-DP scores at age three. Moreover, other independent predictors at age three included poor peer functioning, a child temperament characterised by high surgency, negative emotional intensity and low effortful control, as well as parental hostility and parent history of substance abuse.
The second study predicting DMDD employed a large community sample (Munhoz et al., 2017) to examine perinatal and postnatal risk factors of DMDD at age 11. The authors found that maternal mood symptoms during pregnancy, maternal depression during the first years after birth, and low maternal level of education were associated with a higher risk of developing DMDD. However, in this study, mood symptoms during pregnancy were identified with just an affirmative answer to a single question, and the instrument to ascertain DMDD in the community, the DMDD module from the DAWBA (see section 1.5.2), has not yet been validated.
In terms of predictions of ODD irritability dimension, as far as I know, only one study has used ODD dimensions as outcomes. In this study, the authors examined neuropsychological correlates at age three as predictors of ODD dimensions at age six (Griffith, Arnold, Rolon-Arroyo, & Harvey, 2017). The authors found that working memory, control inhibition and sustained attention at age three were associated with the ODD dimension of negative affect (i.e., irritability) at age six, whereas delay aversion predicted the oppositional behaviour dimension (i.e., headstrong).

Certainly, more studies are needed that examine predictors of chronic severe irritability.

1.6.5 Physiological correlates of irritability

To date, only one study had examined the association between physiological responses to stress and chronic severe irritability, in this case, defined as the ODD irritable dimension (Silva et al., 2014). Chapter 3 in this thesis reports the results of the second study that has examined this association.

Specifically, Silva et al. (2014) examined the association between sympathetic skin response (SSR) to mild electric stimuli and ODD dimensions in a community-based sample of adolescents. They found SSR to be associated with the ODD headstrong dimension but not with the ODD irritability dimension. They suggested that SSR is usually associated with acting-out behaviours such as aggression, which is more linked to headstrong than irritability (Stringaris & Goodman, 2009c).

Although there are no more studies testing associations between stress reactivity and chronic severe irritability, several studies have examined associations between stress reactivity and irritability/anger as a trait. For example, anger manifested inappropriately
based on the context, was associated with lower levels of baseline cortisol and lower cortisol reactivity in boys but not girls in a sample of 6-10-year-old children (Locke, Davidson, Kalin, & Goldsmith, 2009). Similarly, in a sample of boys with ASD, Mikita et al. (2015) found that higher levels of parent- and self-reported irritability were associated to dampened cortisol reactivity to stress and lower overall cortisol levels, respectively. These findings suggest that, at least in boys, the association between stress response and irritability might be negative. The results of Chapter 3 in this thesis suggest similarly.

1.7 Treatment approaches to irritability

Currently, there are no well-established pharmacological or psychosocial treatments specifically designed for DMDD (Benarous et al., 2017; Stringaris et al., 2018; Tourian et al., 2015). However, as I describe below and show in Chapter 5 of this thesis, new interventions are being developed and tested. Nevertheless, there are several evidence-based treatments that have been shown to be effective for disorders in which irritability is a common symptom. For example, parent management training (PMT) has been widely used to treat disruptive behaviours in ADHD or ODD (Barkley, 2013a, 2013b). In this section, I describe psychological and pharmacological treatments that have been tested to manage irritability either in the context of other disorders (e.g. ADHD, ASD) or, as a very preliminary approach, in youth with SMD/DMDD. Figure 1.10 shows an algorithm for the treatment of irritability based on the primary disorder in which it appears. It is recommendable to treat any comorbid condition first using evidence-based treatments. This is because there have been many more trials focused on depression, ADHD or ODD, than irritability itself. This treatment approach relies on secondary analyses that have been conducted with irritability from those trials.
**Figure 1.10.** Treatment algorithm for irritability. These treatments for irritability should only be provided when the comorbidity labelled above has been ascertained after a complete clinical assessment. Continuous arrows show paths to first-line interventions, and dashed arrows show paths to second-line interventions after application of only first-line interventions has proven to be futile. Adapted from Stringaris et al (2017).

### 1.7.1 Psychological treatments

Several psychosocial interventions have been developed for clinical conditions in which irritability is a common symptom. These interventions can be split into two categories, namely, parent management training (PMT) and cognitive behavioural therapy (CBT) (Sukhodolsky, Smith, McCauley, Ibrahim, & Piasecka, 2016).
1.7.1.1 Parent management training (PMT)

As mentioned in section 1.6.3, negative parenting behaviours are one of the factors associated with the development and maintenance of irritability in youth. Since parenting behaviours are modifiable, it is therefore sensible to target them for intervention. Based on principles of instrumental learning and the research in parent-child interactions conducted by Patterson and colleagues (Patterson, 1975), PMT targets negative patterns of family interactions underlying children’s disruptive behaviour and aggression. Specifically, to address irritability and anger, during PMT parents are taught to consistently reinforce children’s prosocial behaviour and to consistently not reinforce maladaptive behaviour such as temper outbursts (Barkley, 2013b). Through positive reinforcement, parents are instructed to focus on youths’ positive behaviours by providing praise or a reward after an identified target behaviour occurs. Problematic child behaviour is addressed through parents using active ignoring and time-outs. Several meta-analytic studies provide evidence for the general efficacy of PMT in decreasing disruptive behaviours and conduct problems in youth with ODD and/or CD (Comer, Chow, Chan, Cooper-Vince, & Wilson, 2013; Eyberg, Nelson, & Boggs, 2008; Furlong et al., 2012; Knapp, Chait, Pappadopulos, Crystal, & Jensen, 2012; Morrison, Pikhart, Ruiz, & Goldblatt, 2014; NICE, 2013; Sandler, Schoenfelder, Wolchik, & MacKinnon, 2011). In the case of youth with ODD, those youth with irritability, but not headstrong symptoms, seem to be particularly responsive to parenting interventions (Scott & O’Connor, 2012). There is also evidence for the usefulness of parenting interventions in addressing disruptive behaviours in children with ASD (Bearss et al., 2015). Moreover, some studies suggest that improvements in prosocial behaviour following PMT are stable and associated with fewer risk of antisocial behaviours (Scott, Briskman, & O’Connor, 2014).
1.7.1.2 Cognitive behavioural therapy (CBT)

The principles of CBT are based on social information processing and behavioural theories. From this perspective, disruptive behaviours are conceptualised as the results of aberrant interpretations of social cues (Crick & Dodge, 1994; Dodge, 1980). It is assumed that these behaviours are learnt and thus modifiable. In CBT, therapists work directly with children and adolescents as they learn and practice more adaptive ways to interpret and respond to social cues, with the aim of reducing anger and aggressive responses. There are several CBT manuals for aggressive behaviour, anger, and irritability, including *Anger Coping Program* (Lochman, Powell, Boxmeyer, & Jimenez-Camargo, 2011) and *CBT for Anger and Aggression in Children* (Sukhodolsky & Scahill, 2012).

1.7.1.3 Development of psychological treatments for severe irritability

In the last years, efforts have been made to develop and test psychological therapies specifically for youth with chronic severe irritability (e.g., SMD or DMDD). For example, a case report of a young girl with DMDD treated with CBT was recently published (Tudor, Ibrahim, Bertschinger, Piasecka, & Sukhodolsky, 2016). In addition, Waxmonsky and colleagues (2016) recently showed the feasibility of a integrative therapy for youth with ADHD and SMD, that combined group-based PMT, CBT and treatment with stimulant medication, after a promising pilot study with a small (N=7) sample (Waxmonsky et al., 2013). Furthermore, a recent randomised controlled trial (RCT) demonstrated the feasibility and preliminary efficacy of a modified Dialectical Behaviour Therapy (DBT) for preadolescent children with DMDD as compared with treatment as usual (TAU) (Perepletchikova et al., 2017). Finally, a small study (N=19) testing Interpersonal Psychotherapy for mood and behaviour dysregulation (IPT-MBD) in adolescents with
DMDD/SMD also showed clinical improvements following a 24-week intervention as compared with TAU (Miller et al., 2018). However, all these studies need replication in larger samples or in youth with DMDD with different comorbidities than ADHD.

Based on the pathophysiological mechanisms described in section 1.6.2. (i.e., aberrant response to reward and threat), the frequent overlap of irritability with anxiety and the efficacy of exposure techniques in treating anxiety, novel therapies are being developed at the National Institute of Mental Health (ClinicalTrials.gov identifier NCT02531893). Specifically, researchers are testing the hypothesis that irritable children will develop an increased tolerance for frustration following graded exposure to frustrating situations, through extinction and more adaptive coping strategies. Exposure is combined with joint parent-child sessions, in which parents are taught PMT strategies, and learn ways to tolerate their own discomfort if the child has an outburst while the parent is engaging in PMT strategies (e.g., active ignoring of child’s outbursts). Preliminary results from an open pilot of exposure-based CBT for severe irritability in 10 youth with DMDD showed improvements after 12-16 weekly sessions, including reduced levels of clinician-rated irritability (Kircanski, Clayton, Leibenluft, & Brotman, 2018a).

Based on the observed attentional biases towards threatening faces in irritable youth and their tendency to rate ambiguous faces as more threatening, preliminary studies are also exploring novel computer-based treatment interventions that focus on modifying these aberrant threat responses. Specifically, interpretation bias training (IBT) is a cognitive retraining approach that requires youth to make forced, two-choice (happy/angry) judgments of facial expressions on a linear morph continuum from happy to angry. Patients receive feedback encouraging them to learn new, positive associations to
ambiguous face emotions. Evidence suggests that a short training might reduce aggression in youth at risk for criminal behaviour (Penton-Voak et al., 2013). Recently, an open trial training DMDD youth to perceive ambiguous faces in a more benign manner was associated with decreased irritability (Stoddard et al., 2016). RCTs for both of these new therapies, i.e. exposure-based CBT and IBT, are being tested in larger samples.

1.7.2 Pharmacological treatments

Chapter 5 of this thesis presents results from the second pharmacological RCT that has ever been conducted in youth with severe irritability. So far, only one RCT for the pharmacological treatment of irritability in youth with SMD has been completed, showing negative results, as I describe below. However, we have evidence of secondary analyses of data from trials treating other psychiatric disorders (e.g., ADHD, depression, ASD) or trials testing irritability-related conditions (e.g. aggression) that suggest that some pharmacological approaches might be effective in reducing irritability. Nevertheless, the evidence from these studies should be interpreted in the context in which it is examined; that is, for example, a medication that shows to be effective in reducing irritability in ADHD might not work in irritable youth without ADHD. Below I describe the evidence available grouped by type of medication.

Lithium:

As I have mentioned above, there has only been one pharmacological RCT for severe irritability in youth, in this case with lithium (Dickstein et al., 2009). As a result of the controversy as to whether chronic severe irritability was a manifestation of paediatric BD, the authors tested the efficacy of lithium in treating irritability in youth with SMD. In this small study (N=25), there was no benefit of lithium over placebo in the treatment of severe
irritability in youth with SMD. It should be noted that lithium may reduce aggression in children and adolescent with conduct disorder (Campbell et al., 1995; Campbell et al., 1984; Malone et al., 2000; Rifkin et al., 1997). However, aggression is only one of the many consequences of irritability, and in the case of CD might not be always associated with irritability.

**Selective serotonin reuptake inhibitors:**

Indirect evidence suggests that serotonin reuptake inhibitors (SRIs) might be efficacious in the treatment of irritability in depressed adults with anger attacks (Fava & Rosenbaum, 1999), and those with intermittent explosive disorder (Coccaro, Lee, & Kavoussi, 2009), or premenstrual syndrome (Dimmock, Wyatt, Jones, & O'Brien, 2000). In children and adolescents, the authors of a recent systematic review on the effects of antidepressants on irritability, aggression and ODD symptoms (Kim & Boylan, 2016), found only two uncontrolled studies reporting irritability as an outcome. Both studies showed substantial improvements of irritability following treatment with SRI, especially with citalopram (Armenteros & Lewis, 2002; Garland & Weiss, 1996).

**Stimulants:**

Meta-analyses have shown medium to large effect sizes in the efficacy of stimulants in reducing aggression and emotional instability in youth with ADHD (Connor, Glatt, Lopez, Jackson, & Melloni, 2002; Shaw et al., 2014). Specifically, stimulant monotherapy in youth with ADHD, in combination with behavioural treatment, decreases aggressive behaviour and temper outbursts (Blader, Pliszka, Jensen, Schooler, & Kafantaris, 2010; Blader et al., 2016), although half of the participants remain refractory to treatment. Post-
hoc analyses in the multimodal treatment study of children with ADHD (MTA) found that stimulant medication alone (ES=0.63) or in combination with behavioural treatment (ES=0.82) was better than behavioural treatment alone (ES=0.42) in reducing irritability symptoms (Fernandez de la Cruz et al., 2015). However, whereas the presence of irritability did not moderate the treatment effects on ADHD symptoms, the magnitude of the effect sizes for the irritability response to treatment was approximately half of that for ADHD symptoms (The MTA Cooperative, 1999).

In the last years, evidence from several open-label trials with stimulant in youth with ADHD and SMD/DMDD has emerged. As mentioned in section 1.7.1.3, Waxmonsky et al. (2016) found that stimulant combined with PMT and CBT might be efficacious to treat irritability in this population. The same group has also shown that stimulant alone might decrease behavioural and mood symptoms in youth with ADHD and comorbid DMDD (Baweja et al., 2016). A recent small open-label trial (N=22) has demonstrated reductions in irritability with medium to large effect sizes in these children following stimulant monotherapy (Winters, Fukui, Leibenluft, & Hulvershorn, 2018). Finally, systematic reviews and meta-analysis have shown that stimulants are not associated with an increase in irritability as a side effect (Ahmann et al., 1993; Manos et al., 2011; Stuckelman et al., 2017).

Antiepileptics:

The anticonvulsants sodium valproate or carbamazepine are commonly employed as mood stabilisers for the treatment of BD in youth. However, results from an RCT showed that mood stabiliser divalproex added to stimulant was also effective in children with ADHD whose aggression had not responded to stimulant treatment alone (Blader et al., 2009).
Nevertheless, caution is warranted in their use in young people, particularly for valproate in young females due to its association with developmental disorders and congenital malformations in children exposed in-utero (MHRA, 2016).

Atypical antipsychotics:

Risperidone has been shown to be effective in reducing irritability-related behaviours such as aggression in children with intellectual disability (Aman et al., 2002), disruptive behaviour disorders such as ADHD, ODD, and CD (Loy, Merry, Hetrick, & Stasiak, 2017) and autism spectrum disorders (ASD) (RUPPAN, 2002; van Schalkwyk et al., 2017). Indeed, risperidone along with aripiprazole both have an FDA indication for the treatment of such behaviours in ASD, with similar effect sizes (d=0.78 and d=0.86, respectively) (Fung et al., 2016). However, it is important to remember that antipsychotics carry with them strong adverse effects, including obesity, iatrogenic diabetes and endocrine abnormalities (Findling, Aman, Eerdekens, Derivan, & Lyons, 2004) even when combined with stimulant medication (Loy et al., 2017). Moreover, these side effects may be more pronounced in those with ASD and young people (Correll et al., 2006; Correll, Sheridan, & DelBello, 2010; De Hert, Dobbelaere, Sheridan, Cohen, & Correll, 2011).

In youth with severe irritability and without ASD (i.e., SMD or DMDD) there have been only two open-label trials to date examining the effects of antipsychotics. First, Krieger and colleagues (2011) examined the efficacy of low doses of risperidone in the treatment of SMD showing reductions in irritability scores. And recently, an open trial of aripiprazole combined with stimulant in youth with ADHD and DMDD showed reductions in irritability after 6-weeks of treatment (Pan, Fu, & Yeh, 2018).
In sum, beyond RCTs in children with ASD, there has only been one RCT with irritability as a primary outcome that showed no benefits of lithium over placebo (Dickstein et al., 2009). Several open-trials have been conducted, but the results of these need replication in larger samples and in controlled designs. Nevertheless, some medications such as stimulants seem to be effective in reducing irritability in at least some youth with ADHD and comorbid DMDD.

1.8 The Aims of this Thesis

This thesis began by highlighting the wide presence of irritability across psychiatric disorders in young people (APA, 2013). However, as mentioned in several sections throughout this chapter, data from individual studies show that severe chronic irritability, either in the form of DMDD/SMD or as ODD dimension, is particularly related to depression and other internalising disorders in longitudinal studies (Copeland et al., 2014; Dougherty et al., 2016; Stringaris et al., 2009; Stringaris & Goodman, 2009a).

Apart from a shared genetic variance (Savage et al., 2015; Stringaris et al., 2012b), the reasons for this distinct relation between irritability and depression are unknown. However, if proven to be robust and consistent, this finding opens the door not only to test other shared risks but also to inform treatment approaches for irritability.

The overarching aims of this thesis are therefore to explore the consistency of the association between irritability and depression, to examine potential shared risks between the two problems, and to test a new treatment approach that is motivated by the close association between depression and irritability. To do so, this thesis employs cross-sectional and longitudinal analyses of epidemiological, physiological, behavioural and
clinical data. For the empirical studies of this thesis, two data sources were used. First, data collected as part of the Wirral Child Health and Development Study, a prospective epidemiological study that, in the current thesis, spans ages 29 weeks to 5 years old. Second, data collected from a clinically referred sample of children and adolescents aged 7 to 20 years old who were recruited at the National Institute of Mental Health in Bethesda (Maryland, United States). Below, I briefly describe the rationale and specific aims for each of the studies included in this thesis.

1.8.1 Longitudinal outcomes of irritability

As mentioned above, findings from several individual studies show that chronic severe irritability is longitudinally associated with internalising psychopathology, including depression and anxiety (Copeland et al., 2014; Rowe et al., 2010; Stringaris et al., 2009). However, there has been no systematic attempt at quantifying this association across studies. Systematically reviewing and quantifying this association is important for several reasons. For example, it might be that studies only test associations with these outcomes, then closing the door to the possibility of finding associations with other types of psychopathology. Also, it is possible that only studies that find positive predictions of internalising problems are published, whereas those with null findings are not due to publication bias.

In addition, even if there were a specific association, it is unclear how strong this association is, and which factors affect its strength. For example, does the strength of this association differ between conceptualisations of chronic severe irritability? That is, does DMDD/SMD predict depression or anxiety differently than the ODD irritability
dimension? Is this association specific for internalising disorders or is it also true for internalising symptoms?

To provide answers to these questions, Chapter 2 of this thesis systematically reviews and meta-analyses longitudinal studies in which chronic severe irritability is the predictor and any type of psychopathology, either in the form of psychiatric disorder or psychiatric symptoms, is the outcome. The aim of this study is to test the consistency in the specific association between irritability and future internalising disorders, against all other psychiatric outcomes ever tested in this context.

1.8.2 Early predictors of irritability

As will be evident in Chapter 2, chronic severe irritability is a robust predictor of depression in longitudinal studies. In other words, youth with severe irritability are at risk of developing depression. However, little is known about the precursors of severe irritability; that is, what are the risk factors for chronic severe irritability? What makes young people develop irritability and depression, concurrently or sequentially? Are such early-life risk factors for irritability the same than those for depression?

Two well-known factors for the development of depressive disorder are being a female and increased reactivity to stress. After puberty, females are twice as likely as males to develop depression. One of the reasons for such a difference is how the impact of stress differs between males and females (Nolen-Hoeksema, 2001). However, it is also possible that females are also more likely to present irritability, thus increasing the likelihood of depression in females. Indeed, some studies show that females present more irritability symptoms than males, whereas the latter present more headstrong symptoms (Stringaris et
al., 2012b; Trepat & Ezpeleta, 2011). In addition, exposure to stress, especially if stress is chronic and early in life, is highly associated with the development of depression (Pizzagalli, 2014). Evidence from animal and human studies suggests that the association between stress and depression is mediated by the impact that stress has on the brain reward circuit (Ironside, Kumar, Kang, & Pizzagalli, 2018; Novick et al., 2018; Stanton, Holmes, Chang, & Joormann, 2018). Dysfunction of reward processing is a hallmark of depression (Keren et al., 2018) and have been shown to predict the development of depression in previously healthy adolescents (Stringaris et al., 2015). As described in section 1.6.2.1, irritability is also associated with altered responses to reward. Therefore, it is possible that stress also plays a role in the development of irritability. However, whether stress and sex interact to predict irritability in early life is still unclear.

As mentioned in section 1.2.2.3, normative irritability peaks during the preschool years (Leibenluft & Stoddard, 2013; Wakschlag et al., 2012), and it is associated with functional impairment, especially at increased levels (Copeland et al., 2015), but also future depression (Dougherty et al., 2013). Surprisingly, few studies have examined chronic severe irritability at this age (Dougherty et al., 2013; Ezpeleta et al., 2012); and none of these employed irritability as outcome. Indeed, very few studies have employed irritability as outcome.

To address this knowledge gap, we present in Chapter 3 a longitudinal study of young children up to age 5 where we test whether sex and physiological stress reactivity interact to predict distinct outcomes in girls and boys. Specifically, we predict that distinct stress responses will be associated with more irritability in girls and more headstrong symptoms in boys.
1.8.3 Cognitive mechanism of irritability

In section 1.6.2.2, I have described one of the main the mechanisms that contribute to the development and maintenance of irritability, namely, aberrant responses to threat. Evidence suggests that youth with irritability show an attentional bias towards threatening faces (Hommer et al., 2014; Salum et al., 2017) and tend to interpret ambiguous or neutral faces as more threatening (Brotman et al., 2010; Stoddard et al., 2016). In addition, compared with healthy volunteers, youth with severe irritability show difficulties in recognizing emotions in faces (Guyer et al., 2007; Rich et al., 2008) and voices (Deveney et al., 2012).

However, bias to negative emotions are not specific to irritability and has also been reported in youth with depression (Armstrong & Olatunji, 2012; Leppanen, 2006; Peckham, McHugh, & Otto, 2010), and people with depression also show deficits in the recognition of emotions (Dalili, Penton-Voak, Harmer, & Munafo, 2015; Kan, Mimura, Kamijima, & Kawamura, 2004). Furthermore, such deficits are predictive of future depressive symptoms (Beevers & Carver, 2003; Vrijen, Hartman, & Oldehinkel, 2016).

Therefore, it is unclear whether such deficits in youth with severe irritability are specific to irritability itself or might be partially explained by co-occurring depressive symptoms. Importantly, it is possible that such deficits increase the risk for depression in youth with severe irritability. To answer these questions, Chapter 4 examines emotion recognition in DMDD youth with different levels of depressive symptoms. This study also employs a longitudinal design to test the hypothesis that these deficits are predictive of depressive symptoms in youth with severe irritability.
1.8.4 Treatment outcomes of irritability

The final part of the results’ section of my thesis reports on the use of a pharmacological treatment, antidepressant citalopram, for severe irritability.

As described in section 1.7.2, treatments specifically designed for severe irritability are scarce. Indeed, only one RCT has been conducted comparing the response to lithium against placebo in youth with SMD (Dickstein et al., 2009). This trial was motivated by the controversy around paediatric BD (Wozniak et al., 1995). That is, the rationale behind the trial was that if chronic severe irritability was a manifestation of BD in children, therefore, as it does in adults, lithium should be the treatment of choice. However, the results of the trial showed no benefit of lithium over placebo in the treatment of chronic irritability (Dickstein et al., 2009).

Since then, no other pharmacological RCTs for the treatment of severe irritability have been published. However, several secondary analysis of data and a number of open-trials suggest that stimulant medication might reduce irritability in those youth with ADHD symptoms and severe irritability (Baweja et al., 2016; Fernandez de la Cruz et al., 2015; Waxmonsky et al., 2016; Winters et al., 2018). Nevertheless, about half of them remain refractory to treatment (Blader et al., 2016).

As evidenced in Chapters 2 and 4, chronic irritability is specifically associated, and shares risks factors, with depression. Therefore, following a similar rationale as in the first RCT with lithium, it is plausible that if chronic irritability is especially linked to depression, then the treatment of choice for depression (i.e. serotonin reuptake inhibitors) should also work
for chronic irritability. Indeed, some evidence suggests that SRIs might reduce irritability in youth (Kim & Boylan, 2016).

To this end, the aim of Chapter 5 is to test whether adding an SRI (i.e., citalopram) to stimulant medication (i.e., methylphenidate) is more effective in reducing irritability in youth with SMD/DMDD than adding placebo.
CHAPTER 2: Longitudinal correlates of chronic severe irritability – a meta-analysis

2.1 Background

As mentioned in the introduction, irritability is one of the few symptoms to cut across internalising and externalising disorders. In the DSM-5, irritability is listed as a cardinal or associated symptom of several psychiatric conditions, particularly in children and adolescents (APA, 2013). However, several individual studies highlight the specific association between chronic severe irritability and future internalising disorders, including depression and anxiety, as opposed to externalising disorders such as ADHD or CD. Yet, a systematic account of these associations and their quantification is lacking.

The association between chronic irritability and future depression/anxiety has been demonstrated using the two conceptualisations of chronic irritability described in section 1.2.1; these are DMDD and the irritability dimension of ODD. In a prospective population-based study, children and adolescents with DMDD were more likely to suffer from depression and anxiety disorders in young adulthood than those without a psychiatric disorder or those with a psychiatric disorder other than DMDD. However, no differences were found in rates of antisocial behaviours or substance abuse disorders (Copeland et al., 2014). Similarly, using data from the BCAMHS, Stringaris and Goodman (2009a) showed that the irritability dimension of ODD was predictive of depressive and GAD three years later, whereas the headstrong dimension of ODD was predictive of ADHD and CD.

Although most studies have found specific associations between irritability and internalising disorders, some of them have not replicated these results. For example, in a
large clinical sample, children aged 6-12 years who were retrospectively diagnosed with DMDD were not at increased risk of developing any new psychiatric disorder at 12- and 24-month follow-ups than those children without DMDD (Axelson et al., 2012). Using the data from the Great Smoky Mountains Study (GSMS), Rowe and colleagues (2010) showed that the irritability dimension of ODD in children aged 9-13 years was associated with ODD and anxiety at age 16; whereas the headstrong dimension was associated with depression along with CD, ODD, and substance disorder.

It is therefore unclear why most studies have found the specific association between irritability and internalising disorders, whereas others have not. First, it is possible that other outcomes than depression and anxiety have not been tested, thus decreasing the probability of finding significant associations with other disorders. Second, it is possible that there is publication bias. That is, only studies that find positive associations with depression and anxiety are published, whereas those with null results are not. Finally, the differences of findings across studies could indicate the presence of moderating factors inherent to each study, which may include the type of sample, whether it is a community-based or a clinical sample, or the age of the participants at which irritability or psychopathological outcomes were assessed. Furthermore, even if the association between irritability and future internalising disorders is consistent across studies, it is unknown how strong this association is overall, and which factors impact the strength of the association.

In this chapter, I set out to answer these questions by conducting a systematic review and meta-analysis of longitudinal studies in which chronic severe irritability is the predictor and psychopathology is the outcome.
2.1.1 Aims and hypotheses

The primary aim of the current study is to carry out a quantitative analysis of the association between chronic severe irritability, defined either as DMDD/SMD-like irritability or ODD irritable dimension, and future psychopathology, either in the form of psychiatric disorders or psychiatric symptoms. In other words, I aim to examine the strength of the association between chronic severe irritability and future psychiatric disorders and symptoms.

The secondary aims of this study include:

- To examine the extent to which associations between irritability and psychiatric outcomes other than depression and anxiety have been tested.
- To examine whether there might be publication bias in reporting these associations.
- To examine factors that might contribute to differences across studies.
- To qualitatively examine the association between irritability and future functional impairment, including suicide.

2.2 Methods

2.2.1 Data sources and search strategies

We searched PubMed and Web of Science through December 2014 (updated January 2018) using the terms (“irritability” OR “irritable” OR “disruptive mood dysregulation” OR “severe mood dysregulation” OR (“oppositional” OR “ODD”) AND (“dimension” OR “subdimension” OR “class” OR “factor” OR “subfactor’)) AND (“predict” OR “longitudinal” OR “prospective”). No limits were applied for language or date of publication. All articles obtained from the search were then manually assessed for inclusion
or exclusion, based on the presence of search terms by two independent investigators with Master-level education, who included me and another colleague. In addition, the reference list of each relevant article was reviewed for further studies, as well as papers citing these articles in Google Scholar.

2.2.2 Inclusion and exclusion criteria

We included longitudinal studies where chronic non-episodic irritability was the predictor of future psychopathology. For the purposes of this meta-analysis, chronic non-episodic irritability was defined as a dimensional measure of irritability based on criteria defining ODD symptomatology, or alternatively, as a categorical measure like DMDD or SMD. Future psychopathology was defined as any type of psychiatric disorder identified through well validated and (semi-)structured interviews, or otherwise, psychopathological scores in well-validated measures.

Prospective, longitudinal studies were included if they were published in peer-reviewed journals. No limits were applied to the age of participants at baseline or the years of follow up, nor to the type of sample (i.e., community-based or clinical). Studies using the same cohort were included as long as the baseline and follow up time points were not the same.

On the other hand, studies were excluded if the predictor was episodic irritability (e.g., in the context of depression disorder, bipolar disorder or premenstrual syndrome) and/or the outcome was other than psychopathology.

2.2.3 Study selection

Our search strategy results are shown in Figure 2.1. The search returned 230 articles after duplicates were removed, for which the titles were reviewed for relevance. Following this
review, 75 articles were excluded. After a second review, 121 further articles were excluded based on the following exclusion criteria: 12 articles were cross-sectional; 1 article was retrospective; 16 articles were focused on episodic irritability; in 19 articles, irritability was the outcome; in 35 articles, outcome was other than psychopathology; in 16 articles, the outcome was psychopathology but irritability was not the predictor; 22 articles were actually reviews. The reference sections and citations in Google Scholar of the remaining 34 studies were assessed for other potentially relevant articles, which yielded 5 additional studies to be included. Therefore, 39 studies were included in the full-text review. After this last review, 6 studies were excluded for using an idiosyncratic definition of irritability, 4 studies were excluded because irritability was not the predictor, and 1 study was excluded because it used the same cohort and time points that other study.

2.2.4 Data extraction

Study characteristics were extracted to account for heterogeneity, if any, among studies. Study information extracted included:

a) Type of measured irritability: Dimensional or categorical.

b) Definition of irritability: ODD dimension, DMDD or SMD.

c) Analysis adjustments: Prediction controlled for participant characteristics (i.e., age, sex, family SES), or/and baseline psychopathology, including symptoms comprising the headstrong dimension of ODD.

d) Age: Age at baseline and follow-up, and years between baseline and follow-up assessments. Also, age periods were defined categorically as childhood (ages up to 11 years old), adolescence (ages >11 to 17 years old), and adulthood (ages >17 years old).
Figure 2.1. Review and selection of articles for meta-analysis

167 articles identified from Medline

172 articles identified from Web of Science

230 articles after duplicates removed

Articles excluded: 196
- 75 based on review of title: 12 cross-sectional; 16 episodic irritability in the context of depression, bipolar or premenstrual syndrome; 19 irritability as outcome; 35 outcome is other than psychopathology; 16 irritability is not the predictor; 1 retrospective; 22 reviews/not articles

34 articles relevant for review

5 articles from hand-search of reference lists

39 articles included in full-text review

Articles excluded from meta-analysis and review: 6 differ in definition of irritability; 4 irritability is not the predictor; 1 study overlap in sample and time points.

28 articles included

14 articles included in meta-analysis for the prediction of psychiatric disorders

16 articles* included in meta-analysis for the prediction of psychiatric symptoms

2 articles included for narrative analyses (only provided descriptive data)

* 4 articles also predicted psychiatric disorders
e) **Sex:** Defined as proportion of males in the sample (%).

f) **Sample characteristics:** Clinical or community-based; geographic origin of the sample (i.e., US or other countries), and cohort.

### 2.2.5 Data analysis

#### 2.2.5.1 Estimation of pooled effect sizes

Given that the studies predicting psychiatric disorders and symptoms included in the meta-analysis varied in methodology and design, a random-effects model was calculated using Stata 14. We conducted the meta-analysis for findings where outcome data from two or more studies of different cohorts could be combined. We calculated pooled odds ratios (OR) and standardised effect sizes (ES), along with 95% CI, for each of the outcomes.

#### 2.2.5.2 Sensitivity analysis

Each of the studies was individually excluded from the analysis and the pooled OR/ES and 95% CI recalculated. If an individual study contributed heavily to the pooled OR/ES, a change in the magnitude or significance of the pooled OR/ES would be observed. In addition, when two or more studies from the same cohort predicted the same outcome, analyses were re-run with only one study allowed for each cohort.

#### 2.2.5.3 Heterogeneity analysis

We tested for between-study heterogeneity using the $I^2$ statistic, which is the percentage of variation attributable to heterogeneity. The values of $I^2$ lie between 0% and 100%, with larger values showing increasing heterogeneity. Higgins et al. (2003) suggest that $I^2$ values between 25%-50% are low, moderate for 50%-75%, and high for ≥75%.
In addition, to explore sources of heterogeneity and examine whether pooled estimates changed based on study features, subgroup analyses with categorical covariates and meta-regression with continuous covariates were employed.

2.2.5.4 Publication bias

Given the possibility of publication bias (i.e., significant findings are more likely to be published), we examined whether there was asymmetry in funnel plots (a scatterplot of the estimates from individual studies against a measure of a study size) and calculated the Egger’s coefficient bias. Egger’s test is a regression-based test for formally detecting skewness in the funnel plots and, therefore, publication bias in the data (Egger, Davey Smith, Schneider, & Minder, 1997). This test evaluates whether the intercept deviates significantly from zero in a regression of standardised effect estimates against their precision. It is recommended that a test for publication bias in meta-analysis should only be used when there are 10 or more studies, especially if there is substantial heterogeneity between studies (Sterne et al., 2011). This is because the power of the test is usually too low to distinguish chance from real asymmetry in funnel plots.

2.3 Results

2.3.1 Characteristics of studies for the meta-analysis

The search and review of studies resulted in 28 studies meeting inclusion criteria. Of those, 14 studies were included in the meta-analysis for the prediction of psychiatric disorders (Althoff et al., 2014; Axelson et al., 2012; Brotman et al., 2006; Copeland et al., 2014; Dougherty et al., 2015; Dougherty et al., 2016; Dougherty et al., 2013; Ezpeleta, Granero, de la Osa, Trepat, & Domenech, 2015; Kolko & Pardini, 2010; Leibenluft et al., 2006; Pickles et al., 2010; Rowe et al., 2010; Stringaris et al., 2009; Stringaris & Goodman,
2009a); 16 studies were included in the meta-analysis for the prediction of psychiatric symptoms (Barker & Salekin, 2012; Burke, 2012; Burke et al., 2010; Dougherty et al., 2015; Dougherty et al., 2016; Dougherty et al., 2013; Herzhoff & Tackett, 2016; Kolko & Pardini, 2010; Lavigne et al., 2014; Leadbeater & Homel, 2015; Mikolajewski et al., 2017; Savage et al., 2015; Stringaris et al., 2012b; Wakschlag et al., 2015; Whelan et al., 2015; Whelan et al., 2013); and 2 studies were included for narrative analysis, since these only provided descriptive data (Deveney et al., 2015; Stringaris et al., 2010a). Of note, 4 studies were included in both meta-analyses, because provided predictions for both psychiatric outcomes and symptoms (Dougherty et al., 2015; Dougherty et al., 2016; Dougherty et al., 2013; Kolko & Pardini, 2010). Finally, 9 out of the 28 studies provided information on future functional outcomes, including suicidal behaviour (Copeland et al., 2014; Dougherty et al., 2015; Dougherty et al., 2016; Dougherty et al., 2013; Ezpeleta et al., 2015; Kolko & Pardini, 2010; Pickles et al., 2010; Stringaris et al., 2009; Wakschlag et al., 2015).

In some articles, measures of estimate variance (e.g., standard errors or 95% confidence intervals) were not reported. In these cases, we contacted the authors by email; and all the authors but one provided this information. For that reason, only 15 studies predicting psychiatric symptoms could be analysed. We provide a description of the results of the study that could not be analysed (Barker & Salekin, 2012).

The overall characteristics of studies included in the meta-analyses are displayed in Table 2.1. Detailed information about the 14 studies included in the meta-analysis for the prediction of psychiatric disorders and the 16 studies included in the meta-analysis for the prediction of psychiatric symptoms are presented in Table 2.2 and Table 2.3, respectively. Table 2.3 also provides information on the 2 studies providing descriptive data.
The fourteen studies predicting psychiatric disorders were composed of 9 cohorts and comprised 8,946 unique participants. The analyses included the prediction of depression, anxiety disorder, bipolar disorder (BD), attention deficit hyperactivity disorder (ADHD), oppositional defiant disorder (ODD), conduct disorder (CD) (antisocial personality disorder in adulthood), and substance disorder (included alcohol abuse and cannabis use).

The sixteen studies predicting psychiatric symptoms were composed of 12 cohorts and comprised 16,029 unique participants. For the prediction of psychiatric symptoms, outcomes analysed included depressive symptoms, anxiety symptoms, ADHD symptoms, ODD symptoms, and CD symptoms. None study predicted manic symptoms, and only one study predicted substance abuse symptoms (Mikolajewski et al., 2017).

Of note, one study examining associations with psychiatric disorders (Stringaris & Goodman, 2009a) and four studies examining associations with psychiatric symptoms predicted internalising disorders/symptoms (Herzhoff & Tackett, 2016; Kolko & Pardini, 2010; Leadbeater & Homel, 2015; Savage et al., 2015). And two studies examining associations with psychiatric symptoms predicted externalising symptoms (Herzhoff & Tackett, 2016; Kolko & Pardini, 2010). That is, depressive and anxiety disorders/symptoms were lumped under internalising disorders/symptoms; and ADHD, ODD, and CD symptoms were lumped under externalising symptoms. In these cases, the studies were included in the meta-analysis of each relevant symptom (e.g. effect sizes of studies examining internalising symptoms were pooled with both, those examining depression and those examining anxiety); we then re-ran the analyses removing the studies using these broad definitions.
Table 2.1. Characteristics of studies included in the meta-analyses

<table>
<thead>
<tr>
<th>Studies predicting psychiatric disorders (n=14)</th>
<th>Studies predicting psychiatric symptoms (n=16)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of observations:</td>
<td></td>
</tr>
<tr>
<td>- <em>Total, n</em></td>
<td>13,178</td>
</tr>
<tr>
<td>- <em>Unique, n</em></td>
<td>8,946</td>
</tr>
<tr>
<td>Type of measured irritability</td>
<td></td>
</tr>
<tr>
<td>- <em>Dimensional, n (%)</em></td>
<td>7 (50%)</td>
</tr>
<tr>
<td>- <em>Categorical, n (%)</em></td>
<td>7 (50%)</td>
</tr>
<tr>
<td>Irritability defined as:</td>
<td></td>
</tr>
<tr>
<td>- ODD, n (%)</td>
<td>9 (64%)</td>
</tr>
<tr>
<td>- SMD, n (%)</td>
<td>2 (14%)</td>
</tr>
<tr>
<td>- DMDD, n (%)</td>
<td>3 (21%)</td>
</tr>
<tr>
<td>Analyses adjusted for:</td>
<td></td>
</tr>
<tr>
<td>- Age, n (%)</td>
<td>8 (57%)</td>
</tr>
<tr>
<td>- Sex, n (%)</td>
<td>11 (79%)</td>
</tr>
<tr>
<td>- SES, n (%)</td>
<td>6 (43%)</td>
</tr>
<tr>
<td>- Baseline psychopathology, n (%)</td>
<td>9 (64%)</td>
</tr>
<tr>
<td>- Headstrong dimension, n (%) <em>a</em></td>
<td>4 (44%)</td>
</tr>
<tr>
<td>Age:</td>
<td></td>
</tr>
<tr>
<td>- Age at baseline, years, Mean (range)</td>
<td>9.2 (3.5-14.5)</td>
</tr>
<tr>
<td>- Age at follow-up, years, Mean (range)</td>
<td>17.9 (6-44.5)</td>
</tr>
<tr>
<td>Sex:</td>
<td></td>
</tr>
<tr>
<td>- Males, Mean % (range)</td>
<td>56% (46%-84%)</td>
</tr>
<tr>
<td>Sample characteristics:</td>
<td></td>
</tr>
<tr>
<td>- Type, n (%)</td>
<td>Clinical, 2 (14%); Community, 12 (86%)</td>
</tr>
<tr>
<td>- Geographic origin, n (%)</td>
<td>US, 10 (71%); Other, 4 (29%)</td>
</tr>
<tr>
<td>Disorders/symptoms predicted:</td>
<td></td>
</tr>
<tr>
<td>- Depression, n (%)</td>
<td>11 (79%)           b</td>
</tr>
<tr>
<td>- Anxiety, n (%)</td>
<td>12 (86%)           b</td>
</tr>
<tr>
<td>- Bipolar, n (%)</td>
<td>3 (21%)</td>
</tr>
<tr>
<td>- ADHD, n (%)</td>
<td>8 (57%)</td>
</tr>
<tr>
<td>- CD, n (%)</td>
<td>9 (64%)</td>
</tr>
<tr>
<td>- ODD, n (%)</td>
<td>6 (43%)</td>
</tr>
<tr>
<td>- Drug abuse, n (%)</td>
<td>4 (29%)</td>
</tr>
</tbody>
</table>

SES, socioeconomic status; ODD, oppositional defiant disorder; SMD, severe mood dysregulation; DMDD, disruptive mood dysregulation disorder; ADHD, attention deficit hyperactivity disorder; CD, conduct disorder.

*a* In those studies where irritability is defined as ODD dimension (n=9 for psychiatric disorders; n=15 for psychiatric symptoms). *b* One study predicts internalising disorders. *c* Four studies predict internalising symptoms. *d* Two studies predict externalising symptoms.
<table>
<thead>
<tr>
<th>Study (Cohort)</th>
<th>N</th>
<th>Sex (% males)</th>
<th>Baseline age*</th>
<th>Follow-up age*</th>
<th>Baseline assessment of irritability</th>
<th>Type of Irritability</th>
<th>Outcomes</th>
<th>Follow-up assessment of psychiatric disorder</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Althoff et al., 2014) (Zuid-Holland Longitudinal Study)</td>
<td>1390</td>
<td>49%</td>
<td>3-17 years</td>
<td>18-32 years</td>
<td>Parent report on Child Behaviour Checklist</td>
<td>ODD symptoms</td>
<td>Any mood disorder</td>
<td>Self-report on the Composite International Diagnostic Interview</td>
</tr>
<tr>
<td>(Axelson et al., 2012) (Longitudinal Assessment of Manic Symptoms)</td>
<td>433</td>
<td>67%</td>
<td>6-12 years</td>
<td>8-14 years</td>
<td>Parent report on the Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version</td>
<td>DMDD</td>
<td>Bipolar spectrum Any depressive disorder Anxiety disorder Conduct disorder</td>
<td>Parent report on the Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version</td>
</tr>
<tr>
<td>(Brotman et al., 2006) (Great Smoky Mountains)</td>
<td>1,420</td>
<td>56%</td>
<td>9-13 years</td>
<td>16-22 years</td>
<td>Parent report on the Child and Adolescent Psychiatric Assessment</td>
<td>SMD</td>
<td>Any depressive Any anxiety ADHD CD ODD Substance disorder</td>
<td>Self-report on the Young Adult Psychiatric Assessment</td>
</tr>
<tr>
<td>(Copeland et al., 2014) (Great Smoky Mountains)</td>
<td>1,273</td>
<td>56%</td>
<td>9-13 years</td>
<td>24-26 years</td>
<td>Parent report on the Child and Adolescent Psychiatric Assessment</td>
<td>DMDD</td>
<td>Depressive Anxiety ASPD Alcohol</td>
<td>Self-report on the Young Adult Psychiatric Assessment</td>
</tr>
<tr>
<td>(Dougherty et al., 2013) (Community close to Stony Brook University, NY)</td>
<td>462</td>
<td>54%</td>
<td>3.6 years</td>
<td>6.1 years</td>
<td>Parent report on the Preschool Age Psychiatric Assessment</td>
<td>ODD symptoms</td>
<td>Depression Anxiety ADHD ODD</td>
<td>Parent report on the Preschool Age Psychiatric Assessment</td>
</tr>
<tr>
<td>(Dougherty et al., 2015) (Community close to Stony Brook University, NY)</td>
<td>446</td>
<td>54%</td>
<td>3.5 years</td>
<td>9.3 years</td>
<td>Parent report on the Preschool Age Psychiatric Assessment</td>
<td>ODD symptoms</td>
<td>Depression (null) Anxiety ADHD DBD</td>
<td>Parent and self-report on Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version</td>
</tr>
<tr>
<td>Study Details</td>
<td>Sample Size</td>
<td>Participation Rate</td>
<td>Age Range</td>
<td>Assessment Method</td>
<td>DSM Disorders</td>
<td>Schizophrenia and STAI</td>
<td>Note</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------------------------------------------</td>
<td>-------------</td>
<td>--------------------</td>
<td>-----------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>---------------</td>
<td>------------------------</td>
<td>----------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>(Dougherty et al., 2016) (Community close to Stony Brook University, NY)</td>
<td>473</td>
<td>55%</td>
<td>6.1 years</td>
<td>Parent report on the Preschool Age Psychiatric Assessment</td>
<td>DMDD</td>
<td>Depression Anxiety</td>
<td>ADHD DBD Parent and self-report on Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version</td>
<td></td>
</tr>
<tr>
<td>(Ezpeleta et al., 2015) (Community in Barcelona)</td>
<td>218</td>
<td>50%</td>
<td>3.8 years</td>
<td>Parent report on the Diagnostic Interview of Children and Adolescents for Parents of Preschool Children</td>
<td>ODD</td>
<td>ADHD Anxiety ODD</td>
<td>ADHD ODD Parent report on the Diagnostic Interview of Children and Adolescents for Parents of Preschool Children</td>
<td></td>
</tr>
<tr>
<td>(Kolko &amp; Pardini, 2010) (Patients from the University of Pittsburgh Medical Center)</td>
<td>177</td>
<td>84%</td>
<td>6-11 years</td>
<td>Parent and child report on the Schedule for Affective Disorders and Schizophrenia for School-Age Children</td>
<td>ODD</td>
<td>ADHD CD ODD</td>
<td>Parent report on the Schedule for Affective Disorders and Schizophrenia for School-Age Children</td>
<td></td>
</tr>
<tr>
<td>(Leibenluft et al., 2006) (Community in New York)</td>
<td>776</td>
<td>50%</td>
<td>9-19 years</td>
<td>Parent and child report on the Diagnostic Interview Schedule for Children</td>
<td>ODD</td>
<td>MDD GAD ADHD CD ODD Mania</td>
<td>Self-report on the Diagnostic Interview Schedule for Children</td>
<td></td>
</tr>
<tr>
<td>(Pickles et al., 2010) (Isle of Wight)</td>
<td>2,226</td>
<td>50%</td>
<td>14-15 years</td>
<td>Parent and child report on a structured interview</td>
<td>SMD-like</td>
<td>MDD GAD Substance disorder</td>
<td>Self-report on Schedule for Affective Disorders and Schizophrenia</td>
<td></td>
</tr>
<tr>
<td>(Rowe et al., 2010) (Great Smoky Mountains)</td>
<td>1,420</td>
<td>56%</td>
<td>9-13 years</td>
<td>Parent report on the Child and Adolescent Psychiatric Assessment</td>
<td>ODD</td>
<td>Depression anxiety CD ODD Substance disorder</td>
<td>Parent report on the Child and Adolescent Psychiatric Assessment</td>
<td></td>
</tr>
<tr>
<td>(Stringaris et al., 2009) (Community in New York)</td>
<td>631</td>
<td>46%</td>
<td>13.8 years</td>
<td>Parent and child report on the Diagnostic Interview Schedule for Children</td>
<td>ODD</td>
<td>MDD GAD Bipolar disorder</td>
<td>Self-report on the Structured Clinical Interview for DSM-IV Axis I Disorders</td>
<td></td>
</tr>
<tr>
<td>(Stringaris &amp; Goodman, 2009a) (British Child and Adolescent Psychiatric Assessment)</td>
<td>1,833</td>
<td>59%</td>
<td>9.8 years</td>
<td>Parent report on the Development and Well-Being Assessment</td>
<td>ODD</td>
<td>Internalising disorders</td>
<td>Parent report on the Development and Well-Being Assessment</td>
<td></td>
</tr>
<tr>
<td>Adolescent Mental Health Survey)</td>
<td></td>
<td></td>
<td></td>
<td>(Depressive + Anxiety) ADHD CD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>-------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDD: Major depressive disorder. GAD: Generalized anxiety disorder. CD: Conduct disorder. ADHD: Attention deficit hyperactivity disorder. ODD: Oppositional defiant disorder. ASPD: Antisocial personality disorder. SMD: Severe Mood Dysregulation. DMDD: Disruptive Mood Dysregulation Disorder. * Age: Range, or mean if range not indicated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study (Cohort)</td>
<td>N</td>
<td>Sex (% males)</td>
<td>Baseline age</td>
<td>Follow-up age</td>
<td>Baseline assessment of irritability</td>
<td>Type of Irritability</td>
<td>Outcomes</td>
<td>Follow-up assessment of psychiatric symptoms</td>
</tr>
<tr>
<td>---------------</td>
<td>---</td>
<td>---------------</td>
<td>--------------</td>
<td>--------------</td>
<td>---------------------------------</td>
<td>----------------------</td>
<td>----------</td>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>(Burke et al., 2010) (Pittsburgh Girls Study)</td>
<td>2451</td>
<td>0%</td>
<td>5-8 years</td>
<td>11.5 years</td>
<td>Self-report on Child Symptom Inventory – IV</td>
<td>ODD symptoms</td>
<td>Depressive symptoms Conduct symptoms</td>
<td>Self-report on Child Symptom Inventory – IV</td>
</tr>
<tr>
<td>(Burke, 2012) (Developmental Trends Study)</td>
<td>165</td>
<td>100%</td>
<td>7-12 years</td>
<td>17 years</td>
<td>Parent report on the Diagnostic Interview Schedule for Children</td>
<td>ODD symptoms</td>
<td>Depressive symptoms Anxiety symptoms</td>
<td>Parent report on the Diagnostic Interview Schedule for Children</td>
</tr>
<tr>
<td>(Dougherty et al., 2013) (Community close to Stony Brook University, NY)</td>
<td>462</td>
<td>54%</td>
<td>3.6 years</td>
<td>6.1 years</td>
<td>Parent report on the Preschool Age Psychiatric Assessment</td>
<td>ODD symptoms</td>
<td>Depressive Anxiety ADHD ODD</td>
<td>Parent report on the Preschool Age Psychiatric Assessment</td>
</tr>
<tr>
<td>(Dougherty et al., 2015)(Community close to Stony Brook University, NY)</td>
<td>446</td>
<td>54%</td>
<td>3.5 years</td>
<td>9.3 years</td>
<td>Parent report on the Preschool Age Psychiatric Assessment</td>
<td>ODD symptoms</td>
<td>Depressive Anxiety ADHD DBD</td>
<td>Parent and self-report on Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version, Child Depression Inventory, and Screen for Child Anxiety Related Disorders</td>
</tr>
<tr>
<td>(Dougherty et al., 2016)(Community close to Stony Brook University, NY)</td>
<td>473</td>
<td>55%</td>
<td>6.1 years</td>
<td>9.2 years</td>
<td>Parent report on the Preschool Age Psychiatric Assessment</td>
<td>DMDD</td>
<td>Depressive Anxiety ADHD DBD</td>
<td>Parent and self-report on Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version, Child Depression Inventory, and Screen for Child Anxiety Related Disorders</td>
</tr>
<tr>
<td>(Herzhoff &amp; Tackett, 2016)</td>
<td>275</td>
<td>49%</td>
<td>9-10 years</td>
<td>11-12 years</td>
<td>Parent report on Computerized Diagnostic</td>
<td>ODD symptoms</td>
<td>Internalizing symptoms</td>
<td>Parent report on Child Behavior Checklist</td>
</tr>
<tr>
<td>Studies</td>
<td>Sample Size</td>
<td>Age Range</td>
<td>Interview Schedule</td>
<td>Externalizing Symptoms</td>
<td>Parent Report</td>
<td>Self-report</td>
<td>Other Instruments</td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>-------------</td>
<td>------------</td>
<td>--------------------</td>
<td>------------------------</td>
<td>---------------</td>
<td>-------------</td>
<td>------------------</td>
<td></td>
</tr>
<tr>
<td>(Kolko &amp; Pardini, 2010) (Patients from the University of Pittsburgh Medical Center)</td>
<td>177</td>
<td>6-11 years</td>
<td>9-14 years</td>
<td>ODD symptoms dimension</td>
<td>Parent report on Child Behavior Checklist</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lavigne et al. 2014</td>
<td>796</td>
<td>4.4 years</td>
<td>6.5 years</td>
<td>ODD symptoms</td>
<td>Depressive symptoms Anxiety symptoms</td>
<td>Parent report on DISC and Child Symptom Inventory</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Leadbeater &amp; Homel, 2015) (Victoria Healthy Youth Survey)</td>
<td>464</td>
<td>22-23 years</td>
<td>24-25 years</td>
<td>ODD symptoms</td>
<td>Internalizing symptoms Conduct symptoms</td>
<td>Self-report on Brief Child and Family Phone Interview</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Mikolajewski et al 2017) (The Minnesota Twin Family Study)</td>
<td>2230</td>
<td>11.78</td>
<td>17</td>
<td>ODD symptoms</td>
<td>Depressive Anxiety Substance use Adult antisocial behaviour</td>
<td>Self-report on Structured Clinical Interview for DSM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Savage et al., 2015) (Swedish Twin Study of Child and Adolescent Development)</td>
<td>576</td>
<td>16-17 years</td>
<td>19-20 years</td>
<td>ODD symptoms</td>
<td>Anxious/depressed symptoms</td>
<td>Self-report on Adult Behavior Checklist</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Stringaris et al., 2012b) (G1219)</td>
<td>1597</td>
<td>12-12 years</td>
<td>14-23 years</td>
<td>ODD symptoms</td>
<td>Depressive symptoms Delinquent behaviour</td>
<td>Depressive symptoms – Self-report on Short Mood and Feelings Questionnaire Delinquent behaviour- Self-report on ASEBA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Wakschlag et al 2015) (Multidimensional Assessment of Preschoolers Study)</td>
<td>497</td>
<td>2.9-6 years</td>
<td>3.8-8.5 years</td>
<td>ODD symptoms</td>
<td>Depressive symptoms GAD symptoms ADHD symptoms ODD symptoms CD symptoms</td>
<td>Parent report on the Preschool Age Psychiatric Assessment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Sample Size</td>
<td>Prevalence</td>
<td>Age at First Assessment</td>
<td>Age at Last Assessment</td>
<td>Research Instruments</td>
<td>ODD Symptoms</td>
<td>Depressive Symptoms</td>
<td>Conduct Symptoms</td>
</tr>
<tr>
<td>-----------------------------------------</td>
<td>-------------</td>
<td>------------</td>
<td>--------------------------</td>
<td>------------------------</td>
<td>----------------------</td>
<td>--------------</td>
<td>---------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>(Whelan et al., 2013) (Avon Longitudinal Study of Parents and Children)</td>
<td>6328</td>
<td>51%</td>
<td>13 years</td>
<td>16 years</td>
<td>Parent report on the Development and Well Being Assessment</td>
<td>ODD symptoms</td>
<td>Depressive symptoms</td>
<td>Conduct symptoms</td>
</tr>
<tr>
<td>(Whelan et al., 2015) (Avon Longitudinal Study of Parents and Children)</td>
<td>3963</td>
<td>41%</td>
<td>8 years</td>
<td>13 years</td>
<td>Parent report on the Development and Well Being Assessment</td>
<td>ODD symptoms</td>
<td>Depressive symptoms</td>
<td>Conduct symptoms</td>
</tr>
<tr>
<td>(Stringaris et al., 2010a)</td>
<td>84</td>
<td>67%</td>
<td>11.6 years</td>
<td>13.6 years</td>
<td>Parent and child report on The Kiddie Schedule for Affective Disorders—Present and Lifetime Version</td>
<td>SMD</td>
<td>Manic episodes</td>
<td>Depressive episodes</td>
</tr>
<tr>
<td>(Deveney et al., 2015)</td>
<td>200</td>
<td></td>
<td>7-17.6 years</td>
<td>11-21 years</td>
<td>Parent and child report on The Kiddie Schedule for Affective Disorders—Present and Lifetime Version</td>
<td>SMD</td>
<td>Depression disorder</td>
<td>Anxiety disorder</td>
</tr>
</tbody>
</table>

ODD: Oppositional defiant disorder. SMD: Severe Mood Dysregulation. ADHD: Attention deficit hyperactivity disorder. DBD: Disruptive behaviour disorder. GAD: Generalized anxiety disorder. CD: Conduct disorder. *Age: Range, or mean if range not indicated
Below I present the results of the meta-analysis for the prediction of psychiatric disorders and psychiatric symptoms separately. Then, I discuss the studies that only report descriptive data. Finally, I report the results from studies predicting functional impairment.

2.3.2 Irritability as a predictor of psychiatric disorders

Figure 2.2 shows effect sizes and the corresponding 95% confidence intervals for each study in the prediction of future psychiatric disorders from irritability. Results are presented for each psychiatric disorder separately, along with sensitivity analyses, in which each individual study and studies from the same cohort are removed.

2.3.2.1 Depressive Disorder

Eleven studies representing seven cohorts included depressive disorder as an outcome, with one of those including internalising disorders instead. Nine studies (82%) reported significant findings, with irritability predicting depression at follow up (OR=1.85, 95% CI=1.45-2.37, \( p < 0.001 \)). The overall variance of these results was moderate (\( I^2 = 58.4\% \)).

After removing the study predicting internalising disorders (Stringaris & Goodman, 2009a), pooled OR was 1.90 (95% CI, 1.43-2.52, \( p < 0.001 \)), with a moderate heterogeneity (\( I^2 = 62.1\% \)).

The exclusion of each individual study did not alter the results; that is, pooled OR ranged 1.71-1.99, being all significant (all \( p < 0.05 \)). In addition, recalculated pooled odds ratio after removing studies from the same cohort ranged from OR= 1.61, 95% CI 1.35-1.93 to OR=1.95, 95% CI 1.45-2.61 (all \( p < 0.001 \)), with low between-study heterogeneity ranging \( I^2 = 16\% - 50.9\% \).
Figure 2.2. Forest plot of irritability as a predictor of future psychiatric disorders. Points represent the estimated odds ratio of each study; the lines bisecting the point correspond to the 95% confidence intervals (CI). Pooled effect sizes are represented by diamonds. Weights for each study are given in the right column. ADHD: Attention deficit hyperactivity disorder. ODD: Oppositional defiant disorder.
2.3.2.2 Anxiety Disorders

Twelve studies representing seven cohorts included anxiety disorder as an outcome, with one of those including internalising disorders instead. Irritability was a significant predictor of anxiety at follow up (OR=1.58, 95% CI=1.26-1.99, \( p<0.001 \)). The overall variance of these results was also moderate (\( I^2=54.7\% \)). After removing the study predicting internalising disorders (Stringaris & Goodman, 2009a), pooled OR was 1.56 (95% CI, 1.21-2.02, \( p=0.001 \)), with a moderate heterogeneity (\( I^2=57.1\% \)).

As in the case of depression, the exclusion of each individual study predicting anxiety disorders did not alter the results; that is, pooled OR ranged 1.42-1.66, being all significant. Meta-analyses performed after removing studies from the same cohort revealed that the whole heterogeneity was explained by the prediction of Copeland et al (2014). When this study was included, pooled OR ranged from 1.79, 95% CI 1.25-2.56, \( p=0.001 \) to 1.82, 95% CI 1.20-2.76, \( p=0.005 \), with an \( I^2 \) ranging 61% and 67.7%. However, when this study was not included, OR ranged from 1.42, 95% CI 1.21-1.67 to 1.60, 95% CI 1.26-2.02 (all \( p\leq0.001 \)) with an \( I^2 \) of 0% in all cases.

2.3.2.3 Bipolar Disorder

Three studies representing two cohorts included BD as outcome. No study reported significant findings. When all these studies were considered together, chronic irritability was not a significant predictor of BD at follow up (OR=1.09, 95% CI=0.67-1.77, \( p=0.739 \)). The overall variance of these results was low (\( I^2=0\% \)).
Removing individual studies that predicted bipolar disorder did not alter the results; in all cases pooled OR remained not significant (range OR=0.83, 95%CI 0.38-1.80 to OR=1.67, 95%CI 0.49-5.67). Removing studies from the same cohort did not change the results either (all \(p>0.9\); \(I^2=0\%\)).

### 2.3.2.4 Attention Deficit Hyperactivity Disorder

Eight studies representing six different cohorts included ADHD as outcome. Of those, only two studies reported significant findings (Dougherty et al., 2016; Kolko & Pardini, 2010). However, when all the studies were considered together, irritability was a significant predictor of ADHD at follow up (OR=1.29, 95% CI=1.07-1.55, \(p=0.007\)). The overall variance of these results was low (\(I^2=0\%\)).

The exclusion of individual studies predicting ADHD (n=8) revealed that the significant results were driven by a single study. Specifically, when the study from Dougherty et al (2015) was removed from the analysis, the pooled OR became non-significant (OR=1.36, 95%CI 0.99-1.86). As mentioned before, only 2 studies predicting ADHD found significant associations, and the study of Dougherty et al (2015) was not one of them. In this study, the estimate for the association between irritability and future ADHD was OR=1.26, 95%CI 0.99-1.61. However, the percentage of weight attributed to this study in the random effects meta-analysis (57.6%) probably contributed to the significant results.

Similarly, when including only one study from the same cohort, the inclusion of Dougherty et al. (2013) yielded non-significant results (pooled OR=1.25, 95%CI 0.93-1.68, \(p=0.139\)). Including the other two studies from this same cohort one at a time (Dougherty et al., 2015;
Dougherty et al., 2016) resulted in significant pooled odds ratio (range OR=1.31, 95% CI 1.08-1.60 to OR=1.54, 95% CI 1.11-2.15, both $p<0.05$) with a $I^2=0\%$.

### 2.3.2.5 Conduct Disorder

Nine studies representing six cohorts included CD as an outcome. Of those, no study reported significant findings. When all these studies were considered together, irritability was not a significant predictor of CD at follow up (OR=1.11, 95% CI=0.92-1.34, $p=0.269$). The overall variance of these results was low ($I^2=2.6\%$). Excluding individual studies did not alter the main results; that is, in all cases, the pooled odds ratio remained not significant (range=1.01-1.16, all $p>0.05$). When removing studies from the same cohort results ranged from OR=0.89, 95% CI 0.60-1.32, $p=0.556$, $I^2=14\%$ to OR=1.05, 95% CI 0.81-1.36, $p=0.714$, $I^2=29\%$.

### 2.3.2.6 Oppositional Defiant Disorder

Six studies from six different cohorts included ODD as an outcome. When all the studies were considered together, irritability was a significant predictor of ODD at follow up (OR=2.62, 95% CI=1.41-4.85, $p=0.002$). However, the overall variance of these results was high ($I^2=83.2\%$). Excluding individual studies did not alter the main results; that is, in all cases pooled odds ratio remained significant (range=1.90-3.53, all $p<0.05$).

When removing the study of Brotman et al. (2006) from the Great Smoky Mountains Study cohort the overall effect size was OR=2.59, 95% CI=1.37-4.90, $p=0.003$, $I^2=86.5\%$, whereas when removing the study of Rowe et al. (2010) using the same cohort the pooled effect size was OR=3.53, 95% CI=1.33-9.37, $p=0.011$, $I^2=85.2\%$. 

130
The high between-study heterogeneity in the prediction of ODD was totally explained by two studies. The first of these studies was Ezpeleta et al. (2015) who compared children with a trajectory of high persistent irritability to children without irritability. Given that the sample size of the first group was small (n=23) and the prevalence of ODD at follow up was much higher in comparison to children without irritability (58.4% vs 3.36%), this resulted in an outlier value in the prediction of ODD (OR=81.88, 95% CI=14.68-456.69, p<0.05). In addition, this study did not control for baseline rates of ODD and the followed-up sample was enriched with cases of ODD. The second study contributing to heterogeneity was the single clinical study (Kolko & Pardini, 2010), which consisted of a clinical trial of patients with either CD or ODD. Removing these two studies from the analysis resulted in a heterogeneity of \(I^2=0\%\) with a pooled OR of 1.57, 95% CI=1.28-1.92, \(p<0.001\).

### 2.3.2.7 Substance Abuse/Dependence

Four studies representing two cohorts included substance disorder as an outcome. Of those, no study reported significant findings. When all the studies were considered together, irritability was not a significant predictor of drug abuse at follow up (OR=1.11, 95% CI=0.74-1.65, \(p=0.613\)). The overall variance of these results was low (\(I^2=27.2\%\)).

All results after removing individual studies remained non-significant (range OR=0.92-1.53, all \(p>0.05\)) When removing studies from the same cohort results ranged from OR=0.21, 95% CI 0.59-2.47, \(p=0.608\), \(I^2=71.6\%\) to OR=1.71, 95% CI 0.95-3.09, \(p=0.076\), \(I^2=0\%).
2.3.2.8 Covariates analysis in the prediction of psychiatric disorders

To explore whether the covariates could account for some of the variability among the effect sizes for the prediction of depression, anxiety, ADHD and ODD, we conducted subgroup analyses where separate pooled OR and \( p \)-values were calculated for each level of the categorical covariates. Meta-regressions were used to explore continuous covariates (i.e., mean age at baseline, mean age at follow-up, the difference between these two, and percentage of males). Given the small number of studies predicting each outcome, and hence in each category of the covariate in question, these results are intended to be hypothesis-generating as opposed to identifying conclusive reasons for heterogeneity among studies. We only describe results with at least two studies per covariate category.

For the prediction of depression, no significant differences in pooled OR were found for any covariate between sub-groups, or in relation to continuous covariates. However, the pooled ORs of those studies that adjusted for headstrong symptoms (\( n=2 \)) (OR=1.42, 95% CI 0.97-2.07, \( p=0.068 \)) (Rowe et al., 2010; Stringaris & Goodman, 2009a) as well as those studies that did not adjust for sex (\( n=2 \)) (OR=2.90, 95% CI 0.53-15.96, \( p=0.222 \)) (Axelson et al., 2012; Copeland et al., 2014) were not significant.

For the prediction of anxiety, no significant differences in pooled OR were found for any covariate between sub-groups. However, in contrast to studies that employed a dimensional measure of irritability, the pooled OR of those studies that employed categorical definitions of irritability was not significant (\( n=6 \), OR=1.78, 95% CI 0.79-4.00, \( p=0.166, I^2=72.2\% \)). Similarly, pooled odds ratio of those studies in which irritability was defined as DMDD or SMD (\( n=5 \)) were not significant (OR=1.81, 95% CI 0.70-4.69, \( p=0.222, I^2=77.3\% \)), as
opposed to pooled estimates of those studies using ODD criteria to define chronic irritability (OR= 1.44, 95% CI 1.26-1.64, p<0.001, I²=0%). In addition, non-significant pooled odds ratio emerged for those studies that did not adjust for baseline disorder (n=4, OR=2.39, 95% CI 0.85-6.69, p=0.098), age (n=5, OR=2.10, 95% CI 0.82-5.36, p=0.122), or sex (n=3, OR=2.95, 95% CI 0.72-11.98, p=0.131). No relations were found with continuous covariates.

For the prediction of ADHD, no significant differences in pooled OR were found for any covariate between sub-groups, or in relation to continuous covariates. However, the pooled odds ratios of those studies that did not adjust for age (n=2, OR=2.14, 95% CI 0.42-10.87, p=0.358), or socioeconomic status (n=3, OR=1.19, 95% CI 0.74-1.91, p=0.464) were not significant. In addition, only the pooled estimate from those studies that used ODD criteria to define chronic irritability was significant (n=6, OR=1.26, 95% CI 1.04-1.52, p=0.017), in contrast to those using DMDD/SMD criteria.

Finally, for the prediction of ODD, as in the other cases, no differences in OR were found for any covariate between sub-groups. However, the pooled OR of those studies that employed categorical definitions of irritability (n=3) (6.64, 95% CI 0.29-152.47, p=0.236) was not significant. No associations were found with continuous covariates.

2.3.2.9 Test of publication bias for the prediction of psychiatric disorders

Test of publication bias was only examined for the prediction of depression and anxiety, as these were the only outcomes with 10 studies, which is the minimum number of studies
recommended for a test of publication bias (Sterne et al., 2011). A detailed description of the examination of publication bias for depression and anxiety disorders is provided below.

**Figure 2.3a** shows the funnel plot of the 11 studies predicting depression. Three studies in the lower right of the funnel, Brotman et al (2006), Copeland et al (2014), and Dougherty et al (2016), contributed to asymmetry. The Egger’s bias coefficient also suggested the presence of asymmetry and publication bias (bias=2.48, p=0.001). However, since most of the individual effect estimates were above zero, the effect of publication bias, if any, would be to inflate the estimate rather than to lead to an incorrect conclusion about the existence of an effect. Moreover, when removing these three studies from the analyses, not only did the pooled OR remained significant and heterogeneity decreased (OR=1.61, 95% CI 1.35-1.93, p<0.001, $I^2=34.6\%$), but the Egger’s bias coefficient was no longer significant (bias=2.21, p=0.067). As can be seen in **Figure 2.3b**, the funnel plot after excluding these three studies showed more symmetry. These three studies provided larger odds ratio and larger variance that the remaining studies. This could be explained because these were the only studies that captured irritability categorically based on SMD or DMDD criteria in community samples. Therefore, the prevalence of positive cases was very low (i.e., 3.3% in Brotman et al (2006), 4.1% in Copeland et al (2014), and 7.7% in Dougherty et al (2016)) compared with the negative cases (>90%), then leading to large OR and variances.

In the prediction of anxiety, only the study from Copeland et al (2014) was in the mid-right region outside the funnel (**Figure 2.4**). Overall, the funnel plot shows symmetry across studies, and the Egger’s bias coefficient also suggested the absence of asymmetry and publication bias (bias=0.63, p=0.449).
Figure 2.3. Funnel plot with pseudo 95% confidence limits, using data from 11 studies predicting depression (A) and 8 of these studies after excluding outlier estimates (B).

Figure 2.4. Funnel plot with pseudo 95% confidence limits, using data from 12 studies predicting anxiety disorder.
Tests of publication bias in the prediction of the remaining outcomes – although these included less than 10 studies – yielded non-significant results.

2.3.3 Irritability as a predictor of psychiatric symptoms

Figure 2.5 shows effect sizes and the corresponding 95% confidence intervals for each study in the prediction of future psychiatric symptoms from irritability. Results are presented for each type of psychiatric symptom separately, along with sensitivity analyses, in which each individual study and studies from the same cohort are removed.

2.3.3.1 Depressive Symptoms

Fifteen studies representing twelve cohorts included depressive symptoms as outcome, with four of those studies including internalising symptoms as outcome, instead. Thirteen studies (87%) reported significant findings, with irritability predicting depressive symptoms at follow up (ES=0.17, 95% CI 0.11-0.23, p<0.001). The overall variance of these results was high ($I^2$=91.5%). After removing studies predicting internalising symptoms, pooled ES was 0.16 (95% CI, 0.09-0.23, p<0.001), with high heterogeneity ($I^2$=93%).

The pooled ES of studies predicting internalising symptoms was also significant (ES=0.27, 95% CI 0.08-0.47, p=0.006, $I^2$=85.3%).

The exclusion of each individual study did not alter the results; that is, pooled ES ranged 0.14-0.19, being all significant (all p<0.05). In addition, the recalculated pooled ES after removing studies from the same cohort ranged from ES= 0.17, 95% CI 0.10-0.24 to ES=0.19, 95% CI 0.12-0.26 (all p<0.001), with high between-study heterogeneity ranging $I^2$=92.8%-93.1%.
Figure 2.5. Forest plot of irritability as a predictor of future psychiatric symptoms. Points represent the estimated effect size of each study; the lines bisecting the point correspond to the 95% confidence intervals (CI). Pooled effect sizes are represented by diamonds. Weights for each study are given in the right column. ADHD: Attention deficit hyperactivity disorder. CD: Conduct disorder. ODD: Oppositional defiant disorder.
2.3.3.2 Anxiety Symptoms

Eleven studies representing nine cohorts included anxiety symptoms as outcome, with four of those studies including internalising symptoms as outcome instead. Irritability was a significant predictor of anxiety symptoms at follow up (ES=0.09, 95% CI=0.03-0.14, \(p=0.002\)). The overall variance of these results was moderate-high (\(I^2=73.5\%\)). After removing studies predicting internalising symptoms, the pooled ES was 0.05 (95% CI, -0.001-0.10, \(p=0.057\)), showing a non-significant trend, with a moderate heterogeneity (\(I^2=56.5\%\)).

As in the case of depression, the exclusion of each individual study predicting anxiety symptoms did not alter the results; that is, the pooled ES ranged 0.07-0.10, being all significant (\(p<0.05\)). Estimated pooled ES of meta-analyses performed after removing studies from the same cohort ranged from ES= 0.10, 95%CI, 0.04-0.17 to ES=0.09, 95%CI, 0.03-0.15 (all \(p<0.01\)), with high between-study heterogeneity ranging \(I^2=77.2\%-77.8\%\).

2.3.3.3 Attention Deficit Hyperactivity Disorder Symptoms

Six studies representing four different cohorts included ADHD symptom outcome, with two of those studies including externalising symptoms as outcome, instead. Only two studies reported significant findings (Dougherty et al., 2016; Wakschlag et al., 2015). However, when all studies were considered together, irritability was a significant predictor of ADHD symptoms at follow up (ES=0.07, 95% CI, 0.01-0.13, \(p=0.033\)). The overall variance of these results was low (\(I^2=41\%\)). When excluding studies that predicted externalising symptoms, results were not significant (ES=0.07, 95% CI, -0.01-0.14, \(p<0.05\)).
Pooled ES estimates that included only studies predicting externalising symptoms (n=2) were not significant either (ES=0.09, 95% CI, -0.09-0.26, \( p=0.326 \)). The exclusion of individual studies predicting ADHD revealed that the significance of pooled ES estimates disappeared when three studies, one at a time, were excluded (Dougherty et al., 2016; Kolko & Pardini, 2010; Wakschlag et al., 2015). Similarly, when excluding studies from the same cohort, the exclusion of Dougherty et al (2015) and Dougherty et al (2016) yielded non-significant results.

2.3.3.4 **Conduct Disorder Symptoms**

Nine studies representing eight cohorts included CD symptoms as outcome, with two of those studies including externalising symptoms as outcome, instead. Four studies reported significant findings. When all these studies were considered together, irritability was a significant predictor of CD symptoms at follow up (ES=0.09, 95% CI, 0.04-0.14, \( p<0.001 \)). The overall variance of these results was moderate (\( I^2=58.4\% \)). Pooled ES were still significant after removing studies that predicted externalising symptoms (ES=0.09, 95% CI, 0.04-0.15, \( p=0.001 \) \( I^2=66.9\% \)). Excluding individual studies did not alter the main results; that is, in all cases pooled ES remained significant (range=0.08-0.10, all \( p>0.05 \)). When removing studies from the same cohort results ranged from ES=0.08, 95% CI 0.03-0.12, \( p=0.001 \), \( I^2=43.6\% \) to ES=0.08, 95% CI 0.03-0.13, \( p=0.003 \), \( I^2=56.1\% \).

2.3.3.5 **Oppositional Defiant Disorder Symptoms**

Four studies from four different cohorts included ODD symptoms as outcome, with two of these studies predicting externalising symptoms. When all the studies were considered
together, irritability was a significant predictor of ODD symptoms at follow up (ES=0.11, 95% CI, 0.04-0.147, \( p=0.003 \)). The overall variance of these results was low (\( I^2=0\% \)). Excluding the two studies with externalising symptoms also yielded significant results (ES=0.11, 95% CI, 0.03-0.18, \( p=0.005 \)). Excluding individual studies did not alter the main results; that is, in all cases pooled ES remained significant (range=0.10-0.11, all \( p<0.05 \)). Finally, the authors of one study (Barker & Salekin, 2012) were contacted for further information in order to pool their data with that from the other studies but did not reply; however the prediction of internalising disorders in this study (Barker & Salekin, 2012) was significant (ES=0.21, \( p<0.05 \)).

### 2.3.3.6 Covariates analysis in the prediction of psychiatric symptoms

To explore whether the covariates could account for some of the variability among the effect sizes for the prediction of psychiatric symptoms, we conducted subgroup analyses where separate pooled ES and \( p \)-values were calculated for each level of the categorical covariates. Meta-regressions were used to explore continuous covariates (i.e., mean age at baseline, mean age at follow-up, the difference between these two, and percentage of males). Given the small number of studies predicting each outcome, and hence in each category of the covariate in question, these results are intended to be hypothesis-generating as opposed to identifying conclusively reasons for heterogeneity among studies. We only describe results with at least two studies per covariate category.

For the prediction of depressive symptoms, no significant differences in pooled ES were found for any covariate between sub-groups, or in relation to continuous covariates.
For the prediction of anxiety symptoms, no significant differences in pooled ES were found for any covariate between sub-groups. However, pooled estimates for those studies that adjusted for sex and socioeconomic status were not significant (ES=0.10, 95% CI, -0.001-0.19, p=0.053). By contrast, only pooled estimates of those studies that adjusted for baseline symptoms were significant (ES=0.08, 95% CI, 0.02-0.14, p=0.013). In addition, only pooled estimates from community-based samples were significant (ES=0.06, 95% CI, 0.01-0.12, p=0.021). No associations were found with continuous covariates.

For the prediction of ADHD symptoms, no significant differences in pooled ES were found for any covariate between sub-groups, or in relation to continuous covariates. However, when analyses were split by the adjustment of baseline symptoms, the adjustment of headstrong symptoms or the type of sample, no pooled estimate was significant in any of the sub-groups (all p>0.05).

For the prediction of CD symptoms, no significant differences in pooled ES were found for any covariate between sub-groups, or in relation to continuous covariates. However, only pooled estimates of studies that adjusted for age (ES=0.12, 95% CI, 0.06-0.19 p<0.001) and those that did not adjust for headstrong symptoms (ES=0.08, 95% CI, 0.06-0.10, p<0.001) were significant. In contrast, pooled estimates of those studies that adjusted for socioeconomic status were not significant (ES=0.11, 95% CI, -0.001-0.22, p=0.053). Finally, only pooled estimates from US studies were significant (ES=0.11, 95% CI, 0.07-0.16, p<0.001).
Finally, for the prediction of **ODD symptoms**, as in the other cases, no differences in ES were found for any covariate between sub-groups. However, only the pooled estimates of those studies that adjusted for baseline symptoms (ES=0.11, 95% CI, 0.02-0.20, \( p=0.016 \)), and those that did not adjust for headstrong symptoms (ES=0.11, 95% CI, 0.03-0.18, \( p=0.005 \)) were significant.

### 2.3.3.7 Test of publication bias for the prediction of psychiatric symptoms

As in the case of psychiatric disorders, tests of publication bias was only examined for the prediction of depressive and anxiety symptoms, as these were the only outcomes with at least 10 studies, which is the minimum number of studies recommended for a test of publication bias (Sterne et al., 2011). A detailed description of the examination of publication bias for depressive and anxiety symptoms is provided below.

**Figure 2.6** shows the funnel plot of the 14 studies predicting **depressive symptoms**. The Egger’s bias coefficient suggested the presence of asymmetry and publication bias (bias=3.14, \( p=0.005 \)). Even after removing the most far-right outliers in **Figure 2.6** (Herzhoff & Tackett, 2016; Kolko & Pardini, 2010; Lavigne et al., 2014), there was still evidence of asymmetry and bias (bias=2.30, \( p<0.001 \)). In the prediction of **anxiety symptoms**, two studies were located outside the funnel plot seen in **Figure 2.7** (Herzhoff & Tackett, 2016; Kolko & Pardini, 2010). Of note, these two studies predicted internalising symptoms and not just anxiety symptoms. In any case, Egger’s bias coefficient suggested the absence of asymmetry and publication bias (bias=1.55, \( p=0.211 \)). And when these two outliers were removed, Egger’s bias coefficient became -0.69, \( p=0.585 \).
Figure 2.6. Funnel plot with pseudo 95% confidence limits, using data from 14 studies predicting depressive symptoms.

Figure 2.7. Funnel plot with pseudo 95% confidence limits, using data from 11 studies predicting anxiety symptoms.
Test of publication bias in the prediction of the remaining outcomes – although these included less than 10 studies – yielded non-significant results.

2.3.4 Descriptive studies

As mentioned before, two studies could not be included in the meta-analysis because they provided only descriptive data. The summary of the results of these studies is described below.

In one study, two-hundred youth meeting criteria for SMD were followed up for 2 (n=84) and 4 years (n=46) (Deveney et al., 2015). The authors found that rates of depression and anxiety increased at 2 and 4 years follow up, whereas rates of ODD and ADHD were high but unchanged. In addition, no participant met criteria for a hypomanic/manic episode at the 2-year follow-up, and a single person (2.2%) met criteria for BD (type II) by the 4-year follow-up.

A second study compared the rates of onset of bipolar disorder in youth with chronic vs episodic irritability (Stringaris et al., 2010a). In a referred sample, 84 youths with SMD and 93 youths with DSM-IV bipolar disorder were followed over a median of 28.4 months. At follow up only one patient (1.2%) with SMD exhibited one or more manic, hypomanic, or mixed episode, compared to 58 patients (62.4%) with bipolar disorder.

2.3.5 Irritability as a predictor of functional impairment

Nine studies provided information on future functional impairment associated with irritability. Chronic severe irritability in children and adolescents is significantly associated with worse health outcomes, poorer social functioning, more risky/illegal behaviours
(Copeland et al., 2014), and lower financial and educational attainment in adulthood (Copeland et al., 2014; Stringaris et al., 2009). Similarly, chronic severe irritability in children is longitudinally associated with increased use of educational support services, outpatient treatment and psychotropic medication, as well as poorer social functioning at school (Dougherty et al., 2016). Several studies have found significant associations between chronic irritability in children and future functional impairment assessed with dimensional measures, even after controlling for baseline psychiatric disorders and symptoms (Dougherty et al., 2015; Dougherty et al., 2013; Ezpeleta et al., 2015; Kolko & Pardini, 2010; Wakschlag et al., 2015). Finally, a large population-based study found a strong association between chronic irritability in adolescence and suicidal behaviours in adulthood, including plans and suicide attempts, independent of affective disorders (Pickles et al., 2010).

Although criminal behaviours are not strictly speaking psychiatric disorders, four studies examined how this was predicted by irritability (Aebi et al., 2013; Althoff et al., 2014; Copeland et al., 2014; Stringaris et al., 2012b). Only the study with categorically defined irritability (i.e., DMDD) found a significant association (Copeland et al., 2014). By contrast, studies examining ODD dimensions found that the headstrong/hurtful dimension of ODD, but not irritability, predicted criminal offences (Aebi et al., 2013; Althoff et al., 2014; Copeland et al., 2014; Stringaris et al., 2012b).
2.4 Discussion

2.4.1 Summary of findings

Chronic, severe irritability appears to be associated with future depressive and anxiety disorders. Evidence also suggests some associations between chronic, severe irritability and future ODD. Conversely, the evidence for the association with BD, ADHD, CD and substance abuse, is either weak (in the case of ADHD) or non-existent. In addition, irritability consistently predicted not only depressive and anxiety symptoms, but also CD and ODD symptoms. The evidence for the association between irritability and ADHD symptoms was weak. Finally, irritability appears to be associated with future impairment in different areas, even after adjusting for baseline psychopathology. These findings should be interpreted with caution given the number of studies analysed and their methodological shortcomings.

2.4.2 Association between irritability and future psychiatric outcomes

In terms of psychiatric disorders, chronic severe irritability seems to be specifically associated with depression (OR=1.85, p<0.001) and anxiety (OR=1.58, p<0.001) as opposed to BD, ADHD, CD and substance abuse. Some evidence also points to associations with future ODD (OR=2.62, p=0.002), but the strength of this association decreased when removing outliers and clinical samples of youth with ODD (OR=1.57, p<0.001) (Ezpeleta et al., 2015; Kolko & Pardini, 2010). Nevertheless, it is unclear whether the association between chronic irritability and future ODD is robust or artificial due to using ODD items to measure irritability (Dougherty et al., 2013; Rowe et al., 2010).
Unlike its associations with psychiatric disorders, the associations between irritability and future psychiatric symptoms are not specific. That is, chronic severe irritability seems to be associated with both internalising and externalising symptoms (with the exclusion of ADHD symptoms). It is possible that depressive and anxiety disorders that develop following chronic severe irritability, are presented with more behavioural symptoms without reaching the threshold for a diagnosis. Indeed, the presence of irritability in youth with depression doubles the rate of comorbid behavioural disorders such as ODD or CD, compared with youth with depression without irritability (Stringaris et al., 2013). Future studies should compare symptom profiles between depression and anxiety disorders that are or not preceded by chronic irritability.

### 2.4.3 Variety of psychiatric outcomes tested

The robust association between irritability and future internalising disorders is supported by the close number of comparisons made with externalising disorders. In other words, the association between irritability and internalising disorders have been tested almost as many times as the association between irritability and externalising disorders. Therefore, the probability of finding an association by chance should be nearly the same in both cases. Specifically, if we put together depression, anxiety and internalising outcomes on one side, and ODD, ADHD, CD and externalising outcomes on another, then the association between irritability and internalising disorders have been tested 22 times, compared with 17 times in the case of externalising disorders. In comparison, the associations between irritability and the development of BD have been rarely tested quantitatively (n=3), although two further studies have given descriptive rate estimates (Deveney et al., 2015; Stringaris et al., 2010a).
Moreover, BD takes an episodic form and may be missed in follow-up assessments. Although this is also true for depression, BD in young people is rare compared with rates of depression, which might give rise to potential Type II errors.

In the context of symptoms, the association between irritability and internalising symptoms has been tested nearly twice (n=23) as much as the association between irritability and externalising symptoms (n=12). This probably occurred because most of the studies testing associations with internalising symptoms employed hypothesis-driven analyses aimed at understanding further the association between irritability and internalising outcomes (Burke, 2012; Lavigne et al., 2014; Savage et al., 2015; Stringaris et al., 2012b). Further studies are needed that test the associations between irritability and externalising symptoms. Moreover, studies testing associations with both psychiatric disorders and psychiatric symptoms will help us to examine whether irritable youth who develop depressive and anxiety disorders are at higher risk of externalising symptoms without meeting criteria for externalising disorders.

2.4.4 Evidence for publication bias

For the prediction of depressive disorder, there was a suggestion of bias toward larger effect sizes driven by three studies (Brotman et al., 2006; Copeland et al., 2014; Dougherty et al., 2016). However, when these outliers were removed, results for the association between irritability and depressive disorder were still significant with no evidence of publication bias. No evidence for publication bias was seen for the significant association between irritability and anxiety disorders.
For the association between irritability and depressive or internalising symptoms, there was evidence for publication bias. This is not surprising given that 87% of studies (13/15) reported a significant association. No publication bias was evident for the association with anxiety.

The results from tests of publication bias should be treated with caution due to the small number of individual studies included in this meta-analysis. There were only 11 studies predicting depressive disorder, and eight when excluding the outliers, and the pooled OR had moderate heterogeneity. Fifteen studies predicted depressive or internalising symptoms and the pooled ES had high heterogeneity (>90%). Therefore, more studies are needed to conclude whether there is publication bias or not. In addition, publication bias is only one of many sources of asymmetry found in funnel plots (Egger et al., 1997; Sterne et al., 2011).

2.4.5 Factors contributing to differences between studies

As mentioned earlier, the results from subgroup analyses with categorical features (e.g., community sample vs. clinical sample) and meta-regression with continuous features (e.g., the proportion of males in the sample) should be taken with caution given the small number of studies predicting each outcome. However, it should be noted that there were no significant differences between subgroups in any of the analyses, and no significant associations were found with continuous features. However, non-significant differences (i.e., overlapping 95%CI) were found between subgroups. For example, for the prediction of depressive disorder, the pooled OR of those studies that did not adjust analyses for sex or headstrong ODD dimension were not significant. In the prediction of ODD, the pooled OR
of those studies using a categorical definition of irritability like SMD or DMDD was not
significant (Brotman et al., 2006; Leibenluft et al., 2006). Although the current meta-
analysis is not sufficiently powered to test these differences, these might be suggestive of
sources of heterogeneity and disparity between studies. For example, in the case of ODD,
only studies using a dimensional ODD dimension found significant associations with ODD,
which might support the artificial overlap due to shared item variance. Further studies are
needed to extract conclusions from these analyses.

2.4.6 Strengths and limitations

The results of the current study are supported by the methods employed to retrieve and
analyse the data across studies. First, we retrieved and selected the studies by
systematically reviewing all the literature available. This approach minimises the exclusion
of studies that might reject the alternative hypothesis, and thus reduces reporting bias in the
results. Second, we analysed the data using a random-effects model, thus accounting for
heterogeneity across studies (which was large in some cases) and providing less stringent
effect sizes.

However, this study also has several limitations. First, the number of studies that could be
meta-analysed for each psychiatric outcome was low. This had a number of consequences.
First, the interpretation of the results of the regression and sub-groups analyses could be
unreliable because it decreased even more the number of studies that were compared.
Second, in order to increase power, we merged studies that predicted broad definitions of
psychiatric disorders and symptoms (i.e., internalising and externalising). However,
irritability may have different clinical correlates with more specific constructs such as
depression and anxiety, or ADHD, CD and ODD. Third, the problem of having few studies
was accentuated by the relatively few research groups that examined associations as well as
the overlap in samples and cohorts tested. Clearly, more studies are needed, especially from
independent research groups and different samples.

There were also methodological limitations, which could contribute to the high
heterogeneity across studies. First, the instruments to assess both irritability and psychiatric
outcomes differed substantially across studies. In the case of different definitions of
irritability, the most visible difference was that some studies used ODD dimensions,
whereas other employed DMDD or SMD as irritability. However, even within the ODD
dimension or DMDD definitions, there were also differences, not only due to different
instruments with different items but also because no instruments specifically designed to
assess DMDD were employed. Therefore, studies used items from ODD or MDD sections,
sometimes a combination of both. Another limitation for the interpretability of the results is
the disparity of ages and years at follow-ups. The meta-analysis assumes that irritability
measured in pre-school might have the same longitudinal correlates than the irritability
measured in adolescence. Unfortunately, due to the low number of studies, we could not
formally test whether these associations differed by age at baseline or follow-ups. Another
potential source of heterogeneity is the type of analyses employed in each study; that is,
whereas some studies used continuous scores of irritability to predict future outcomes,
other studies compared groups with high and low irritability (e.g., DMDD vs HV) in their
future outcomes. Finally, 36% of studies predicting psychiatric disorders and 56% of
studies predicting psychiatric symptoms did not adjust their longitudinal predictions for
baseline measures of the outcome, or in the case of ODD dimensions, for the presence of headstrong symptoms, which might inflate or confound the estimates.

2.5 Conclusion

Chronic severe irritability is prospectively associated with depressive and anxiety disorders, as opposed to externalising disorders. The specificity for psychiatric symptoms, as opposed to disorders, is not as clear. More studies are needed to robustly examine the existence of publication bias, especially for disorders different than depression and anxiety, and sources of heterogeneity between studies.

The reasons for the specific association between irritability and internalising disorders are unknown. So far, twin and family studies suggest that both conditions share genetic variance. However, no other factors have been examined. In Chapters 3 and 4, I examine potential factors for the association between irritability and depression.
CHAPTER 3: Sex differences in the associations between vagal reactivity and ODD symptoms

3.1 Background

The results from Chapter 2 highlight the specific association between chronic severe irritability and future depressive disorder. Therefore, it can be claimed that chronic irritability in youth is a risk factor for the development of depression (Stringaris & Goodman, 2009a). However, little is known about the risk factors for chronic irritability itself. Given the close link between irritability and depression, it is plausible to think that both entities share, to some extent, the same risk factors. This has been shown at the genetic and family level (Stringaris et al., 2012b; Whelan et al., 2015), but other potential factors remain to be tested, especially in early life. As mentioned in section 1.8.2, two of the most well-known risk factors for the development of depression are being female and having aberrant stress reactivity (Nolen-Hoeksema, 2001). In the current study, I examine how sex and stress reactivity interact to predict distinct dimensions of ODD symptoms in 29-month infants up to age 5.

3.1.1 Respiratory sinus arrhythmia as estimator of vagal reactivity

One method to measure stress reactivity is by looking at changes in vagal tone in response to a challenge. Commonly, changes in vagal tone have been assessed through changes in respiratory sinus arrhythmia (RSA). RSA is the natural variation in heart rate that occurs in each breathing cycle leading to increased and decreased heart rate during inspirations and...
expiration, respectively. Under stressful conditions, heart rate is expected to increase, which is translated in a reduction of RSA (i.e. vagal reactivity).

According to Porges’ Polyvagal Theory, vagal reactivity plays a pivotal role in emotional and behavioural regulation (Porges, 2007). The Polyvagal Theory asserts that engagement with the environment is physiologically mediated by the ventral vagal complex. The ventral vagal complex consists of pathways originating in the nucleus ambiguous, which regulates vagal influence on the heart and are linked to those that regulate muscles controlling facial expressions and vocalizing. The theory hypothesises that an adaptive response to stress involves a reduction in vagal tone (i.e., vagal reactivity), which activates physiological and psychological resources that result in efficient actions without activating flight-fight responses associated with the sympathetic nervous system. In other words, a reduction in vagal tone under stress conditions is translated in a better emotional and behavioural regulation.

This hypothesis has been supported by several studies, in which a reduction in RSA in response to a challenge (i.e., vagal reactivity) in children has been associated with a better emotional and behavioural regulation. For example, increased vagal reactivity have been related most consistently to fewer externalising (Calkins, Blandon, Williford, & Keane, 2007; Calkins & Keane, 2004; El-Sheikh, Harger, & Whitson, 2001; Graziano, Keane, & Calkins, 2007) and internalising problems (Calkins et al., 2007; Calkins & Keane, 2004; Gentzler, Santucci, Kovacs, & Fox, 2009; Graziano & Derefinko, 2013). However, a meta-analysis that included 44 studies and 4996 children concluded that the evidence for the effect of vagal reactivity on externalising and internalising symptoms was small;
furthermore, there was substantial heterogeneity across studies (Graziano & Dereffinko, 2013).

### 3.1.2 Sex differences in the association between vagal reactivity and psychiatric outcomes

One possible explanation for both the low effect estimates and the heterogeneity across studies might be that the association between vagal reactivity and emotional/behavioural regulation differs by sex. More than 20 years ago, Eisenberg et al. (1995) found that higher vagal tone was associated with improved social competence and emotion regulation in boys; however, this association had the opposite direction in girls. Recent studies have found the same pattern of sex differences (Hinnant & El-Sheikh, 2013; Morales, Beekman, Blandon, Stifter, & Buss, 2015). For example, Hinnant and El-Sheikh (2013) found that girls with higher vagal reactivity during a frustration task were more likely to be characterised by high externalising and internalising symptoms, whereas this was true for boys with weaker vagal reactivity or vagal tone augmentation during the same task. Similarly, Morales et al. (2015) found that an exuberance temperament was predictive of externalising symptoms in girls only when they showed vagal reactivity at 24 and 42 months, and in boys only when they augmented vagal tone at 24 months.

These sex differences are potentially of great interest in the context of development of psychopathology, whereby rates of neurodevelopmental and externalising disorders are higher in boys than girls before puberty (Lewinsohn, Hops, Roberts, Seeley, & Andrews, 1993), followed by a female predominance mainly for affective disorders (Angold & Rutter, 1992). The reasons for these differences are not well understood, although they are
probably explained by, in part, differential exposure to risks (Rutter, Caspi, & Moffitt, 2003), including prenatal influences (Glover & Hill, 2012), or differences in susceptibility to environmental stressors and adversity (Nolen-Hoeksema, 2001).

For example, longitudinal studies in large population-based samples have shown associations between low birth weight (Costello, Worthman, Erkanli, & Angold, 2007) and prenatal maternal depression (Quarini et al., 2016) with adolescent depression in females but not males. Relatedly, previous findings in the cohort investigated in this work showed that prenatal anxiety predicted internalising symptoms in girls but not in boys in the presence of low maternal stroking soon after birth (Sharp, Hill, Hellier, & Pickles, 2015). One possibility is that, in females, prenatal risks are associated with elevated physiological and emotional reactivity that may, in turn, generate vulnerability to stress and affective disorders (Murray, Halligan, Adams, Patterson, & Goodyer, 2006). In contrast, externalising disorders in males may arise from a failure to inhibit evolved sex-typical behaviours such as aggression through physiological and emotional arousal (Glover & Hill, 2012). A further study using the same cohort as the present work, showed, for example, that low birth weight and prenatal maternal anxiety were associated with increasing vagal reactivity assessed at 27 weeks in females, and with decreasing vagal reactivity in males (Tibu et al., 2014). In the present study, I will test whether this sex difference in vagal reactivity is associated with a distinct development of psychiatric symptoms, particularly, ODD symptoms.
3.1.3 Distinct symptom dimensions in ODD

In the last years, ODD in young people has been conceptualised as an emotional dysregulation disorder (Cavanagh, Quinn, Duncan, Graham, & Balbuena, 2017) rather than purely an externalising manifestation (Cavanagh et al., 2017; Rowe, Maughan, Pickles, Costello, & Angold, 2002). Moreover, as described in section 1.3.2 of this thesis, several studies have revealed the existence of different ODD dimensions representing both behavioural and emotional dysregulation (Burke et al., 2014; Ezpeleta et al., 2012; Stringaris & Goodman, 2009c). The two dimensions, i.e., “irritability” and “headstrong”, are associated with different longitudinal outcomes (Althoff et al., 2014; Burke et al., 2010; Ezpeleta et al., 2015; Leadbeater & Homel, 2015; Rowe et al., 2010; Stringaris & Goodman, 2009a; Whelan et al., 2013) and, possibly, distinct aetiologies (Stringaris et al., 2012b). Specifically, the irritability dimension has been mostly associated with future depression in adolescents (Althoff et al., 2014; Burke, 2012; Burke et al., 2010; Stringaris & Goodman, 2009a; Whelan et al., 2013) and pre-school children (Dougherty et al., 2013; Ezpeleta et al., 2015), whereas the headstrong dimension has been associated with future conduct disorder (CD) and ADHD (Burke et al., 2010; Stringaris & Goodman, 2009a).

Mirroring these associations, findings from the GSMS showed that ODD in 9-year girls was associated with higher odds of future depression and anxiety at age 16, whereas in boys, the association was with CD (Rowe et al., 2010). In addition, girls with ODD show more comorbidity with internalising problems whereas boys with ODD present higher rates of ADHD (Trepat & Ezpeleta, 2011). In terms of sex differences in irritability and other ODD symptoms, the results are mixed. Some studies have not found differences in rates of
irritability between males and females (Copeland et al., 2015). However, those studies reporting sex differences have found that girls score higher in irritability symptoms (Riglin et al., 2017; Silva et al., 2014; Stringaris et al., 2012b) whereas boys report more frequently symptoms within the headstrong dimension such as ”annoying people deliberately” and ”blaming others” (Stringaris et al., 2012b; Trepot & Ezpeleta, 2011).

3.1.4 Irritability in pre-school years

Whereas the literature on irritability during childhood and adolescence is substantial, few research groups have focused on examining irritability during the pre-school years (i.e., 3 to 5 years of age). This is surprising given that irritability peaks at these ages in the general population (Leibenluft & Stoddard, 2013).

Several studies have investigated chronic irritability as a part of ODD dimension in pre-school children (Ezpeleta et al., 2012; Ezpeleta et al., 2015; Ezpeleta & Penelo, 2015; Lavigne et al., 2001). Cumulatively, findings from these studies suggest that, first, ODD dimensions are already evident in the pre-school years (Ezpeleta et al., 2012; Lavigne et al., 2015; Lavigne et al., 2014). Second, there is measurement invariance across gender for ODD dimensions in pre-school children (Ezpeleta & Penelo, 2015; Lavigne et al., 2015); that is, the same construct is being measured in boys and girls. Finally, similar with older children, the headstrong dimension correlates with disruptive disorders (Ezpeleta et al., 2012) whereas the irritability dimension correlates with internalising problems (Ezpeleta et al., 2012; Lavigne et al., 2014) and, in those with ODD, also with poorer functional outcomes overall (Ezpeleta et al., 2015).
Nevertheless, these studies either used cross-sectional designs or, when employing longitudinal analyses, treated ODD dimensions as predictors instead of outcomes. Therefore, it is unclear what the precursors of each ODD dimension are. To my knowledge, only one recent study has used ODD dimensions as outcomes using distinct neuropsychological correlates as predictors (Griffith et al., 2017). In this study, the authors found that working memory, control inhibition and sustained attention at age three were associated with the ODD dimension of negative affect at age six, whereas delay aversion predicted the oppositional behaviour dimension (Griffith et al., 2017). Thus, there are precursors of ODD dimensions that remain to be tested.

### 3.1.5 Aims and hypotheses

Few studies have studied chronic irritability in pre-school children and examined predictors of irritability or ODD dimensions. In the current study, we explored the predictors of ODD dimensions, including irritability, at ages 2.5, 3.5 and 5 years. We set the following aims and hypotheses.

First, there is evidence to suggest the existence of sex differences in the association between vagal reactivity and future psychiatric symptoms. Specifically, higher vagal reactivity in girls, but lower in boys, has been shown to be associated with higher psychiatric symptoms later in life. Therefore, our first aim is to test whether this is also the case for the prediction of ODD symptoms. We hypothesised that increased vagal reactivity in females and decreased vagal reactivity in males would be associated with later ODD symptoms.
Second, there is evidence to suggest that girls and boys differ in their phenotype presentation and outcomes of ODD. Specifically, girls with ODD symptoms are more prone to irritability and internalising problems; and boys with ODD symptoms are more prone to headstrong symptoms and externalising problems. Therefore, our second aim was to test whether sex differences in the association between vagal reactivity and ODD symptoms were also associated with specific ODD dimensions. We hypothesised that in girls vagal reactivity would be associated with risk for later depression, namely the irritability ODD dimension, while, in boys, with risk for later disruptive disorders, namely the headstrong ODD dimension.

3.2 Methods

3.2.1 Participants

The participants were members of the Wirral Child Health and Development Study, a prospective epidemiological longitudinal study of consecutive first-time mothers who booked for antenatal care at 12 weeks gestation between 12/02/2007 and 29/10/2008 with surviving singleton-births.

The study design has been previously described (Sharp et al., 2015; Sharp et al., 2012; Tibu et al., 2014). Briefly, this study involves a series of assessments, some of the whole cohort of 1233 children and some involving a stratified random sub-sample of 316 children. This enables intensive measurement to be employed efficiently within a subsample while retaining the ability to recover estimates relating to the general population. Approval for the procedures was obtained from the Cheshire North and West Research Ethics Committee (UK). Potential participants were introduced to the study by research midwives who
obtained informed consent. Representation of psychosocial risk in the intensive sample was increased through stratification by psychological abuse between partners (Moffitt et al., 1997). The stratifier was associated with ODD symptoms yielding correlations at 2.5, 3.5 and 5 years of 0.18 (p=0.006), 0.13 (p<0.001) and 0.19 (p<0.001), respectively.

Figure 3.1 shows a participant flow diagram of this study. The cohort of 1233 mothers-to-be had a mean age at recruitment of 26.8 years (SD=5.8, range 18-51). There were 316 recruited to the intensive sub-sample at 32 weeks pregnancy, and 270 mothers and infants provided data for vagal reactivity when the infants were 29 weeks old at least in one procedure (see measures section) and 244 in all five procedures. Maternal reports of ODD symptoms were available on 253 of the intensive sub-cohort at 2.5 years (Mean=2.1, SD=0.3, range 2-3), on 826 of the whole cohort at 3.5 years (M=3, SD=0.2, range 3-4) and on 770 of the whole cohort at 5 years (M=4.3, SD=0.5, range 4-6). Mothers whose infants were assessed at 29 weeks were slightly older than those in the whole cohort, mean age 27.9 years (SD= 6.2; range= 18-51).

3.2.2 Measures

3.2.2.1 Respiratory Sinus Arrhythmia (RSA)

RSA was computed from an ECG recording made during five procedures. These five methods were 1) the helper-hinderer (Hamlin, Wynn, & Bloom, 2007), 2) the novel toy exploration (Calkins & Dedmon, 2000), and three sequences of the ‘Still-Face’ 3) face to face engagement with mother, 4) the still-face , and 5) repair (Tronick, Als, Adamson, Wise, & Brazelton, 1978). These procedures are briefly described below.
Procedure 1, The Helper-Hinderer is an experimental paradigm developed to assess whether infants favour prosocial acts (Hamlin et al., 2007). The infant is seated on the mother’s lap and views a large display (3x5 feet) situated in front of him/her approximately 6 feet away in which a coloured shape (square, circle, triangle) with googly eyes is shown either helping another up a slope (helper trial) or hindering another’s progress up the slope (hinderer trial). Helping trials and hindering trials are alternated throughout and the series of learning trials are ended once the infant has shown a predetermined level of habituation to the stimuli, or when the maximum number of pre-set trials has been reached (14 trials). Once the learning trials have ended the infant are given a preference task, between the helper-shape or hinderer-shape. The duration of the learning procedure is not standard and varies depending on how quickly the infant habituates to the presentation of the stimuli. In the current study, the mean duration of the procedure was 3.74 minutes, SD 1.20, minimum 0.88 minutes, maximum 8.09 minutes. RSA was calculated for the last 2 minutes of this procedure to ensure standardisation of infants’ looking times.
Figure 3.1. Participant flow diagram. Note: Although not relevant for this study, there were two assessments points before ‘birth’ and two more assessments points between ‘birth’ and ‘29 weeks’.

Procedure 2 - The Novel Toy Exploration Procedure is a 2-minute episode in which the infant is presented at a table with a 4-facet triangular pyramid-shaped toy to explore for two minutes while sitting on the mother’s knees. This has been used in previous studies to assess baseline vagal tone (Calkins & Dedmon, 2000).
Procedures 3, 4, and 5 - Still-face procedures were conducted with the infant in a high chair facing the mother. They comprised two minutes of face to face playful interactions without toys, followed by two minutes during which the mother was asked to be unresponsive to her child’s communications (the ‘Still Face’), after which she became again responsive (the ‘repair’) (Tronick et al., 1978). The Still Face has been used extensively in studies of vagal reactivity (Moore, 2009; Moore & Calkins, 2004).

The patterns of RSA across the five procedures were very similar in males and females, and there were no significant sex differences in any procedure (Tibu et al., 2014). A principal components analysis yielded one factor with an eigenvalue of 3.54 that explained 70.73% of the total variance (Sharp et al., 2012). Post hoc tests in repeated measures ANOVA showed that RSA during the Still Face was significantly lower than in each of the other four procedures (all values of p <0.01) consistent with vagal reactivity to the stressor.

3.2.2.2 Oppositional defiant disorder symptoms

Maternal reports of child ODD symptoms were assessed at 2.5, 3.5 and 5 years using the preschool Child Behavior Checklist (CBCL), which has been extensively employed in studies of child and adolescent emotional and behavioural disorders (Achenbach & Rescorla, 2000). ODD dimensions of irritability and headstrong symptoms were generated following the results of previous confirmatory factor analyses (CFA) in adolescents and adults (Aebi et al., 2013; Althoff et al., 2014; Stringaris et al., 2012b) across the items of the ODD subscale.
The items employed by Stringaris et al. (2012b) were drawn from the Youth Self-Report (ages 11-18) and Adult Self-Report (ages 18-59) versions of the ASEBA family of instruments. These included “have a hot temper”; “stubborn” (for adolescents) or “stubborn, sullen or irritable” (for adults), and “mood/feelings change suddenly” for irritability; and “disobey parents” (only for adolescents), “mean to others”, “destroy others’ things”, “disobey at school”, and “tease others a lot” for headstrong. In the present study we use the Preschool version of the CBCL and, as some items differ, we provide here the results of CFA from our own study. Based on the CFA from Stringaris et al. (Stringaris et al., 2012b) Irritability comprised the items “Angry moods“, “Stubborn, sullen or irritable“, and “Temper tantrums or hot temper”; Headstrong comprised the items “Defiant“, “Disobedient“, and “Uncooperative“. Each item is scored using a 3-point Likert scale (0-Not true, 1-Somewhat or sometimes true, and 2-Very true or often true).

We used multivariate probit confirmatory factor analysis for ordinal data in Mplus using the WLSMV estimator. To consider a model as showing ‘acceptable’ fit, we required a Confirmatory Fit Index (CFI)>0.90, and Root Mean Square Error Approximation (RMSEA) <0.08, and for a ‘good’ fit, we required a CFI>0.95, and RMSEA<0.06 (Brown, 2006). The $\chi^2$ difference test was used to compare nested models (i.e. one single factor models against a two-factor model), where improvements in model fit by the nested –less constrained and more parsimonious– model are tested.

### 3.2.2.3 Confounders

Previous studies have shown that younger maternal age (Fergusson & Woodward, 1999; Harden et al., 2007) socioeconomic status (Bradley & Corwyn, 2002), and maternal
depressive symptoms (Brennan et al., 2000) are associated with emotional and behavioural problems in offspring. Therefore, we employed these measures as covariates in the analyses.

Maternal age was recorded in the 20 weeks gestation questionnaire with the extensive sample (N = 1233).

Socioeconomic status was measured using the revised English Index of Multiple Deprivation (IMD); (Noble et al., 2004) based on data collected from the UK Census in 2001. According to this system, postcode areas in England are ranked from most deprived (i.e. IMD of 1) to least deprived (i.e. IMD of 32,482) based on deprivation in seven domains: income, employment, health, education and training, barriers to housing and services, living environment, and crime. All mothers were given IMD ranks according to the postcode of the area where they lived and assigned to a quintile based on the UK distribution of deprivation.

Maternal depression was assessed by self-report using the Edinburgh Postnatal Depression Scale (Cox et al. 1996) at 2.5 and 3.5 years, and the Center for Epidemiologic Studies Depression Scale (Radloff, 1977) at 5 years. Depression scores at these three assessment points were included in analyses to control for possible biasing effects on maternal reports of child ODD symptoms.
3.2.3 Statistical analyses

The sample design and attrition were accounted for in the estimates of descriptive statistics by the use of inverse probability weighting (Lehtonen & Pahkinen, 2004). Pickles, Dunn & Vasquez-Barquero (1995) provide a description specific to this type of study design.

To test our developmental hypotheses we fitted two structural equation models (SEM) estimated using full maximum likelihood and the auxiliary command, a method which enables participants with incomplete observations to be included under the assumption of the missingness being missing-at-random (MAR) (Graham, 2003). This method yields estimates that account for the sample stratification and for attrition associated with covariates and observed values of outcomes (e.g., RSA under still-face or ODD symptoms).

The five measures of RSA were incorporated into the model as dependent indicators of vagal tone and vagal reactivity factors, enabling their observation only in the intensive sample to be accounted for under the MAR assumption. For examining sex differences, we made use of the “knownclass” approach for the two groups and Wald tests of main-effect and interaction contrasts formed as linear functions of parameters defined within model constraints.

In the first model, we examined relationships to overall ODD symptoms and in the second to the ODD dimensions of irritability and headstrong. The standardized estimates reported are from the model in which all observed variables and factors are scaled to have unit variance. Sex difference contrasts were defined in the traditional effect size manner by semi-standardized effects where the predicted variable had unit variance for both males and females.
3.3 Results

3.3.1 Descriptive Statistics

Table 3.1 provides summary statistics for each of the measures for males and females separately. In most cases, there were no significant differences between males and females. However, mothers of boys reported higher levels of depressive symptoms on themselves than mothers of girls when the children were 5 years old. Also, mothers reported higher levels of overall ODD symptoms in boys than in girls at 3.5 and 5 years. This difference was accounted mainly by higher levels of irritability symptoms in boys at ages 3.5 and 5 years, though there was a trend for higher levels of headstrong symptoms as well.

3.3.2 Confirmatory factor analyses of ODD symptoms

Results from the CFA are shown in Table 3.2. According to goodness-of-fit indices, a two-factor model that decomposed ODD into irritability and headstrong showed significant improvements compared to the one-factor model.

Although a one-factor model for the ODD items generally gave a satisfactory CFI across all three ages and both sexes, the RMSEAs were less so. Chi-square difference tests comparing these one-factor models to two-factor models that decomposed ODD into irritability and headstrong showed significant improvements in all cases, with the two-factor models having CFI all >0.95 and RMSEAs all less than .07, except among the youngest girls. Correlation between the two factors across the 3 points of assessment ranged 0.70-0.86 in boys and 0.75-0.92 in girls, in both cases the correlation increasing with age. These correlations are in line with previous studies (Ezpeleta et al., 2012; Stringaris & Goodman, 2009c).
Table 3.1. Summary of measures in boys and girls

<table>
<thead>
<tr>
<th></th>
<th>Boys</th>
<th></th>
<th></th>
<th></th>
<th>Girls</th>
<th></th>
<th></th>
<th>T-test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean (SD)</td>
<td>Min</td>
<td>Max</td>
<td>N</td>
<td>Mean (SD)</td>
<td>Min</td>
<td>Max</td>
<td></td>
</tr>
<tr>
<td>Mother’s age at 20</td>
<td>599</td>
<td>26.84 (5.80)</td>
<td>18</td>
<td>51</td>
<td>634</td>
<td>26.79 (5.86)</td>
<td>18</td>
<td>43</td>
<td>0.862</td>
</tr>
<tr>
<td>week scan (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Socioeconomic</td>
<td>598</td>
<td>33.31 (20.84)</td>
<td>1.69</td>
<td>80.62</td>
<td>632</td>
<td>32.57 (19.61)</td>
<td>1.69</td>
<td>79.73</td>
<td>0.519</td>
</tr>
<tr>
<td>status (IMD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maternal depression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.5 years</td>
<td>115</td>
<td>4.91 (5.05)</td>
<td>0</td>
<td>24</td>
<td>127</td>
<td>5.46 (4.93)</td>
<td>0</td>
<td>23</td>
<td>0.391</td>
</tr>
<tr>
<td>3.5 years</td>
<td>389</td>
<td>5.16 (4.59)</td>
<td>0</td>
<td>24</td>
<td>423</td>
<td>4.92 (4.20)</td>
<td>0</td>
<td>19</td>
<td>0.423</td>
</tr>
<tr>
<td>5 years</td>
<td>352</td>
<td>8.09 (8.67)</td>
<td>0</td>
<td>55</td>
<td>386</td>
<td>6.95 (6.85)</td>
<td>0</td>
<td>38</td>
<td><strong>0.046</strong></td>
</tr>
<tr>
<td>RSA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Helper-hinderer</td>
<td>131</td>
<td>3.28 (0.83)</td>
<td>1.09</td>
<td>5.76</td>
<td>139</td>
<td>3.37 (0.85)</td>
<td>0.95</td>
<td>5.46</td>
<td>0.366</td>
</tr>
<tr>
<td>Novel toy</td>
<td>128</td>
<td>2.93 (0.75)</td>
<td>1.47</td>
<td>4.68</td>
<td>138</td>
<td>3.01 (0.86)</td>
<td>0.39</td>
<td>5.75</td>
<td>0.464</td>
</tr>
<tr>
<td>Engagement</td>
<td>123</td>
<td>3.38 (0.77)</td>
<td>1.58</td>
<td>5.19</td>
<td>134</td>
<td>3.27 (0.89)</td>
<td>0.61</td>
<td>6.04</td>
<td>0.301</td>
</tr>
<tr>
<td>Still face</td>
<td>121</td>
<td>2.80 (0.73)</td>
<td>0.56</td>
<td>4.56</td>
<td>132</td>
<td>2.86 (0.79)</td>
<td>0.73</td>
<td>5.03</td>
<td>0.550</td>
</tr>
<tr>
<td>Repair</td>
<td>119</td>
<td>3.40 (0.91)</td>
<td>1.28</td>
<td>5.84</td>
<td>128</td>
<td>3.30 (0.99)</td>
<td>0.55</td>
<td>5.89</td>
<td>0.421</td>
</tr>
<tr>
<td>Vagal reactivity</td>
<td>116</td>
<td>0.49 (0.55)</td>
<td>-0.74</td>
<td>2.08</td>
<td>128</td>
<td>0.42 (0.56)</td>
<td>-0.91</td>
<td>1.98</td>
<td>0.320</td>
</tr>
</tbody>
</table>
Table 3.1. Summary of measures in boys and girls (Continued)

<table>
<thead>
<tr>
<th></th>
<th>Boys</th>
<th></th>
<th></th>
<th></th>
<th>Girls</th>
<th></th>
<th></th>
<th></th>
<th>T-test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean (SD)</td>
<td>Min</td>
<td>Max</td>
<td>N</td>
<td>Mean (SD)</td>
<td>Min</td>
<td>Max</td>
<td>p-value</td>
<td></td>
</tr>
<tr>
<td>ODD Symptoms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.5 years</td>
<td>123</td>
<td>3.28 (2.23)</td>
<td>0</td>
<td>12</td>
<td>126</td>
<td>2.98 (2.08)</td>
<td>0</td>
<td>12</td>
<td>0.269</td>
<td></td>
</tr>
<tr>
<td>3.5 years</td>
<td>395</td>
<td>3.12 (2.50)</td>
<td>0</td>
<td>12</td>
<td>431</td>
<td>2.76 (2.38)</td>
<td>0</td>
<td>12</td>
<td><strong>0.035</strong></td>
<td></td>
</tr>
<tr>
<td>5 years</td>
<td>369</td>
<td>2.90(2.63)</td>
<td>0</td>
<td>12</td>
<td>401</td>
<td>2.46 (2.23)</td>
<td>0</td>
<td>12</td>
<td><strong>0.017</strong></td>
<td></td>
</tr>
<tr>
<td>ODD Irritability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.5 years</td>
<td>122</td>
<td>1.62 (1.26)</td>
<td>0</td>
<td>6</td>
<td>130</td>
<td>1.55 (1.23)</td>
<td>0</td>
<td>6</td>
<td>0.624</td>
<td></td>
</tr>
<tr>
<td>3.5 years</td>
<td>395</td>
<td>1.54 (1.40)</td>
<td>0</td>
<td>6</td>
<td>431</td>
<td>1.34 (1.22)</td>
<td>0</td>
<td>6</td>
<td><strong>0.035</strong></td>
<td></td>
</tr>
<tr>
<td>5 years</td>
<td>369</td>
<td>1.45 (1.50)</td>
<td>0</td>
<td>6</td>
<td>401</td>
<td>1.20 (1.31)</td>
<td>0</td>
<td>6</td>
<td><strong>0.012</strong></td>
<td></td>
</tr>
<tr>
<td>ODD Headstrong</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.5 years</td>
<td>123</td>
<td>1.37 (1.33)</td>
<td>0</td>
<td>6</td>
<td>130</td>
<td>1.44 (1.19)</td>
<td>0</td>
<td>6</td>
<td>0.136</td>
<td></td>
</tr>
<tr>
<td>3.5 years</td>
<td>395</td>
<td>1.57 (1.36)</td>
<td>0</td>
<td>6</td>
<td>431</td>
<td>1.42 (1.34)</td>
<td>0</td>
<td>6</td>
<td>0.090</td>
<td></td>
</tr>
<tr>
<td>5 years</td>
<td>369</td>
<td>1.44 (1.42)</td>
<td>0</td>
<td>6</td>
<td>401</td>
<td>1.26 (1.34)</td>
<td>0</td>
<td>6</td>
<td><strong>0.063</strong></td>
<td></td>
</tr>
</tbody>
</table>

RSA, Respiratory sinus arrhythmia. ODD, Oppositional defiant disorder. *p*<0.05 in bold
Table 3.2. Model fit of ODD symptoms in Confirmatory Factor Analyses

<table>
<thead>
<tr>
<th></th>
<th>One–Factor</th>
<th>Two–Factor</th>
<th>Diff χ²</th>
<th>Diff df</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>CFI</td>
<td>RMSEA</td>
<td>CFI</td>
<td>RMSEA</td>
</tr>
<tr>
<td>Boys</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.5 years</td>
<td>123</td>
<td>0.907</td>
<td>0.187</td>
<td>0.957</td>
<td>0.138</td>
</tr>
<tr>
<td>3.5 years</td>
<td>395</td>
<td>0.971</td>
<td>0.105</td>
<td>0.992</td>
<td>0.058</td>
</tr>
<tr>
<td>5 years</td>
<td>369</td>
<td>0.976</td>
<td>0.125</td>
<td>0.994</td>
<td>0.065</td>
</tr>
<tr>
<td>Girls</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.5 years</td>
<td>130</td>
<td>0.943</td>
<td>0.131</td>
<td>0.971</td>
<td>0.093</td>
</tr>
<tr>
<td>3.5 years</td>
<td>431</td>
<td>0.980</td>
<td>0.091</td>
<td>0.997</td>
<td>0.040</td>
</tr>
<tr>
<td>5 years</td>
<td>401</td>
<td>0.981</td>
<td>0.076</td>
<td>0.986</td>
<td>0.071</td>
</tr>
</tbody>
</table>

3.3.3 SEM applied to the analysis of RSA predicting ODD symptoms.

The models of interest are shown in Figure 3.2 for the single ODD symptoms dimension and in Figure 3.3 for the two ODD dimensions, with different parameters being allowed for boys and girls. The RSA measures were decomposed into a general vagal tone factor and vagal reactivity defined by differential response under the Still-Face condition. Separate effects of vagal tone and vagal reactivity on ODD were then estimated. To account for the sample stratification, the psychological abuse variable was included as an auxiliary variable and allowed to correlate with all other variables; and the potential confounders of maternal age and socioeconomic status were included as covariates influencing baseline vagal tone, vagal reactivity and ODD symptoms. For simplification, the psychological abuse variable and the covariates are not shown in Figures 3.2 and 3.3.
3.3.4 Vagal tone, vagal reactivity and ODD symptoms

The model shown in Figure 3.2 fitted satisfactorily (Overall: CFI=0.950, RMSEA=0.031; Males: CFI=0.926, RMSEA=0.038; Females: CFI=0.967, RMSEA=0.025).

Figure 3.2. Summary of structural equation modelling analysis of baseline vagal tone and vagal reactivity predicting overall ODD symptoms. Standardized coefficients with standard errors in bold, $p<0.05$. G, girls; B, boys.

Following our first hypothesis, we tested whether vagal reactivity was differently associated with ODD symptoms in boys and girls. Wald test showed that there was a significant sex difference ($\chi^2(1)=3.98; p=0.046$) arising from the negative association in males (-0.17, 95%CI -0.40, 0.07, p=0.159) and positive association in females (0.18, 95%CI -0.04, 0.40, p=0.100). No overall main effect of vagal reactivity was evident in the whole sample (0.03, 95%CI -0.13, 0.19; p=0.725).
For the association between baseline vagal tone and ODD symptoms there was no sex difference ($\chi^2(1)=0.45; \ p=0.504$) and associations were close to zero, both in boys (-0.09, 95%CI -0.31, 0.13; $p=0.418$) and girls (-0.01, 95%CI -0.22, 0.21; $p=0.947$), and there was no main effect of vagal tone in the whole sample (-0.03, 95%CI -0.18, 0.11, $p=0.669$).

3.3.5 Vagal tone, vagal reactivity and ODD irritability and headstrong dimensions

The model shown in Figure 3.3 examined dimensional specific effects and fitted a slightly better than the model with the single factor (Overall: CFI=0.968, RMSEA=0.027; Males: CFI=0.955, RMSEA=0.032; Females: CFI=0.978, RMSEA=0.022).

**Figure 3.3.** Summary of structural equation modelling analysis of baseline vagal tone and vagal reactivity predicting ODD dimensions of irritability and headstrong symptoms. Standardized coefficients with standard errors in bold, $p<0.05$. G, girls; B, boys.
For girls, the model estimated a significant positive association of vagal reactivity with irritability (0.20, 95%CI 0.003, 0.40, p=0.047), and a smaller non-significant association with headstrong (0.14, 95%CI -0.09, 0.38, p=0.225). For boys, negative associations were estimated for both dimensions though neither was significant (irritability -0.14, 95%CI -0.37, 0.10, p=0.250 and headstrong -0.16, 95%CI -0.38, 0.07, p=0.172). These four coefficients were decomposed into an overall average effect, a difference by ODD dimension, a difference by sex, and an ODD dimension by sex interaction. Only the sex difference was significant (coefficient = 0.70, p=0.038) there being no evidence for a difference by dimension (p=0.457) nor any interaction (p=0.701). The pattern of associations is illustrated in Figure 3.4 where the simple linear fits to the estimated irritability, headstrong and vagal reactivity factor scores are displayed. The parallel effects on the two symptom dimensions are clear together with a marked difference for boys and girls that is consistent across dimensions.
Figure 3.4. Interplay between sex and respiratory sinus arrhythmia (RSA) reactivity in the prediction of irritability and headstrong ODD symptom dimensions.

In the whole sample, vagal reactivity was neither associated with headstrong symptoms (0.01, 95%CI -0.16, 0.17, p=0.926) nor with irritability symptoms (0.05, 95%CI -0.10, 0.20, p=0.505). Baseline vagal tone was neither associated with headstrong (Males: -0.12, 95%CI -0.32, 0.08, p=0.255; Females: -0.06, 95%CI -0.28, 0.17, p=0.630) nor with irritability (Males: -0.05, 95%CI -0.29, 0.18, p=0.658; Females: 0.03, 95%CI -0.17, 0.24, p=0.745).
3.4 Discussion

3.4.1 Summary of findings

Our first hypothesis was that vagal reactivity, assessed as a reduction in RSA to a stressor at 29 weeks, would be associated with future ODD symptoms in opposite direction in boys and girls. Specifically, we predicted that increased vagal reactivity would be positively associated with ODD symptoms in girls, and negatively associated with ODD symptoms in boys. The results of the current study supported this hypothesis.

Our second hypothesis was that, in girls, this effect would be predominant for irritability symptoms, whereas in boys this effect would be predominant for headstrong symptoms. The results of the current study did not fully support this hypothesis. Although the main effect of vagal reactivity on irritability was significant for girls, we found that the sex difference was true for each dimension, with no dimension-specific effects.

3.4.2 Sex differences in vagal reactivity as a predictor of ODD symptoms

Our findings are in line with previous studies that have found a sex by vagal reactivity interaction in the prediction of future socioemotional outcomes (Hinnant & El-Sheikh, 2013; Morales et al., 2015). Furthermore, the association was in the same direction; that is, higher vagal reactivity was related to poorer outcomes in girls (i.e., more ODD symptoms) and better outcomes in boys (i.e., less ODD symptoms).

In a previous report using the same cohort as in the current study, low birth weight and prenatal anxiety were associated with increasing vagal reactivity in females, and decreasing vagal reactivity in males (Tibu et al., 2014). In the present study, we have found that these
distinct patterns of vagal reactivity were associated with increased ODD symptoms. Altogether, these findings expand on our understanding of sex differences in the vulnerability for psychopathology. It has been demonstrated that some sex differences are explained by differential exposure to risks that operate in the same way in males and females (Rutter et al., 2003). In addition, there is growing evidence for sex-specific risks; for example, low birth weight and prenatal maternal depression have been found to increase the risk for depression in females but not males (Costello et al., 2007; Quarini et al., 2016). The findings from the current study add a further possibility; that even when the risks for psychiatric outcomes (e.g., ODD symptoms) are the same in males and females (e.g., prenatal maternal anxiety), the mechanisms through which these risks impact the outcomes may differ (e.g., increasing/decreasing vagal reactivity). That is, our findings are consistent with the possibility that males respond to stressors with reduced, and females with increased, arousal and emotional reactivity, and that each confers vulnerability in different ways. The risk of ODD symptoms in boys might arise from a failure to sufficiently activate inhibitory processes, but in girls might be explained from dysregulation of autonomic responses that are probably adaptive under some conditions (Obradovic, 2012). Another but not exclusive possibility is that parents might have different expectations from girls and boys in terms of their coping strategies under stressful conditions. Such expectations might result in differential reinforcement of coping strategies in boys and girls and, consequently, in different parenting styles and child-parent interactions (Chang, Olson, Sameroff, & Sexton, 2011). Therefore, even when boys and girls show the same responses to the same stressors, these child-parent interactions might lead to the development of different
emotional and behavioural regulation trajectories later in life (Kopala-Sibley et al., 2015).

Finally, as mentioned at the beginning of this paragraph, some evidence also suggests that differences in foetal programming might impact the response to stress challenges (Tibu et al., 2014). It is thus possible that prenatal risk factors are associated with later development of psychiatric symptoms through the mediation role of reactivity to stressful life experiences.

3.4.3 Sex differences in vagal reactivity as a predictor of ODD symptom dimensions

Our second set of analyses (i.e., the prediction of ODD dimensions by vagal reactivity) was an attempt to explain the ODD differences seen in boys and girls (Trepat & Ezpeleta, 2011) and how these sex differences might explain the risk for the development of externalising and internalising disorders later in life (Angold & Rutter, 1992; Lewinsohn et al., 1993). These analyses were motivated by the specific longitudinal association seen between irritability and depression (Vidal-Ribas et al., 2016). Prevalence rates of depression are higher in females after puberty, and females with ODD symptoms show higher levels of irritability and higher rates of internalising disorders than males (Riglin et al., 2017; Rowe et al., 2010; Stringaris et al., 2012b; Trepat & Ezpeleta, 2011). In contrast, males with ODD show more headstrong symptoms and higher rates of disruptive disorders than females (Rowe et al., 2010; Stringaris et al., 2012b; Trepat & Ezpeleta, 2011). Therefore, we hypothesised that girls would be at higher risk of irritability and boys at higher risk of headstrong symptoms. However, this hypothesis was not fully supported by the results.

It can be argued that we did not find dimension-specific effects because ODD symptoms at these early ages are not multidimensional, but instead a single unitary construct. However,
the results of the CFA in the current study suggested that this was not the case. The CFA in the current study showed that a two-factor model fitted better the data than a single factor model, in line with previous reports. Irritable and headstrong dimensions have been previously found in pre-schoolers employing other instruments (Ezpeleta et al., 2012; Lavigne et al., 2014), and these dimensions were associated with different correlates (Ezpeleta et al., 2015), confirming that the heterogeneity of ODD is present from early in life. Nevertheless, the CFA findings should be taken with caution because this was the first study to perform a CFA of ODD dimensions in children as young as 2.5 years old, and the first study to employ the pre-school version of the CBCL to generate these dimensions.

It is important to note that irritability and headstrong symptoms co-occur, and the correlation between both dimensions was high in our sample as in other studies (Stringaris & Goodman, 2009c). Specifically, correlations between irritable and headstrong dimensions in our sample ranged r=0.70-0.92 and tended to be non-significantly higher in girls. Therefore, it was expected that an increase in symptoms of one dimension will be accompanied by an increase in symptoms of the other dimension. Studies using latent class analyses (LCA), in which participants are grouped according to endorsement of symptoms, have not consistently found a pure high-only headstrong class (Althoff et al., 2014; Herzhoff & Tackett, 2016; Kuny et al., 2013) or a pure high-only irritability class (Boylan et al., 2017; Burke, 2012; Drabick & Gadow, 2012; Gadow & Drabick, 2012).

Furthermore, analyses of ODD dimensions trajectories have shown a substantial overlap between dimensions; that is, there is not an independent developmental course for irritability and headstrong symptoms (Boylan et al., 2017). Taken together, these findings
suggest that although ODD dimensions might have distinct longitudinal outcomes, these symptom dimensions are not independent of each other and are highly correlated.

In our study, although not significantly different, the coefficients were larger in magnitude for irritability in girls and for headstrong in boys. In girls, for whom we hypothesised a predominant effect towards irritability, higher stress reactivity was associated with both higher irritability and headstrong symptoms. However, only the main effect on irritability was significant. Thus, it still possible that girls were more prone to develop irritability symptoms than boys. Even though previous studies have not found pure irritable subgroups, those groups of participants with irritability symptoms (independently of having headstrong symptoms) may still present higher rates of mood disorders (Drabick & Gadow, 2012) and are at higher risk of developing these in the future (Althoff et al., 2014; Burke, 2012).

It is possible that sex differences in ODD dimensions are more evident later in life. Also, the conferred risk of each ODD dimension might be different in each developmental period. Reports of ODD dimensions in pre-school children have not found sex differences in the distribution of ODD symptoms (Ezpeleta et al., 2012; Ezpeleta & Penelo, 2015) or in trajectories of irritability (Ezpeleta et al., 2015). However, these reports have found small differences in the correlates of irritability. Specifically, in 3-year-old children, the ODD irritability dimension was associated with anxiety disorders in both boys and girls, but with ADHD in girls only (Ezpeleta et al., 2012). Importantly, it has been suggested that early onset irritability might be more related to ADHD than affective disorders and associated with a male preponderance, whereas irritability with an adolescent onset might be more
related to depression and more common in females (Riglin et al., 2017; Riglin et al., Under review). Therefore, the increased risk of affective disorders due to the presence of irritability might take place in late childhood and early adolescence, just before puberty, and might be more evident in females (Savage et al., 2015). For example, in a recent study in children aged 9-13 years-old, irritability was similarly correlated with externalising symptoms in both males and females; however, the correlation with internalising symptoms was higher in females. Moreover, the same pattern was seen in longitudinal associations with externalising and internalising symptoms 2 years later (Humphreys et al., 2018).

3.4.4 Sex differences in baseline vagal tone as a predictor of ODD symptoms and dimensions

In this study, we found a significant vagal reactivity by sex interaction. However, as opposed to other studies (Eisenberg et al., 1995; Hinnant & El-Sheikh, 2013; Morales et al., 2015), we did not find a baseline vagal tone by sex interaction in the prediction of neither ODD symptoms nor ODD dimensions. However, previous studies have employed single measures of baseline vagal tone. In contrast, we employed a latent variable approach to create a latent vagal tone measure with the shared variance of the vagal tone measured using five tasks and then generated the vagal reactivity measure by including a differential item contrast. This approach overcomes the methodological limitation of having task-specific baselines and the interpretation of vagal reactivity limited to a single task. In addition, the measurement of vagal tone in this study was collected in very young children (i.e., 29 weeks), as opposed to pre-school (Morales et al., 2015) and school-aged children (Eisenberg et al., 1995; Hinnant & El-Sheikh, 2013) in previous studies. Baseline vagal
tone has been shown to be moderately stable from pre-school or middle childhood onwards (Calkins & Keane, 2004; El-Sheikh, 2005) but less so in the first years of life (Bornstein & Suess, 2000), which might also have an impact in the results.

3.4.5 Strengths and limitations

This study had several strengths. First, our findings were based on a population-based sample followed from pregnancy to age 5 years, with good sample retention. Second, a latent variable approach to the analysis of RSA across contrasting procedures provided strong evidence of an overall level of vagal tone evident in low- and high-stress conditions. The differential item functioning approach provided a strong test of the specificity of vagal reactivity for later symptoms by first accounting for associations with the general vagal tone latent variable and then testing for any additional association specifically with vagal reactivity, as defined in the model by the contrast between vagal tone in the still face and the general latent variable. This approach has provided evidence for associations with vagal reactivity to a social stressor, but not with vagal tone under low-stress conditions, of prenatal risks (Tibu et al., 2014) and now of subsequent child behaviours. An additional strength of this study was that the ODD dimensions of irritability and headstrong were extracted by using the CFA independently modelled from a previous study (Stringaris & Goodman, 2009c).

The findings of this study must also be considered in light of its limitations. Although the design allowed us to examine prospective associations over a period of more than 4 years, it has left many questions regarding the intervening autonomic, emotional and behavioural processes yet to be investigated. We cannot assume causality from the reported associations.
because there may be unmeasured confounders, such as genetic factors and other sociodemographic indicators, affecting these relations. Similarly, we cannot assume that the sex differences arose from the hypothesised sex-dependent processes outlined earlier. Instead, it is possible that they arose from differences in the ways boys and girls responded to the Still Face stressor. For example, early social interaction with mothers have been found to have different consequences in boys and girls and boys were found to be more affected by maternal sensitivity in the still-face procedure than girls (Warren & Simmens, 2005; Weinberg, Tronick, Cohn, & Olson, 1999). As a result, maternal sensitivity might be more strongly associated with vagal reactivity in males than in females. A general concern regarding studies reporting statistical interactions is that they run an elevated risk of false positives. This is particularly the case where multiple exploratory analyses of interactions between predictor variables are conducted but less of a concern in studies which, as this study did, examines interactions with sex of infant in the light of the available literature with a hypothesised direction of effect. We sought to minimise the risk further by modelling the outcome as a latent variable requiring contributions from associations with oppositional defiant disorder symptoms across three-time points spanning 2.5 years.

Finally, although this is among the largest studies examining the impact of early vagal reactivity, our power to discriminate the detailed differences in developmental processes for boys and girls was limited. Further studies with longer follow-ups are required to exclude the possibility of boys preferentially responding on the headstrong dimension and girls on the irritability dimension.
3.5 Conclusion

Physiological reactivity to a stressful situation differently predicts ODD symptoms in boys and girls very early in life. Although ODD symptoms seemed to be multidimensional at early ages, the association of vagal reactivity with specific ODD dimensions, irritability in girls and headstrong in boys, was not evident. Nevertheless, the opposite direction in the association with overall ODD symptoms might contribute to our understanding of the several mechanisms involved in the later development of distinct psychiatric disorders in boys and girls. In the future, large longitudinal studies should test whether vagal reactivity to stress measured in infancy is predictive of distinct psychiatric disorders in boys and girls after puberty, when the prevalence rates of these disorders typically arise.
CHAPTER 4: Association between deficits in emotion recognition and depressive symptoms in youth with DMDD

4.1 Background

Evidence from individual studies and Chapter 2 of this thesis suggest that youth with chronic irritability have a substantially increased risk for future depression (Brotman et al., 2006; Stringaris et al., 2009). Our meta-analysis found an odds ratio of 1.85 for the longitudinal association between irritability and depressive disorder (Vidal-Ribas et al., 2016). The overlap between these phenotypes seems to have a genetic origin, as suggested by family (Krieger et al., 2013; Propper et al., 2017) and twin studies (Savage et al., 2015; Stringaris et al., 2012b), and described in section 1.6.1. However, other potential mechanisms involved in this developmental transition have not been tested. Several studies have implicated deficits in emotional processing, including deficits in emotion recognition, as well as attentional and interpretation biases, as a risk factor for mood disorders. Thus, such deficits might be a plausible mechanism for the transition between irritability and depression (Bourke, Douglas, & Porter, 2010; Brotman et al., 2017). In the current study, I use a carefully characterised sample of youth with chronic severe irritability, defined here as DMDD, to test whether deficits in emotion recognition in faces and voices may partially increase the risk for depression.

4.1.1 Emotion processing deficits in irritability and depression

Emotion recognition can be defined as the ability to identify emotions expressed by facial and vocal stimuli. This ability plays a crucial role in the development of social competence and interpersonal well-being (Halberstadt, Denham, & Dunsmore, 2001). Research shows
(see section 1.6.2.2) that youth with chronic severe irritability present deficits in emotion recognition of faces (Guyer et al., 2007; Kim et al., 2013; Rich et al., 2008) and voices (Deveney et al., 2012). Similarly, a recent meta-analysis found that depressed patients have deficits in the recognition of facial emotions compared to healthy volunteers (Dalili et al., 2015), which is consistent with previous reports (Demenescu, Kortekaas, den Boer, & Aleman, 2010; Kohler, Hoffman, Eastman, Healey, & Moberg, 2011). Specifically, depressed patients seem to correctly identify sad faces, but they fail to recognise all other basic facial emotions. Depression has also been associated with deficits in the recognition of emotion in voices (Kan et al., 2004). For example, depressed participants tend to rate voices with negative emotional prosody (i.e., sad, anger, fear) more intensely than healthy volunteers (Naranjo et al., 2011). In addition, they are more likely to rate happy emotional prosody as fearful or sad (Peron et al., 2011). In summary, both youth with chronic severe irritability and youth with depression display deficits in emotion recognition. However, no studies have tested whether these deficits increase the risk for depression in youth with chronic severe irritability.

In addition to deficits in emotion recognition, both chronic severe irritability and depression are associated with other emotion processing deficits, such as interpretation and attentional biases of affective stimuli. As described in section 1.6.2.2, one of the pathophysiological mechanisms of irritability is the aberrant approach towards threatening stimuli (Brotman et al., 2017). In youth with chronic severe irritability, this aberrant approach towards threat is evidenced as an attentional bias towards threatening faces (Hommer et al., 2014; Salum et al., 2017) and as the perception of threat during the viewing of neutral and ambiguous faces (Brotman et al., 2010; Stoddard et al., 2016). Similarly, the attention of depressed patients
is also drawn toward negative facial emotions, especially sadness (Armstrong & Olatunji, 2012; Gotlib, Krasnoperova, Yue, & Joormann, 2004; Leppanen, 2006; Peckham et al., 2010), and people with depression are more likely to interpret neutral and ambiguous faces as sadder or less happy (Leppanen, 2006).

4.1.2 Other factors contributing to emotion recognition: the case of anxiety, age, and sex

If we want to examine how deficits in emotion recognition in youth with irritability relate to depression, we also need to test other factors that might contribute to differences in emotion processing by either testing the specificity of the predictors or by exploring the possible moderating factors that influence the association. In the current study, we examine the effects of anxiety, age and sex on emotion recognition in the presence of depressive symptoms.

Irritability and depression are highly comorbid with anxiety disorders (Althoff et al., 2016; Axelson & Birmaher, 2001; Copeland et al., 2013). As is the case with irritability and depression, anxiety disorders in young people are also associated with deficits in emotional processing (Armstrong & Olatunji, 2012; Bar-Haim et al., 2007). However, recent data suggest that emotional processing deficits in youth with chronic irritability might be independent of co-occurring anxiety symptoms. Specifically, a recent study found that the association between irritability and attentional bias towards threat was still significant even after accounting for the presence of anxiety symptoms, as well as symptoms of ADHD, CD, and ODD. However, this association became non-significant when adjusting for the broad internalising symptoms domain of the CBCL, comprising withdrawing symptoms,
somatic complaints, and anxious/depressed symptoms, which mainly relates to depressive symptomatology (Salum et al., 2017). It is unclear from the study whether the association between irritability and other deficits in emotional processing ability, such as emotion recognition, is also specifically moderated by depressive symptoms and not anxiety.

Another factor associated with variations in emotion recognition is age. The ability to recognise emotions is known to improve from childhood through adolescence and adulthood (Brosgole & Weisman, 1995; Chronaki, Hadwin, Garner, Maurage, & Sonuga-Barke, 2015) with no differences between adolescents and adults (Brosgole & Weisman, 1995). Also, it is well-known that most cases of depression have the onset during adolescence (Kessler et al., 2007). Thus, it is possible that reaching the adolescence period with underdeveloped emotional processing skills might be associated with an increased risk for depression. Indeed, several studies show that, compared with healthy controls, people at risk for depression (e.g. healthy offspring of depressed parents) present more impairments in emotional processing; in other words, deficits in emotional processing might be an endophenotype of depression (Chan, Norbury, Goodwin, & Harmer, 2009; Joormann, Talbot, & Gotlib, 2007; Mannie, Taylor, Harmer, Cowen, & Norbury, 2011; Monk et al., 2008). Moreover, emotional processing deficits are predictive of depression and depressive symptoms in prospective studies (Beevers & Carver, 2003; Vrijen et al., 2016). Since irritability is a strong predictor of depression (Vidal-Ribas et al., 2016), it is plausible that a deficit in emotion recognition in children with severe irritability contributes to increasing their risk for depression. Identifying such deficits as early contributors to depression may inform the development of novel targeted treatment interventions aimed at preventing the development of depression in at-risk children, such as those youth with severe irritability.
Finally, several studies have reported sex differences in emotion recognition. Specifically, most reports suggest that females are better at recognising facial and vocal emotions than males (Kret & De Gelder, 2012; McClure, 2000; Thompson & Voyer, 2014). Both adolescent (Lee et al., 2013) and adult (Wingenbach, Ashwin, & Brosnan, 2018) females perform facial-emotion recognition tasks at higher accuracy than age-matched males. Most studies show that females are better at recognising all emotions; however, some evidence suggests that males might be better in recognising threatening stimuli (Kret & De Gelder, 2012). As discussed in Chapter 3, adolescent girls are at higher risk of depression than boys (Angold & Rutter, 1992; Lewinsohn et al., 1993), and they might show higher rates of irritability than boys (Riglin et al., 2017; Stringaris et al., 2012b). As mentioned before, deficits in emotion processing have been associated with an increased risk of depression, which commonly emerges during adolescence. Therefore, it is possible that sex differences in emotion processing might be associated with differences in the prevalence of psychiatric disorders in this age period. This remains to be tested in youth with irritability.

4.1.3 Aims and hypotheses

In this study, we test the mechanistic role of deficits in emotion recognition in a clinical sample of youth with severe irritability using a well-validated paradigm of facial and vocal emotion recognition.

We have two primary aims. First, we aim to examine how emotion recognition varies in youth with chronic severe irritability with different levels of depressive symptoms. Based on previous reports, we hypothesise that youth with DMDD and high depressive symptoms will be more likely to interpret positive stimuli (i.e. happy faces and voices) as more
negative (i.e., sad, fearful and angry) than those with DMDD and low depressive symptoms, and healthy controls (HC). Second, we aim to test whether deficits in emotion recognition increase the likelihood of depressive symptoms at follow-up. We hypothesise that the misidentification of positive stimuli as negative will be longitudinally associated with higher depressive symptoms over and above any baseline depressive symptomatology.

Our secondary aims include examining the effects of comorbid anxiety disorders, age, and sex on emotion recognition in youth with DMDD and depressive symptoms. We hypothesise that deficits in emotion recognition will be independent of comorbid anxiety disorders in youth with DMDD. In addition, we hypothesise that older youth and girls will be better at emotion recognition than younger youth and boys, respectively.

4.2 Methods

4.2.1 Participants

Participants were 8–20 years old (Mean=13.3, SD=2.8; N=116; n=41 females, 35%; n=75 males, 65%) and included youth with DMDD (n = 77) and HC youths (n = 39). Youth and parents gave written informed assent and consent to participate. This study was approved by the National Institute of Mental Health (NIMH) Institutional Review Board.

Participants were assessed by master’s- or doctoral-level clinicians using the Schedule for Affective Disorders and Schizophrenia for School-Age Children–Present and Lifetime Version (K-SADS–PL) (Kaufman et al., 1997), with a supplemental module to ascertain DMDD. Depressive symptoms in patients with DMDD were assessed by self-report with the Children’s Depression Inventory (CDI) (Kovacs, 1992) while irritability was measured in DMDD and HC groups using the Affective Reactivity Index (ARI) (Stringaris et al.,
General functioning was measured with the Children’s Global Assessment Scale (CGAS) (Shaffer et al., 1983).

To explore the effect of depressive symptoms, we divided the group of youth with DMDD into low (DMDD/LD; \( n = 52 \)) versus high (DMDD/HD; \( n = 25 \)) depressive symptoms based on the cut-off for clinical samples (i.e. 13 points) suggested in the CDI manual (Kovacs, 1992).

A third of the DMDD participants (\( n = 25 \), 33%; DMDD/LD \( n = 19 \); DMDD/HD \( n = 6 \)) had measures of depressive symptoms at follow-up (time between assessments, Mean=1.04 years, SD=0.5, range=0.3-2 years). However, attrition analyses showed no difference in age, sex, race, ethnicity, ARI score, CDI score, or task performance between those participants who had data at follow-up and those participants who had no data.

Control subjects were psychiatrically healthy, based on the K-SADS–PL, and had no first-degree relatives with mood disorders. Exclusion criteria for all groups included IQ <70, as measured with the Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999), pervasive developmental disorder, unstable medical illness, or substance abuse within the past 2 months.

### 4.2.2 Emotion Recognition Task

To examine differences in emotion recognition between all three groups (i.e., DMDD/HD, DMDD/LD, and HC) we used the Diagnostic Analysis of Nonverbal Accuracy (DANVA–2) (Baum & Nowicki, 1998; Rothman & Nowicki, 2004) In this study, we used both the facial expressions subtest and the paralanguage subtest.
The facial expressions subtest includes standardised photographs of children (n=24) and adults (n=24) displaying an expression of happiness, sadness, anger, or fear. After viewing the photograph for 2 seconds, participants indicated by button-press which emotion was expressed. People shown in the photographs were majority white Caucasians (children 92%, adults 79%). However, participants’ ethnicity was not associated with emotion recognition.

Similarly, the paralanguage subtest features a standardised set of recordings of voices of children (n = 24) and adult (n = 24) actors repeating a neutral phrase (e.g. I’m going out of the room now, but I’ll be back later) in happy, sad, angry, or fearful tones. Using a button-press, participants indicated which emotion the actor expressed.

Primary outcome variables included (1) the number of errors in emotion identification in child and adult faces and voices, and (2) the number of each type of misidentification (e.g., number of times that a happy face is rated as angry or sad).

4.2.3 Statistical analysis

4.2.3.1 Sample characteristics

We examined group differences in age, IQ and general functioning using analyses of variance (ANOVA); and differences in sex distribution, ethnicity, and rates of comorbid diagnoses were examined with chi-square tests. Age and sex distribution differed between groups (see section 4.3.1 Sample characteristics); therefore, we decided to include these as covariates in subsequent analyses.
4.2.3.2 Categorical analysis

A Group (DMDD/HD, DMDD/LD, HC) × Emotion (happy, sad, angry, fearful) × Actor (child, adult) × Modality (face, voice) repeated measures analysis of covariance (ANCOVA) with a Greenhouse-Geisser correction, examined the number of errors per emotion category, and subsequent misidentification of emotions. Bonferroni-corrected post hoc analyses were used to examine pairwise comparisons. Partial eta-squared is reported for effect size (small=.01, medium=.06, large=.14) (Cohen, 1988).

4.2.3.3 Dimensional analysis

We aimed to replicate the above categorical analysis using depressive symptoms as a dimensional measure in post hoc analyses in the patient group only (as mentioned in section 4.2.1, CDI was not collected on HC). Then, we performed a CDI × Emotion (happy, sad, angry, fearful) × Actor (child, adult) × Modality (face, voice) repeated measures analysis of covariance (ANCOVA).

4.2.3.4 Effects of comorbid anxiety disorders on emotion recognition

To examine the effects of anxiety disorders, we replicated the categorical analyses in the DMDD participants using a repeated measures ANCOVA with a four-level group (DMDD/LD n=24 vs DMDD/LD with anxiety disorder (+ANX) n=28 vs DMDD/HD n=12 vs DMDD/HD+ANX n=13).

4.2.3.5 Effects of age on emotion recognition

Given the wide range of ages in the current sample (8-20 years old), we examined the effects of age on emotion recognition by (1) performing pairwise correlations between age and number of errors in emotion recognition, and (2) re-running the categorical analyses in
the DMDD participants using a repeated measures ANCOVA with a four-level group (DMDD/LD children n=19 vs DMDD/LD adolescents n=33 vs DMDD/HD children n=14 vs DMDD/HD adolescents n=11). Children with DMDD included participants aged 8-11 years (n=33, 43%), whereas adolescents with DMDD included participants aged 12-20 years (n=44, 57%).

4.2.3.6 Effects of sex on emotion recognition

The effects of sex on emotion recognition were first examined in the whole sample using repeated measures ANCOVA with sex as between-subjects factor. That is, a model including Group (female, male) × Emotion (happy, sad, angry, fearful) × Actor (child, adult) × Modality (face, voice) with age as covariate. In post hoc analyses, independent sample t-tests were used to examine pairwise comparisons between males (n=75) and females (n=41). Second, as we did with age, we re-ran the categorical analyses in the DMDD participants using a repeated measures ANCOVA with a four-level group (DMDD/LD female n=13 vs DMDD/LD male n=39 vs DMDD/HD female n=13 vs DMDD/HD male n=12).

4.2.3.7 Longitudinal analysis

Finally, to test whether emotion recognition accuracy predicted depressive symptoms at follow-up we used linear regression analyses with depressive symptoms at follow-up as outcome, and emotion recognition accuracy as the predictor of interest, adjusting for age, sex, number of days between assessments, and baseline depressive symptoms. Of note, the longitudinal analysis only used a subset of DMDD participants (n=25, 33%) due to attrition.
4.3 Results

4.3.1 Sample characteristics

Groups differed in age ($F[2, 113]=12.81, p<0.0001$) and sex distribution ($\chi^2[2, 114]=5.6, p=0.060$). Specifically, the mean age of healthy volunteers was higher than in both DMDD groups; and the DMDD/LD group had fewer females when compared with the other two groups. However, no differences were found in ethnicity, rates of comorbid diagnoses, prescribed medication, IQ, or general functioning (Table 4.1).

4.3.2 Categorical analysis

A significant main effect indicated that groups differed in emotion recognition errors overall, $F(2,111) = 5.19, p = 0.007, \eta^2 = 0.09$. Bonferroni-corrected post hoc analyses revealed that DMDD/HD youth made more errors than DMDD/LD ($p = 0.010$) and HC ($p = 0.017$), with no differences between DMDD/LD and HC.

While there were no significant four-way interactions, a significant three-way interaction emerged involving Group, Emotion and Modality, $F(6,331.7)=3.95, p = 0.001, \eta^2 = 0.07$. Post hoc analyses revealed that DMDD/HD youth made more errors identifying happy faces and voices than DMDD/LD (happy faces: $p = 0.004$; happy voices: $p = 0.001$), and HC (happy faces: $p = 0.028$; happy voices: $p = 0.001$) with no differences between the latter (Figure 4.1).

Specifically, DMDD/HD youth were more likely to interpret happy stimuli as angry and fearful compared to DMDD/LD (happy as angry: $p = 0.018$; happy as fearful: $p = 0.008$) and HC (happy as angry: $p = 0.014$; happy as fearful: $p = 0.024$). There were no differences in the misidentification of happy stimuli as sad between groups (Figure 4.2).
Table 4.1. Demographic and clinical characteristics

<table>
<thead>
<tr>
<th></th>
<th>HC (n=39)</th>
<th>DMDD/LD (n=52)</th>
<th>DMDD/HD (n=25)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
</tr>
<tr>
<td><strong>Females</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>38</td>
<td>13</td>
</tr>
<tr>
<td><strong>Ethnicity a</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>24</td>
<td>65</td>
<td>39</td>
</tr>
<tr>
<td>Hispanic</td>
<td>2</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>African American</td>
<td>7</td>
<td>19</td>
<td>1</td>
</tr>
<tr>
<td>Asian</td>
<td>2</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Multiple races</td>
<td>2</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td><strong>Medication</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>38</td>
</tr>
<tr>
<td><strong>Comorbid diagnoses</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADHD</td>
<td>-</td>
<td>-</td>
<td>36</td>
</tr>
<tr>
<td>ODD</td>
<td>-</td>
<td>-</td>
<td>30</td>
</tr>
<tr>
<td>CD</td>
<td>-</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>GAD</td>
<td>-</td>
<td>-</td>
<td>16</td>
</tr>
<tr>
<td>SAD</td>
<td>-</td>
<td>-</td>
<td>14</td>
</tr>
<tr>
<td>SOC</td>
<td>-</td>
<td>-</td>
<td>7</td>
</tr>
<tr>
<td>Any mood</td>
<td>-</td>
<td>-</td>
<td>6</td>
</tr>
<tr>
<td>Any externalising</td>
<td>-</td>
<td>-</td>
<td>40</td>
</tr>
<tr>
<td>Any anxiety</td>
<td>-</td>
<td>-</td>
<td>28</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>14.9</td>
<td>2.7</td>
<td>12.7</td>
</tr>
<tr>
<td><strong>IQ b</strong></td>
<td>112.9</td>
<td>11.3</td>
<td>112.7</td>
</tr>
<tr>
<td><strong>CGAS c</strong></td>
<td>-</td>
<td>-</td>
<td>47.4</td>
</tr>
<tr>
<td><strong>ARI d</strong></td>
<td>1.3</td>
<td>2.1</td>
<td>4.1</td>
</tr>
<tr>
<td><strong>CDI</strong></td>
<td>-</td>
<td>-</td>
<td>5.9</td>
</tr>
</tbody>
</table>

HC, healthy controls; DMDD, disruptive mood dysregulation disorder; ADHD, attention deficit hyperactivity disorder; ODD, oppositional defiant disorder; Any mood, includes major depressive disorder, dysthymia, and mood disorder not otherwise specified; GAD, generalised anxiety disorder; SAD, separation anxiety disorder; SOC, social anxiety disorder; Any anxiety, includes GAD, SAD and SOC.

a Ethnicity information was missing for n=2 HC

b Intelligence quotient (IQ) information was available for n=18 HC, n=31 DMDD/LD, and n=17 DMDD/HD.

c Children’s Global Assessment Scale (CGAS) information was available for n=34 DMDD/LD, and n=19 DMDD/HD.

d Affective Reactivity Index (ARI) information was available for n=29 HC, n=46 DMDD/LD, and n=24 DMDD/HD.
Figure 4.1. Number of identification errors by emotion and modality across groups. Bars are 95% confidence intervals. Means adjusted for age and sex, and Bonferroni-corrected for multiple comparisons. *p<0.05, **p<0.01, ***p<0.005
Figure 4.2. Misidentification of happy stimuli as other emotions across groups. Bars are 95% confidence intervals. Means adjusted for age and sex, and Bonferroni-corrected for multiple comparisons. *p<0.05, **p<0.01, ***p<0.005

Additionally, whereas DMDD/HD youth made more errors in the recognition of angry faces than HC (p=0.010), particularly those from child actors (p=0.028), the opposite was true for voices; that is, DMDD/HD youth were better at recognising angry voices than HC (p=0.023) (Figure 4.1), and particularly those from adult actors (p=0.036).

Post hoc analyses showed that DMDD/HD youth misidentified angry faces as sad more frequently than HC (p=0.016). On the other hand, HC youth misidentified angry voices as fearful (p=0.015) and happy (p=0.045) more frequently than DMDD/HD youth (Figure 4.3).
We then examined whether being under prescribed medication would have an effect in task performance. Within the DMDD participants, there were no differences between unmedicated and medicated individuals in task performance (all $p>0.30$). Furthermore, adding medication status as a covariate did not alter the results.

**Figure 4.3. Misidentification of angry stimuli as other emotions across groups.** Bars are 95% confidence intervals. Means adjusted for age and sex, and Bonferroni-corrected for multiple comparisons. *$p<0.05$, **$p<0.01$, ***$p<0.005$

Finally, six participants within the DMDD/LD group had a mood disorder (either depressive disorder, dysthymia, or mood disorder not otherwise specified). We then re-run the analyses after removing these 6 participants but the results remained unchanged.
### 4.3.3 Dimensional analysis

The analysis with the CDI, our dimensional measure of depressive symptoms collected in youth with DMDD, yielded results that were similar to those from the main categorical analyses.

There was a main effect of CDI, $F(1,73) = 7.73, p = 0.007, \eta^2 = 0.10$, suggesting that depressive symptoms were associated with more errors overall in the DMDD group. As in the categorical analyses, a significant three-way interaction emerged involving CDI, Emotion, and Modality, $F(2.9,215) = 3.68, p = 0.014, \eta^2 = 0.05$. Linear regression analyses showed that in youth with DMDD, depressive symptoms were significantly associated with the number of identification errors in happy faces ($\beta=0.31, p=0.011$), happy voices ($\beta=0.41$, $p<0.001$), and angry faces ($\beta=0.26$, $p=0.039$). Although irritability was significantly correlated with depressive symptoms ($r=0.51, p<0.001$), these results were also true in models adjusted for irritability symptoms (happy faces: $\beta=0.32, p=0.029$; happy voices: $\beta=0.35, p=0.014$; angry faces: $\beta=0.30, p=0.045$). Of note, irritability itself did not predict accuracy in these analyses. Furthermore, in models adjusted for clinical impairment as measured with CGAS, impairment was not associated with accuracy, but depressive symptoms were (happy faces: $\beta=0.30, p=0.045$; happy voices: $\beta=0.39, p=0.007$; angry faces: $\beta=0.37, p=0.013$).

Also, in line with the categorical analyses, in youth with DMDD depressive symptoms were associated with the misidentification of happy stimuli as fearful ($\beta=0.38, p=0.001$) and angry ($\beta=0.25, p=0.041$), but not sad stimuli ($\beta=0.12, p=0.336$). However, after
controlling for co-occurring irritability symptoms, only the association with misidentification of happy stimuli as fearful remained significant ($\beta=0.38$, $p=0.006$).

The analyses examining the effects of prescribed medications showed that neither depressive symptoms nor irritability symptoms were associated with medication status (all $p>0.35$) and adding medication as a covariate in the regression analyses did not alter results.

### 4.3.4 Effects of comorbid anxiety disorders on emotion recognition

There were no differences in rates of anxiety disorders between DMDD/LD and DMDD/HD, $\chi^2(2) = 0.02$, $p = 0.879$ (Table 4.1), nor differences in CDI scores between those DMDD youth with anxiety disorders and those DMDD youth without anxiety disorders ($t(75) = 0.025$, $p =0.980$). Furthermore, having an anxiety disorder by itself was not associated with errors in the recognition of emotional stimuli in any modality (all $p>0.20$). However, a repeated measures ANOVA with a four-level group (DMDD/LD n=24, DMDD/LD+ANX n=28, DMDD/HD n=12, DMDD/HD+ANX n=13), revealed a two-way interaction involving Group and Emotion ($F(8,190.4) = 3.49$, $p = 0.040$, $\eta^2 = 0.08$). After controlling for multiple comparisons, post hoc analyses showed that youths in the DMDD/HD+ANX group made more errors in recognising happy stimuli than DMDD/LD ($p<0.001$), DMDD/LD+ANX ($p<0.001$), and DMDD/HD groups ($p=0.015$). Specifically, youth in the DMDD/HD+ANX group were more likely to misidentify happy stimuli as fearful than DMDD/LD ($p=0.005$), DMDD/LD+ANX ($p<0.001$) and DMDD/HD youth ($p=0.003$).
4.3.5 Effects of age on emotion recognition

In the whole sample (N=116) age was negatively associated with number of errors in emotion recognition, overall (r=-0.36, p=0.0001) as well as for happy (r=-0.27, p=0.004), angry (r=-0.21, p=0.02), fearful (r=-0.24, p=0.009) and sad stimuli (r=-0.20, p=0.03). In other words, the older the participants were the better they recognised emotional stimuli.

In participants with DMDD, age was correlated neither with depressive symptoms (r=-0.08, p=0.503) nor with irritability symptoms (r=-0.23, p=0.054), although the latter was a nonsignificant trend.

In a repeated measures ANOVA with a four-level group (DMDD/LD children n=19, DMDD/LD adolescents n=33, DMDD/HD children n=14, DMDD/HD adolescents n=11), a two-way interaction emerged involving Group and Emotion (F(8,191.3) = 3.37, p = 0.001, η² = 0.12). After controlling for multiple comparisons, post hoc analyses showed that both children (p=0.002) and adolescents with DMDD/HD (p=0.004) made more errors in recognising happy stimuli than DMDD/LD adolescents. In addition, DMDD/HD children made more errors in recognising fearful stimuli than DMDD/LD children (p=0.024) and DMDD/LD adolescents (p=0.028). Finally, within the DMDD/LD group, children made more errors in recognising sad stimuli than adolescents (p=0.025).

4.3.6 Effect of sex on emotion recognition

Analyses in the whole sample with repeated measures ANCOVA revealed a main effect of sex (F(1, 30.9) = 11.6, p = 0.001, η² = 0.09). Overall, males made more errors than females (t( 114)=-3.39, p=0.001). A four-way interaction emerged involving Emotion, Actor, Modality and sex (F(2.9, 330.2) = 3.13, p = 0.027, η² = 0.03). Post hoc analyses revealed
that male made more errors in recognising child fearful faces \((t(114)=-3.23, p=0.0016)\), as well as child and adult fearful voices (child voices: \(t(114)=-3.40, p=0.0009\); adult voices: \((t(114)=-4.14, p=0.0001)\). Males also made more errors than females in recognising child sad voices \((t(114)=-2.49, p=0.014)\).

In participants with DMDD, females had higher levels of depressive symptoms \((t(75)=-3.07, p=0.003)\) and irritability symptoms \((t(68)=2.05, p=0.044)\) than males.

In a repeated measures ANOVA with a four-level group (DMDD/LD female \(n=13\) vs DMDD/LD male \(n=39\) vs DMDD/HD female \(n=13\) vs DMDD/HD male \(n=12\)), a significant three-way interaction emerged involving Group, Emotion and Modality, \(F(8.8,211.8) = 2.15, p = 0.028, \eta^2 = 0.08\). Post hoc analyses revealed that DMDD/HD males made more errors in recognising happy voices than DMDD/LD males \((p=0.015)\) and DMDD/LD females \((p=0.033)\). DMDD/HD males also made more errors in recognising fearful faces than DMDD/LD females \((p=0.035)\). Interestingly, both DMDD/HD males and DMDD/LD males made more errors in recognising fearful voices than DMDD/HD females (vs DMDD/HD males, \(p=0.012\); vs DMDD/LD males, \(p=0.043\)) and DMDD/HD females (vs DMDD/HD males, \(p=0.009\); vs DMDD/LD males, \(p=0.025\)).

### 4.3.7 Longitudinal analysis

We used significant associations at baseline as predictors of interest for future depressive symptoms in a subset of DMDD participants. Given the high level of attrition (66%) these analyses are exploratory, and the results should be interpreted with caution. All results mentioned below are adjusted for age, sex, days between assessments, and baseline depressive symptoms.
In youth with DMDD, the misidentification of happy stimuli as fearful was associated with higher depressive symptoms at follow-up ($\beta=0.41$, $p=0.016$), even when adjusting for baseline irritability symptoms ($\beta=0.43$, $p=0.017$). Misidentification of happy stimuli as angry, or angry faces as sad was not associated with future depressive symptoms ($\beta=-0.20$, $p=0.302$; $\beta=-0.12$, $p=0.441$, respectively).

By contrast, misidentifying angry voices as fearful was associated with higher depressive symptoms at follow-up ($\beta=0.34$, $p=0.046$); a finding that became a trend after controlling for irritability symptoms at baseline ($\beta=0.34$, $p=0.051$).

4.4 Discussion

4.4.1 Summary of findings

To our knowledge, this is the first study to examine a behavioural deficit (i.e., emotion recognition) as a potential mechanism for explaining the association between irritability and depression (Vidal-Ribas et al., 2016).

Cross-sectionally, we have found that depressive symptoms in youth with DMDD were mainly associated with impaired recognition of happy stimuli across modalities (i.e., faces and voices). Specifically, DMDD youth with higher levels of depressive symptoms were more likely to misinterpret happy stimuli as fearful and angry, independent of current irritability symptoms.

Longitudinally, we found that deficits in emotion recognition predicted higher levels of depressive symptoms at follow-up approximately one year later, independent of baseline depressive and irritability symptoms.
In addition, comorbid anxiety disorder was not associated with deficits in emotion recognition unless high levels of depressive symptoms were present. As expected, the ability to recognise facial and vocal emotions improved with age and was higher in females.

4.4.2 Deficits in emotion recognition in youth with severe irritability are associated with depressive symptoms

The findings of the current study suggest that deficits in emotion recognition in youth with DMDD are associated with the presence of depressive symptoms rather than irritability symptoms. Emotion recognition accuracy in healthy volunteers only differed from those DMDD youth with high depressive symptoms, but not from DMDD youth with low levels of depressive symptoms. Furthermore, we also found that the association between CDI scores and emotion recognition was significant even after accounting for current irritability symptoms, with the latter not being associated with emotion recognition accuracy.

These findings are in contrast with previous studies which have found deficits in emotion recognition in youth with severe irritability relative to healthy volunteers (Deveney et al., 2012; Guyer et al., 2007; Kim et al., 2013; Rich et al., 2008). The discrepancy from previous studies could be explained by the different definition of chronic severe irritability. That is, previous works were conducted in youth with SMD which differs slightly from DMDD (See Table 1.1 in Chapter 1). Specifically, the definition of SMD includes hyperarousal symptoms and states that the abnormal mood between outbursts can be that of anger or sadness. Therefore, although most cases present with angry mood, it is unclear whether the effects on emotion recognition were explained by angry or sad mood.
The differences from previous literature could also be explained by the distinct approaches for assessing emotion recognition (Kim et al., 2013; Rich et al., 2008). For example, Kim et al. (2013) employed black and white photographs of facial emotions taken from the Picture of Facial Affect set (Ekman & Friesen, 1976) with different levels of emotional intensity. In addition, in Kim’s study, the primary outcome was fixation as measured with eye-tracking instead of emotional labelling accuracy. Our approach is also distinct from Rich and colleagues (2008) who used an emotional multimorph task in which each facial emotion morphed gradually from neutral to emotional through 39 incremental stages, with the aim of assessing the intensity of emotion required before participants correctly identified the emotion displayed.

In addition, in the current study, we decided to be conservative and apply Bonferroni correction for multiple comparisons in post-hoc analyses. In previous studies, differences in emotion recognition between SMD and HC were found using less conservative analyses (Deveney et al., 2012; Guyer et al., 2007) and became non-significant when corrections were applied (Guyer et al., 2007). Another reason why previous studies did not report associations with depressed mood was simply because the effects of depressive symptoms on emotion recognition were not tested (Kim et al., 2013) or could not be tested due to low rates of depressed mood (Guyer et al., 2007; Kim et al., 2013; Rich et al., 2008).

Further to the finding that the deficits in emotion recognition were associated with depressive symptoms in youth with DMDD, our longitudinal analysis showed that regardless of current depressive symptoms, emotion recognition deficits in youth with DMDD were predictive of more depressive symptoms at the follow-up. It is important to
Note that most DMDD youth that were included in the longitudinal analyses (n=19, 76%) were part of the DMDD group with low depressive symptoms at baseline, which emphasises the role of deficits in emotion recognition as a risk for increased depressive symptoms. The findings from the longitudinal analysis should be interpreted with caution since levels of attrition were high; however, these suggest that emotion recognition deficits act as a risk factor for depression in youth with DMDD, in the same way it does in healthy youth (Vrijen et al., 2016). This conclusion would be strengthened if these findings were replicated in a larger sample.

Unexpectedly, compared to HC, youth with DMDD and high depressive symptoms had deficits in recognising angry faces, especially those of children, but were better in recognising angry voices, especially those of adults. It should be noted that in the current study all emotions were more difficult to identify in voice modality than facial modality, except for anger; DMDD participants found it easier to identify angry voices than angry faces, and such distinction was not observed in HC. The reason why DMDD youth with depressive symptoms were better at recognising angry voices but worse at recognising angry faces is unclear. It is possible that this study is unpowered to accurately identify interactional effects. Another plausible explanation is that, on the one hand, youth with DMDD are frequently involved in arguments with their parents and peers, hence frequently exposed to angry stimuli (making these easier to be recognised). However, depressive symptoms might lead to social withdrawal, loss of contact with peers, isolation within the home environment (e.g., spending more time alone in their bedrooms), and more conflicts at home. Consequently, attention might be drawn inward - thus relying on voices to identify emotion - and negative thoughts about not being liked by others may emerge. The
combination of these changes with the frequent exposure to angry voices from parents, along with the decreased exposure to the angry faces from friends, might explain this finding. Nonetheless, these results would need replication in a larger sample and should be supported by collecting information on the change in social relationships due to depressive symptomatology.

4.4.3 Underlying mechanisms that might explain the association between deficits in emotion processing, irritability and depression

Here we have examined a behavioural deficit (i.e., deficits in emotion recognition) to try to explain the specific association between irritability and depression. However, our findings need to be considered in the larger context of pathophysiological mechanisms underlying deficits in emotion processing, including neural and genetic factors, that might be shared between irritability and depression.

Functional magnetic resonance imaging (fMRI) paradigms examining face emotion processing can probe conscious processing of face emotions (i.e., unmasked faces presented ≥40ms) or non-conscious (i.e. masked faces presented <40ms). Unmasked processing can be implicit, in which research participants focus on a stimulus feature other than the face emotion (e.g., reporting nose width or gender), or explicit, in which the task directs attention toward the face emotion (e.g. rating fear or hostility). No fMRI studies have directly compared irritability and depression in their neural correlates during emotion processing. However, evidence from independent fMRI studies has shown that both irritability and depression are associated with dysfunctional activity and connectivity of fronto-limbic regions, including amygdala and orbitofrontal cortex, during explicit
(Almeida et al., 2009; Beesdo et al., 2009; Brotman et al., 2010; Carballedo et al., 2011; Coccaro, McCloskey, Fitzgerald, & Phan, 2007; Gaffrey, Barch, Singer, Shenoy, & Luby, 2013; Hall et al., 2014; Keedwell, Andrew, Williams, Brammer, & Phillips, 2005; Lee et al., 2008; Peluso et al., 2009; Wiggins et al., 2016; Yang et al., 2010; Zhong et al., 2011) and implicit (Fu et al., 2008; Stoddard et al., 2017; Thomas et al., 2013) processing of unmasked facial emotions, and also during processing of masked faces (Suslow et al., 2010; Tseng et al., 2016; Victor, Furey, Fromm, Ohman, & Drevets, 2010). Of note, dysfunctional fronto-limbic response during explicit (Chai et al., 2015; Mannie et al., 2011; Monk et al., 2008) and implicit processing of emotional faces (Goulden et al., 2012; Monk et al., 2008) has also been found in remitted depressed patients and people at high risk for depression. These findings suggest that aberrant neural responses underlying the processing of emotional stimuli might also be a risk factor for depression in both healthy youth and in those with severe irritability. Future prospective neuroimaging studies should examine to what extent neural responses to emotional stimuli in youth with DMDD are predictive of depression.

As described in section 1.6.1 in Chapter 1, the longitudinal overlap between irritability and depression is mainly explained by genetic factors (Savage et al., 2015; Stringaris et al., 2012b). Therefore, the shared genetic variance between both phenotypes may also underlie shared deficits in emotional processing. Indeed, behavioural and neurobiological deficits of emotional processing seem to be partially heritable. Twin studies show heritability estimates to be around 0.50 for event-related brain potential (ERP) responses to emotional stimuli (Anokhin, Golosheykin, & Heath, 2010; Weinberg, Venables, Proudfit, & Patrick, 2014). To our knowledge, there are no published twin studies examining the genetic
association between emotional processing and depression or irritability. However, preliminary work in a large community-based twin sample suggests a moderate genetic correlation ($r = -0.26$, $p<0.001$) between emotion recognition and irritability, and especially with happy stimuli (Rappaport et al., 2018b). Future work should expand these analyses and examine the degree to which associations among emotion recognition, irritability and depression can be explained by shared genetic factors.

4.4.4 Anxiety is not directly related with a deficit in emotion recognition in youth with DMDD

Our finding that DMDD youth misperceive happiness as fear or anger is consistent with previous studies showing misinterpretation of neutral or ambiguous stimuli as threat in both youths with severe irritability (Brotman et al., 2010; Stoddard et al., 2016) and people with depression (Leppanen, 2006; Peckham et al., 2010). However, the misidentification of happy stimuli was not associated with anxiety disorders in youth with DMDD unless they also had high levels of depressive symptoms. While the evidence for attentional bias toward threat in children with anxiety disorders is strong (Bar-Haim et al., 2007), previous studies suggest that depression in youth might be more associated with deficits in emotion recognition than anxiety (Demenescu et al., 2010; McClure, Pope, Hoberman, Pine, & Leibenluft, 2003; Morningstar, Dirks, Rappaport, Pine, & Nelson, 2017). In addition, the pathophysiological mechanisms associated with threat bias seem to differ between irritability and anxiety (Kircanski et al., 2018b). It remains to be seen whether these also differ between irritability and depression.
4.4.5 Effects of age and sex on emotion recognition

In keeping with previous literature (Brosgole & Weisman, 1995; Chronaki et al., 2015; Morningstar et al., 2017), we found that emotion recognition increased with age. Regardless of depressive symptoms, younger children were more likely to make errors in emotion recognition than adolescents. However, both children and adolescents with high depressive symptoms made more errors in the recognition of happy stimuli. This suggests that the misidentification of happy stimuli by DMDD youth is closely related to depressive symptoms irrespective of their age and that this deficit might increase the risk of depression at any developmental period.

As hypothesised, in the whole sample, males made more errors overall than females in the recognition of facial and vocal emotions, which is line with previous reports (Kret & De Gelder, 2012; McClure, 2000; Thompson & Voyer, 2014). This was especially true for fearful stimuli, independent of depressive symptoms. In youth with DMDD, males with high depressive symptoms were worse at recognising happy stimuli than both males and females with low depressive symptoms. In contrast, DMDD females with high depressive symptoms did not differ in emotion recognition accuracy with any group. These results might seem paradoxical given that females are at higher risk of depression after puberty (Wade, Cairney, & Pevalin, 2002), and we have shown that deficits in emotion recognition are also associated with an increased risk of depression. One then would expect that better emotion processing skills in females might protect them for depression outcomes. However, it has been suggested that females’ increased sensitivity to others’ emotions during adolescence puts them at increased risk of experiencing interpersonal conflicts as negative life events and consequently for depressive mood (Cyranowski, Frank, Young, &
Shear, 2000). In contrast, the deficits in emotion recognition might have a greater effect on boys in the prediction of depressive mood. Similar to what we found in Chapter 3, in which vagal reactivity was differently associated with ODD symptoms in boys and girls (Vidal-Ribas, Pickles, Tibu, Sharp, & Hill, 2017), it is possible that sex moderates the relationship between deficits in emotion recognition and future depressive symptoms. Unfortunately, we could not test sex differences in the longitudinal associations with depressive symptoms because only six of the 25 participants included in the analyses were females.

4.4.6 Clinical implications

Our findings have potential implications for the prevention and treatment of depression in youth with severe irritability. Several studies have examined the effects of cognitive bias modification (CBM), targeting both interpretation and attentional bias, on current depression and depressive symptoms. The current evidence suggests that effects are small or non-significant, especially in clinical samples (Cristea, Kok, & Cuijpers, 2015; Cristea, Mogoase, David, & Cuijpers, 2015; Hallion & Ruscio, 2011; LeMoult et al., 2017; Mogoase, David, & Koster, 2014). However, promising results have been found in the prevention of depression in high-risk groups (Browning, Holmes, Charles, Cowen, & Harmer, 2012), and further prevention trials employing this approach are underway (Almeida et al., 2014). In addition, an open trial of interpretation bias training (IBT) in youth with DMDD increased happy, as opposed to angry, judgments of ambiguous faces, and this was associated with a reduction in irritability symptoms (Stoddard et al., 2016). A larger-scale RCT of IBT for DMDD is ongoing (ClinicalTrials.gov identifier NCT02531893). Given our current results, one plausible hypothesis is that this treatment
might not only reduce irritability symptoms but also prevent the development of future
depression.

With regards to pharmacological approaches, several studies suggest that serotonin
reuptake inhibitors (SRIs) might normalise deficits in emotional processing (Harmer et al.,
2003; Harmer et al., 2009), even before changes in mood occur (Pringle & Harmer, 2015).
This effect may be explained by the influence of antidepressants on neural circuits involved
in emotional processing, such as the amygdala and prefrontal areas (Fu et al., 2004;
Godlewska, Norbury, Selvaraj, Cowen, & Harmer, 2012; Roiser et al., 2012; Victor et al.,
2010). Whether this effect would be seen in youth with severe irritability remains to be
seen. However, there is indirect evidence that SRIs can be effective in the treatment of
irritability (Coccaro et al., 2009; Fava & Rosenbaum, 1999). Indeed, in Chapter 5 of this
thesis, we employ a randomised controlled trial (RCT) to examine whether adding
citalopram, an SRI, to stimulant medication decreases irritability in youth with DMDD
compared with placebo. Future studies should examine whether SRIs normalises emotional
processing deficits in youth with irritability, including aberrant neural responses, and
whether this is associated with lower rates of future depression.

4.4.7 Strengths and limitations

The findings of this study are supported by the inclusion of participants with a careful
clinical characterisation, and the use of a standardised behavioural task to assess emotion
recognition. Moreover, we employed a conservative multiple comparisons correction to
minimise Type I error. Also, this is the first study in youth with irritability that examines
distinct modalities of emotional stimuli at the same time.
Our findings should also be considered in light of its limitations. First, there was considerable attrition at follow-up. Thus, our longitudinal findings should be interpreted with caution and need further replication in larger samples. However, it should be noted that attrition analyses showed no baseline differences between the participants with follow-up data and those without.

Second, information about depressive symptoms was not collected in healthy participants; having this information would have strengthened the dimensional and follow-up analyses. Moreover, having this information would have allowed us to test interactions between irritability and depressive symptoms and its association with emotion recognition.

Third, the output of the analyses examining the effects of anxiety disorders, age, and sex on emotion recognition should also be interpreted with caution given the small number of observations in some groups.

Finally, although not strictly a limitation, in the current study we used an emotion recognition task. Therefore, our study is not directly comparable to other studies that used paradigms to examine attentional bias (Hommer et al., 2014; Salum et al., 2017) or interpretation bias (Brotman et al., 2010; Guyer et al., 2007; Stoddard et al., 2016), the latter usually tested with neutral or ambiguous stimuli. However, the misidentifications in the current task were with regards to specific emotions, such as fear and anger, and not randomly distributed across all possible emotions; this suggests that errors were emotion-specific, like the ones seen in attentional and interpretation bias paradigms (Bourke et al., 2010). Regardless, future studies in irritable youth should examine associations between attentional and interpretation biases and depressive symptoms.
4.5 Conclusion

In summary, deficits in emotion recognition are associated, cross-sectionally and longitudinally, with depressive symptoms in youth with severe irritability. Future studies should examine the genetic and neural mechanisms that contribute to such associations and test whether treatments targeting emotion recognition deficits could prevent the development of depression in youth with severe irritability.
CHAPTER 5: A double-blind randomised controlled trial of adjunctive citalopram in youth with chronic severe irritability treated with stimulant

5.1 Background

Chronic severe irritability is among the most common reasons for referral to child psychiatric services (Mikita & Stringaris, 2013; Stringaris et al., 2018) and is associated with substantial concurrent and future impairment (Althoff et al., 2010; Copeland et al., 2013; Copeland et al., 2014; Vidal-Ribas et al., 2016). In an attempt to capture youth whose main problem is the manifestation of chronic severe irritability, the American Psychiatric Association recently introduced the diagnosis of DMDD into the DSM-5 (APA, 2013). As described in depth in section 1.7 of this thesis, rigorous testing of treatments for chronic severe irritability is still at a nascent state despite the increasing recognition of its importance. Indeed, very few clinical trials have been conducted for DMDD, or its precursor, SMD (Benarous et al., 2017; Stringaris et al., 2018; Tourian et al., 2015). Moreover, within the existing trials, there has been only one pharmacological randomised controlled trial (RCT), with lithium, that yielded null results (Dickstein et al., 2009). In this Chapter, I present the results of an RCT of citalopram versus placebo as add-ons to open-label stimulant optimisation for the treatment of irritability in SMD, the precursor to DMDD.
5.1.1 Psychological approaches to the treatment of chronic severe irritability

The most supported psychological approaches to treat irritability are parent management training (PMT) and cognitive behavioural therapy (CBT). However, evidence for the efficacy of these interventions comes from studies focusing on disorders in which irritability is a common symptom, and none in SMD or DMDD alone (Barkley, 2013a, 2013b; Comer et al., 2013; Scott & O’Connor, 2012; Sukhodolsky & Scahill, 2012). Only one small RCT has tested these approaches in youth with ADHD and comorbid SMD (Waxmonsky et al., 2016). In this trial, the authors found that adding an integrative therapy (a combination of group-based PMT, and CBT) to optimised stimulant-medication treatment resulted in a reduction in irritability. However, these improvements were not maintained at follow-up.

In the last years, new or modified psychological therapies are being tested for DMDD. For example, there is encouraging preliminary evidence of efficacy of Dialectical Behaviour Therapy (DBT) (Perepletchikova et al., 2017) and Interpersonal Psychotherapy (IPT) (Miller et al., 2018) against treatment as usual (TAU) in adolescents with DMDD. In addition, evidence suggests that people with irritability in the context of DMDD may be responsive to computer-based interpretation bias re-training (Stoddard et al., 2016). Furthermore, the evaluation of transdiagnostic (Sukhodolsky et al., 2016) and exposure-based CBT (Kircanski et al., 2018a) for youth with DMDD is currently underway.

Whilst promising, the clinical efficacy of psychological treatments for DMDD requires further trials in larger samples. Pharmacological treatments have a clear role in most common psychiatric disorders in youth, including anxiety (Walkup et al., 2008), depression
Moreover, pharmacological interventions may work synergistically with psychological treatments (Walkup et al., 2008). Therefore, it is imperative to identify pharmacological agents for the treatment of children experiencing chronic irritability (Stringaris et al., 2018).

### 5.1.2 Pharmacological approaches to the treatment of chronic severe irritability

As mentioned before, so far there is only one pharmacological RCT designed specifically for the treatment of severe chronic irritability (Dickstein et al., 2009). Prior literature suggested a therapeutic effect of lithium on aggression (Campbell et al., 1995; Malone et al., 2000). Within the context of the controversy around paediatric bipolar disorder (BD) (Wozniak et al., 1995), it was thought that since chronic severe irritability was a manifestation of BD in children, lithium should be the treatment of choice. However, this small trial (N=24) (Dickstein et al., 2009) showed that lithium was ineffective for the treatment of severe irritability in children and adolescents with SMD, the precursor to DMDD (Leibenluft et al., 2003). Although there have not been other pharmacological RCTs for SMD or DMDD since, and open trial with risperidone showed a reduction in irritability symptoms in youth with SMD (Krieger et al., 2011).

Further evidence for pharmacological approaches to irritability comes from treatment trials in youth with ADHD. Stimulant medication, the first-line treatment for ADHD, has been shown to reduce irritability, aggression and emotional instability in youth with ADHD (Connor et al., 2002; Shaw et al., 2014). A post-hoc analysis of the Multimodal Treatment Study of Children with ADHD (Fernandez de la Cruz et al., 2015) suggested that stimulant treatment in those with ADHD may be effective in reducing irritability. In a retrospective
analysis of data from a cross-over, placebo-controlled trial of methylphenidate (MPH), Waxmonsky et al. (2008) reported a significant reduction of irritability/aggression in ADHD. In the last years, open-trials using stimulant monotherapy in youth with ADHD and comorbid DMDD/SMD have shown reductions of irritability (Baweja et al., 2016; Winters et al., 2018). Nevertheless, a significant proportion (~50%) of those with aggression and ADHD remain refractory to stimulant medication even when combined with parent training/behavioural treatments (Blader et al., 2010; Waxmonsky et al., 2008). Therefore, alternative approaches such as adjunctive pharmacological treatments are worth testing.

Adjunctive treatments to stimulant have been tested on the related construct of aggression. The Treatment of Severe Childhood Aggression (TOSCA) study group found that add-on risperidone and placebo were no different to each other when combined with parent-training plus stimulant medication in children with ADHD (Aman et al., 2014) with similar results obtained after a 12-week follow up (Findling et al., 2017). In terms of other mood-stabilising medication, Blader and colleagues reported that add-on valproate reduced aggression in children with ADHD who did not respond to optimised stimulant plus family education (Blader et al., 2009). Yet, this study was conducted in a small number of participants (n=14).

One promising approach to treating chronic irritability might be serotonin reuptake inhibitors (SRIs). Evidence derived from adult samples indicates that SRIs might be efficacious in the treatment of irritability in depression (Fava & Rosenbaum, 1999), as well as in patients with intermittent explosive disorder (Coccaro et al., 2009) and those with premenstrual dysphoria (Dimmock et al., 2000). In a recent systematic review on the effect
of antidepressants on irritability in young people, Kim and Boylan (2016) identified two uncontrolled studies of SRIs that reported on irritability as an outcome (Armenteros & Lewis, 2002; Garland & Weiss, 1996); both studies indicated substantial improvement of irritability with SRI treatment, especially citalopram (Armenteros & Lewis, 2002). In addition, as shown in Chapter 2, irritability is a specific and robust predictor of future anxiety and depressive disorders (Vidal-Ribas et al., 2016) Moreover, findings from genetically-informative studies suggest shared pathophysiological mechanisms among irritability, anxiety and depression (Savage et al., 2015; Stringaris et al., 2012b), thus indicating that agents effective for these disorders could also be useful for irritability. Taken together, these data suggest that SRIs is a promising candidate for the treatment of children with severe chronic irritability.

5.1.3 Aims and hypotheses

In the current study, we report the results of an RCT of adjunctive citalopram, an SRI, against placebo in youth treated with stimulant who were originally recruited because they fulfilled criteria for SMD. Prior to randomisation, children took part in an open-label stimulant optimisation lead-in phase. This lead-in phase was motivated by the observation that a majority of youth with DMDD also suffer from ADHD (Deveney et al., 2015) and, as mentioned before, increased evidence suggests that stimulants appear efficacious for irritability in ADHD (Fernandez de la Cruz et al., 2015) even when comorbid with SMD/DMDD (Baweja et al., 2016; Winters et al., 2018). Our RCT, thus, was designed to provide a rigorous test of SRI effects on irritability over and above stimulant optimisation in youth with SMD.
Our primary aim was to test the efficacy of citalopram plus methylphenidate vs. placebo plus methylphenidate in decreasing irritability in youth with SMD using a Phase II RCT. We hypothesised that add-on citalopram will be associated with a greater improvement in irritability than add-on placebo.

Our secondary aims were to test the effects of citalopram plus methylphenidate vs. placebo plus methylphenidate on overall functional impairment, anxiety symptoms, and depressive symptoms. We hypothesised that add-on citalopram will be associated with a greater improvement in these secondary outcomes than add-on placebo.

A further secondary aim of this study was to assess the tolerability and adverse effects of citalopram plus methylphenidate vs. placebo plus methylphenidate.

5.2 Methods

5.2.1 Participants

This study was conducted at the National Institute of Mental Health Division of Intramural Research Programs (NIMH DIRP) from November 2008 until January 2018 and was approved by the National Institute of Health’s Combined Neuro-Science Institutional Review Board (CNS-IRB). Prior to participation, the study was explained to parents and patients. All children gave written assent and parents gave written informed consent. Subjects with ages 7–17 years were recruited through advertisements that were placed in local parenting magazines, on support groups’ websites, and distributed to psychiatrists nationwide. Also, information about the study was provided in talks to local practitioners, and advocacy/parent groups. The study was registered on ClinicalTrials.gov (Identifier: NCT00794040). The final protocol of this study is available upon request.
5.2.2 Inclusion and exclusion criteria

All participants met criteria for SMD, the precursor of DMDD (Leibenluft et al., 2003). SMD criteria originally were designed to capture those youth with non-episodic irritability and hyperarousal symptoms who were frequently diagnosed with bipolar disorder in clinical settings (Leibenluft et al., 2003). While DMDD was introduced in the DSM after the study began, all but one randomised participant (n=48, 98%) also met criteria for DMDD; one participant did not meet DMDD criteria because the onset of severe irritability was after 10 years old but before age 12.

Inclusion criteria for the trial were: (1) irritability operationally defined as markedly increased reactivity to negative emotional stimuli manifest verbally or behaviorally at least three times weekly; (2) abnormal mood (anger or sadness), present at least half of the day most days; (3) three hyperarousal symptoms (insomnia, agitation, distractibility, racing thoughts or flight of ideas, pressured speech, intrusiveness); (4) symptoms cause severe impairment in at least one setting (home, school, or peers) and at least mild impairment in a second setting; (5) symptom onset before age 12 and currently present for at least 12 months without any symptom-free period greater than 2 months; (6) failing treatment as defined by current CGAS score <60, outpatient treatment provider agrees that the child’s response to treatment is no more than minimal 7) On the basis of record review and interviews with child and parent, the research team agrees that the child’s response to his/her current treatment is no more than minimal. Having a score of > 12 on the irritability subscale of the Aberrant Behavior Checklist (ABC) (Aman, Singh, Stewart, & Field, 1985) was initially an inclusion criterion but the collection of the ABC was stopped for several reasons (see below). All the patients had histories of failed response to treatment either
pharmacological, psychological or both. None of the patients was medication naive at the time of enrolment. For all patients, their providers in the community endorsed their patient’s participation in the study and acknowledged that the current treatment was only minimally successful and other options should be explored.

Exclusion criteria for the trial were: (1) presence of cardinal bipolar symptoms of elevated, expansive mood, grandiosity, inflated self-esteem, or episodically decreased need for sleep; (2) distinct episodes of hypo/manic symptoms greater than 1 day; (3) current Major Depressive Disorder; (4) Autism Spectrum Disorder; (5) psychosis; (6) Post-Traumatic Stress Disorder; (7) substance abuse within 3 months; (8) medical illnesses that require medications, are chronic, unstable, could cause SMD symptoms or are contraindications to treatment with SRI or stimulant (e.g. liver, seizure, renal, platelet disorder), or require medications that are contraindicated with SRIs or methylphenidate; (9) intelligence quotient (IQ) <70; (10) pregnant, lactating, or sexually active without barrier method of contraception; (11) failed adequate trial (defined as four weeks of consecutive treatment with no less than 20 mg citalopram or 10 mg escitalopram) or severe ill effects while on citalopram or escitalopram; (12) history of hypersensitivity or severe adverse reaction to methylphenidate or serious adverse reactions (psychosis, severely increased activation compared to baseline) to methylphenidate or amphetamines.

Co-morbid ADHD, ODD, or anxiety disorders were not exclusionary. Full-scale intelligence quotient (FSIQ) was measured using the Wechsler Abbreviated Scale of Intelligence (WASI) for all subjects (Wechsler, 1999).
Following a telephone interview to screen for relevant inclusion or exclusion criteria, record review, and consultation with the child’s treating clinician, candidates for the study were invited to the NIMH IRP (n=311). On-site screening included the Kiddie Schedule for Affective Disorders Present and Lifetime Version (K-SADS-PL) (Kaufman et al., 1997) with an additional supplement for SMD, designed in collaboration with Joan Kaufman, Ph.D., to ascertain whether children met criteria for this syndrome. All diagnostic measures were administered to the parent and child individually by trained post-graduate level clinicians with established interrater reliability (k=0.9, including distinguishing SMD subjects from those with distinct hypo/manic episodes) (Kaufman et al., 1997; Leibenluft et al., 2003). Diagnoses were based on best-estimate procedures (Leckman, Sholomskas, Thompson, Belanger, & Weissman, 1982) generated in a consensus conference led by two psychiatrists with extensive experience evaluating children with mood disorders.

5.2.3 Enrolment

Figure 5.1 shows an overview of the trial. After an on-site screening, subjects who met inclusion criteria were offered enrolment in the randomised, placebo-controlled trial (RCT). To maximise safety and observational data collection while they were tapered off medication and began open treatment with MPH, patients received inpatient care on a research child psychiatric unit at the NIH Clinical Center. Upon admission to the hospital, participants were assessed as described below (Baseline at admission) and then gradually tapered off psychotropic medications (Phase I). They were monitored closely for clinical deterioration and side effects; weekend passes with family were offered weekly or every other weekend depending on proximity to NIH. Medication tapering (Median=3 weeks) was individually tailored and prioritised medications with longer half-lives (e.g., atypical
neuroleptics, antidepressants) before those with short half-lives (e.g., psychostimulants).

Care in the hospital for all patients included weekday on-site school, weekday rounds meeting with a psychiatrist or nurse practitioner, weekly meetings with parent(s), individual and group activities, and milieu treatment that are common in inpatient paediatric psychiatry units. Concomitant therapy and any other psychological treatment were not allowed during participation in the protocol. During medication taper, 12 participants were discharged for meeting the following exclusion criteria: brain abnormalities found during magnetic resonance imaging (MRI) (n=2), presented autism spectrum disorder (ASD) symptoms or diagnosis (n=5; one with possible psychotic symptoms), too much aggression (n=3), conduct disorder (CD) (n=1, with possible psychotic symptoms), posttraumatic stress disorder (PTSD) (n=1) (Figure 5.2)

5.2.4 Treatment of ADHD symptoms

After completing up to two medication-free weeks (Phase II), participants completed up to 5 weeks of open treatment with MPH to achieve optimal control of ADHD symptoms (Phase III).

Only oral administration of MPH (immediate-release, intermediate release, or extended release forms) was used. Commercial preparations of MPH (in its various forms) were obtained from local distributors. MPH dosing was based on clinical judgement optimizing effects on attention, restlessness, and impulsivity while balancing adverse effects on sleep and appetite. The dose was not fixed by protocol, but a dose range was allowed based on efficacy and tolerability. The goal was 0.8 mg/kg/day to 1.2 mg/kg/day not to exceed 80 mg or 2 mg/kg/day. Treatment decisions about MPH drew from reports from nursing staff and
classroom teachers at the NIH Clinical Center School which patients attended during their inpatient care. MPH treatment began with immediate release forms twice daily and then was converted to intermediate or extended release forms of MPH to achieve once-a-day dosing to the greatest degree possible. Morning dosing with immediate plus longer release forms of MPH was permitted.

Prior to randomisation, those who no longer met SMD criteria due to decreased irritability were discharged (n = 11), and one participant was discharged due to excessive anxiety about randomisation (Figure 5.2). Those who continued to meet SMD criteria while on MPH (n = 53) had a second set of baseline measures collected (Baseline at randomisation) and then were randomised to add-on treatment with either CTP or PBO for an 8-week double-blind RCT (Phase IV) (Figure 5.1). Treatment allocation was randomised by the Pharmaceutical Development Service of the National Institutes of Health Clinical Center Pharmacy Department using a random numbers table. It was randomised in alternating blocks of six and four in a 1:1 ratio.

During the randomised phase, every prescriber and everyone who came into contact with the children was blind to their treatment assignment. Commercial preparations of CTP were obtained from local distributors and investigational (blinded) capsules were compounded and supplied by the NIH Clinical Center Pharmaceutical Development Section.

The CTP/PBO dosing schedule began with one capsule (5 mg) daily; at 5-day intervals, the dose could increase by one-capsule. After roughly 3 weeks of the 8-week trial (Mean=3.1 weeks, SD=1.5, with no difference between groups p=0.899), when patients were receiving four capsules (equivalent to 20 mg/d), they were discharged from inpatient care and
returned home. Once home, they received study medication and were monitored with weekly clinical ratings, as described below, by blind raters who were not prescribers. These ratings were done by telephone, alternating with weekly outpatient visits. Based on a clinical judgment of minimal side effects and ongoing symptoms, weekly increments were permitted until a maximum dose of 8 capsules per day (equivalent to 40 mg) was achieved. During the RCT, lorazepam was available as a PRN medication for agitation; however, no SMD subject received lorazepam during the RCT.

Figure 5.1. Diagram of the trial design
5.2.5 Side Effects

Side effects were ascertained by checklists. These were completed by nurses (during inpatient admission) or parents (after discharge), all of whom were blind to treatment assignment, and instructed to rate what they observed without reference to changes from baseline irritability, time of day or severity of the presentation.

5.2.6 Assessment

Two graduate-level, highly-experienced clinicians completed the weekly Clinical Global Impression’s improvement (CGI-I) and severity (CGI-S) scales (Spearing, Post, Leverich, Brandt, & Nolen, 1997) independently of each other, as well as the Children’s Global Assessment of Severity (CGAS) (Shaffer et al., 1983), the Pediatric Anxiety Rating Scale (PARS) (Riddle et al., 2002), and on the Children’s Depression Rating Scale (CDRS) (Poznanski & Mokros, 1996). Finally, ratings were reached by consensus in a case conference with a senior child and adolescent psychiatrist (Dr Kenneth Towbin); all participants were blind to the treatment condition.

Children in this cohort typically presented with multiple comorbidities (most commonly, anxiety disorders and ADHD, Table 5.1). Since the goal of this trial was to determine the impact of stimulant plus SRI treatment on irritability specifically, the primary outcome was a CGI-I scale focused on irritability; a similar strategy has been used in bipolar disorder (Spearing et al., 1997).

On admission and every week thereafter, clinicians (masters or PhD-level) blind to treatment assignment used a semi-structured interview to ascertain the severity of mood dysregulation symptoms (i.e., temper outbursts, irritable mood between outbursts,
hyperarousal). Informants were parents and the child. When the participant was
hospitalised, the child, nursing staff, and parents were interviewed because patients were
often on pass with their parents on weekends. The “or” rule was applied to collate these
reports. These were used to generate a CGI-I SMD score relative to the defined baseline
(i.e., baseline for the time just prior to admission and baseline before randomisation); in this
score, hyperarousal symptoms were given minimal weight, so that the score largely
reflected temper outbursts and irritable mood between outbursts. To ensure reliability, the
correspondence between each rating and the clinical documentation supporting that rating
was reviewed by a senior child psychiatrist (Dr Kenneth Towbin) who was also blind to
treatment assignment.

Our primary categorical clinical outcome measure was a CGI-I score of 2 [much
improved], or 1 [symptom-free], consistent with the original protocol and common practice
(Brent et al., 2008; March et al., 2004). Sensitivity analyses were also conducted using the
CGI-S, as in other studies (Brent et al., 2008).

The ABC (Aman et al., 1985) was initially designated as a primary outcome measure;
however, it proved inappropriate for the population we studied as it showed minimal
variability. Thus, we stopped collecting it early on.

Secondary outcomes were functional impairment (measured with CGAS), improvement in
anxiety symptoms (measured with PARS), and improvement in depressive symptoms
(measured with CDRS).
5.2.7 Statistical analysis

A Statistical Analysis Plan, available in Appendix A of this thesis, was written after the end of data collection, but prior to the group-label unblinding of the analyst and statistician (Pablo Vidal-Ribas and Andrew Pickles). All children who completed at least one post-randomisation assessment were included in intent-to-treat analyses.

Additionally, to examine the effect of stimulant optimisation before randomisation on irritability severity, we compared changes in CGI-S between baseline at admission and baseline at randomisation with a paired t-test, in everyone randomised. We also performed this analysis for each of the SMD symptom domains (i.e., temper outbursts, mood between outbursts and hyperarousal symptoms).

Primary and secondary outcomes after randomisation were analysed using multilevel models (MLM) estimated by maximum-likelihood enabling the inclusion of participants incompletely observed under the missing at random assumption. This is the recommended analytic approach for repeated measures data (Hamer & Simpson, 2009) over other approaches like last observation carried forward (LOCF). However, a sensitivity analysis was carried out using the LOCF method.

For our primary categorical outcome (i.e. treatment response as defined by a CGI-I score < 3), an efficient estimate was obtained by fitting a growth curve model for the repeated binary response fitted by maximum likelihood in the Stata program gsem, with binomial family and logit link function. The model included a random intercept. The fixed part of the model included treatment group, number of weeks as a measure of time, and the group by week interaction (a test for a quadratic term was also carried out). The post-estimation
are used to estimate the group difference at the 8th week and its 95% confidence interval (CI). To assist in interpretation, this conditional or subject-specific effect estimate (and its 95% CI) was translated to an approximate, but more easily understood, marginal mean estimate (Szmaragd, Clarke, & Steele, 2013). The estimate from the group by week interaction in the model provided a measure of the difference in response rate between groups across the 8 weeks of the trial. Estimated proportions of response were plotted with estimates provided by command margins. We also performed this analysis for each of the SMD symptom domains (i.e., temper outbursts, mood between outbursts and hyperarousal symptoms).

For the analysis of continuous outcomes (i.e., severity as measured by CGI-S, functionality as measured by CGAS, anxiety symptoms as measured by PARS, and depressive symptoms as measured by CDRS) a growth curve model was fitted using the `xtmixed` routine of the Stata statistical program, with restricted maximum likelihood (REML) estimation and an exchangeable covariance matrix for the covariances of the errors of the repeated measures. The fixed part of the model included the baseline values of the outcome variables, treatment group, number of weeks as a measure of time, and the interaction of group by week. The post-estimation `lincom` was used to estimate the group difference at 8th week and its 95% CI. The estimate from the group by week interaction in the model provided a measure of the difference in severity between groups across the 8 weeks of the trial. Distributional assumptions of our primary outcomes were checked by the use of Q-Q plots of residuals.
Frequencies of the most common adverse effects are reported if present in more than one subject in either study group. These were compared using 2-sided Fisher’s exact test.

5.3 Results

5.3.1 Participant sample derivation and characteristics

Figure 5.2 shows the CONSORT diagram. Of the 53 participants eligible for randomisation, 25 were allocated to receive adjunctive citalopram (CTP) and 28 to receive adjunctive placebo (PBO). All subjects completed at least one post-randomisation assessment. However, the first four participants recruited – 2 from each group- were excluded from the analysis because of technical problems with data collection. Of the remaining 49 participants included in the intent-to-treat analysis, 41 completed the trial. One participant was withdrawn by the experimenters at week 7 due to increased levels on Liver Function Test (LFT); specifically, alanine transaminase (ALT) and aspartate transaminase (AST) were elevated to three times the upper limit of normal, but these normalised after cessation of methylphenidate, while the patient continued on open citalopram. Another 7 participants withdrew assent before week 8. There were no differences between groups in the number of weeks in inpatient care before participants were discharged to home (CTP: Mean=3.1 SD=1.1, PBO: Mean=3.1 SD=1.8 t(45)=0.13, p=0.899). Table 5.1 contains the demographic and clinical characteristics of the 49 participants included in the intent-to-treat analysis.
Figure 5.2. CONSORT diagram of the trial. Of the randomised participants (N=53), 4 participants were excluded from analysis because the CGI collected was a different version and thus not comparable. Of the 49 participants included in the analyses, 41 completed the trial; 7 withdrew assent before the 8th week and 1 was ruled out of the study by the experimenters due to high levels of liver function test.
Table 5.1. Sample demographic and clinical characteristics.

<table>
<thead>
<tr>
<th></th>
<th>MPH + CTP n=23 SMD</th>
<th>MPH + PBO n=26 SMD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex, female, n (%)</td>
<td>10 (43)</td>
<td>6 (23)</td>
</tr>
<tr>
<td>Age (years), Mean (SD) range</td>
<td>11.4 (2.5) 7-15</td>
<td>11.7 (2.1), 8-14</td>
</tr>
<tr>
<td>Ethnicity, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>19 (83)</td>
<td>20 (77)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>0 (0)</td>
<td>3 (12)</td>
</tr>
<tr>
<td>Asian</td>
<td>1 (4)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Multiple races</td>
<td>3 (13)</td>
<td>2 (8)</td>
</tr>
<tr>
<td>Unknown</td>
<td>0 (0)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>IQ, Mean (SD)</td>
<td>109.3 (14.8)</td>
<td>113 (9.2)</td>
</tr>
<tr>
<td>Co-occurring diagnoses, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any disorder</td>
<td>23 (100)</td>
<td>26 (100)</td>
</tr>
<tr>
<td>ADHD</td>
<td>20 (87)</td>
<td>24 (92)</td>
</tr>
<tr>
<td>ODD</td>
<td>17 (74)</td>
<td>22 (85)</td>
</tr>
<tr>
<td>CD</td>
<td>0 (0)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Any Anxiety disorder</td>
<td>13 (57)</td>
<td>15 (58)</td>
</tr>
<tr>
<td>Medicated at admission, n (%)</td>
<td>21 (91)</td>
<td>23 (84)</td>
</tr>
<tr>
<td>Number of medications at admission, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>2 (9)</td>
<td>3 (12)</td>
</tr>
<tr>
<td>1</td>
<td>4 (17)</td>
<td>5 (19)</td>
</tr>
<tr>
<td>2</td>
<td>5 (22)</td>
<td>3 (12)</td>
</tr>
<tr>
<td>3 or more</td>
<td>12 (52)</td>
<td>15 (57)</td>
</tr>
<tr>
<td>Type of medication, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atypical antipsychotic</td>
<td>14 (60)</td>
<td>17 (65)</td>
</tr>
<tr>
<td>Typical antipsychotic</td>
<td>0 (0)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Stimulant</td>
<td>8 (35)</td>
<td>9 (35)</td>
</tr>
<tr>
<td>Non-stimulant (Ato, Gua, Clo)</td>
<td>9 (39)</td>
<td>12 (46)</td>
</tr>
<tr>
<td>SRI</td>
<td>7 (30)</td>
<td>7 (27)</td>
</tr>
<tr>
<td>Other antidepressants</td>
<td>3 (13)</td>
<td>3 (12)</td>
</tr>
<tr>
<td>Anti-convulsant</td>
<td>10 (43)</td>
<td>7 (27)</td>
</tr>
<tr>
<td>Lithium</td>
<td>4 (17)</td>
<td>3 (12)</td>
</tr>
<tr>
<td>Prior medication treatments, Mean (SD) range</td>
<td>4.1 (3.5)</td>
<td>5.3 (3.1)</td>
</tr>
<tr>
<td>Prior hospitalizations, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>15 (65)</td>
<td>15 (58)</td>
</tr>
<tr>
<td>1</td>
<td>2 (9)</td>
<td>5 (19)</td>
</tr>
<tr>
<td>2 or more</td>
<td>6 (26)</td>
<td>6 (23)</td>
</tr>
</tbody>
</table>

Table 5.1. Sample demographic and clinical characteristics (Continued)

<table>
<thead>
<tr>
<th></th>
<th>MPH + CTP n=23 SMD</th>
<th>MPH + PBO n=26 SMD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Educational programs, n (%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Mainstream school (MS) | 11 (48) | 4 (15)  
MS with accommodations | 1 (4) | 1 (4)  
In-home school | 1 (4) | 1 (4)  
Special education | 10 (43) | 20 (77)  

Baseline admission, Mean (SD)  
CGI-S | 4.4 (0.6) | 4.6 (0.5)  
CGAS | 44 (6.1) | 41.7 (2.2)  
CDRS | 29.7 (5.9) | 32.0 (7.6)  
PARS | 15.7 (4.5) | 14.8 (4.7)  

Baseline randomisation, Mean (SD)  
CGI-S | 4.0 (0.5) | 4.3 (0.6)  
CGAS | 44.4 (3.3) | 42.9 (3.6)  
CDRS | 29.4 (5.7) | 34.6 (8.6)  
PARS | 13.3 (5.9) | 15.8 (5.2)  

MPH, methylphenidate; CTP, citalopram; PBO, placebo; SMD, severe mood dysregulation; ADHD, attention deficit hyperactivity disorder; ODD, oppositional defiant disorder; CD, conduct disorder; MDD, major depressive disorder; CGI, Clinical Global Impression; CGAS, Children’s Global Assessment Severity; CDRS, Children’s Depression Rating Scale; PARS, Pediatric Anxiety Rating Scale; Ato, Atomoxetine; Gua, Guanfacine; Clo, Clonidine; SRI, Serotonin reuptake inhibitor  
Missing data: a n=1 in MPH+CTP, n=1 in MPH+PBO, b n=1 in MPH+CTP, c n=7 in MPH+CTP, n=8 in MPH+PBO, d n=1 in MPH+PBO  
Note: Any anxiety disorder includes Separation Anxiety Disorder, Generalized Anxiety Disorder, Social Anxiety Disorder, and Panic Disorder.

5.3.2 Response to open-label stimulant optimisation

The severity of irritability and temper outbursts decreased from admission to randomisation (ES=0.60, 95%CI 0.30-0.89, p=0.0002) in the sample included in the intent-to-treat analysis (Figure 5.3). However, of the randomised participants, only one was a responder based on the CGI-I score. This is because, as per protocol, those who responded and no longer met irritability threshold criteria (N=11) were not randomised.
Figure 5.3. Change in irritability severity before and after randomisation in the sample included in the intent-to-treat analysis. The period between baseline at admission (Adm. Baseline) and baseline at randomisation (Rand. Baseline) was variable and included the washout period (Phase I), medication-free period (Phase II), and open-label lead phase with stimulant optimisation (Phase III). Before randomisation, change in irritability severity is shown for the entire sample (N=49). After randomisation, both observed and estimated severity - provided by command `margins` in Stata - are displayed by week and treatment group (CTP, N=23; PBO, N=26).
<table>
<thead>
<tr>
<th>Outcome</th>
<th>Baseline randomisation</th>
<th>8\textsuperscript{th} Week of trial</th>
<th>Between-group difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CTP</td>
<td>PBO</td>
<td></td>
</tr>
<tr>
<td>CGI-I response</td>
<td>% (SE) n</td>
<td>% (SE) n</td>
<td>p-value</td>
</tr>
<tr>
<td>- Estimated</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>35 (10) 8</td>
<td>0.006</td>
</tr>
<tr>
<td>- Observed</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>25 (10) 5</td>
<td>10 (7) 2</td>
<td></td>
</tr>
<tr>
<td>CGI-S</td>
<td>M (SE) n</td>
<td>M (SE) n</td>
<td>p-value</td>
</tr>
<tr>
<td></td>
<td>4.0 (0.4) 23</td>
<td>4.3 (0.4) 26</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td>3.1 (0.3)</td>
<td>3.9 (0.3)</td>
<td></td>
</tr>
<tr>
<td>CGAS</td>
<td>44.4 (0.2) 22</td>
<td>42.9 (0.2) 26</td>
<td>0.109</td>
</tr>
<tr>
<td></td>
<td>52.6 (2.3)</td>
<td>47.2 (2.1)</td>
<td></td>
</tr>
<tr>
<td>CDRS</td>
<td>29.4 (0.5) 16</td>
<td>34.6 (0.7) 18</td>
<td>0.680</td>
</tr>
<tr>
<td></td>
<td>28.6 (1.8)</td>
<td>30.1 (1.8)</td>
<td></td>
</tr>
<tr>
<td>PARS</td>
<td>13.3 (1.2) 23</td>
<td>15.8 (1.0) 25</td>
<td>0.598</td>
</tr>
<tr>
<td></td>
<td>12.0 (1.2)</td>
<td>13.4 (1.2)</td>
<td></td>
</tr>
</tbody>
</table>

MPH, methylphenidate; CTP, citalopram; PBO, placebo; CGI, Clinical Global Impression; CGAS, Children’s Global Assessment Severity; CDRS, Children’s Depression Rating Scale; PARS, Pediatric Anxiety Rating Scale. M, Mean, SE, standard error.

Baseline descriptive statistics are based in observed data. Descriptive statistics at 8\textsuperscript{th} week of the trial and p-values of differences in the randomised controlled trial are based on model-based estimates of the intent-to-treat analysis. For CGI-I response, both estimated (n=49) and observed data (n=41) are provided.
5.3.2.1 Response to open-label stimulant optimization by SMD symptom domains

Here we report the changes in SMD symptom domains (i.e., temper outbursts, irritable mood between outbursts and hyperarousal symptoms) in the sample included in the intent-to-treat analysis between admission and randomisation.

The severity of *temper outbursts* decreased from admission to randomisation (ES=0.54, 95%CI 0.24-0.82, p=0.0006) (**Figure 5.4, depicted as the black dotted line**). The severity of *irritable mood between outbursts* did not significantly change from admission to randomisation (ES=-0.08, 95%CI -0.37-0.21, p=0.585) (**Figure 5.5, depicted as the black dotted line**). The severity of *hyperarousal symptoms* showed the larger decrease between admission and randomisation (ES=1.10, 95%CI 0.81-1.40, p<0.0001) (**Figure 5.6, depicted as the black dotted line**).

**Figure 5.4.** Change in *temper outbursts* severity before (depicted as the black dotted line) and after randomisation in the sample included in the intent-to-treat analysis.
Figure 5.5. Change in mood between outbursts severity before (depicted as the black dotted line) and after randomisation in the sample included in the intent-to-treat analysis.

Figure 5.6. Change in hyperarousal symptoms severity before (depicted as the black dotted line) and after randomisation in the sample included in the intent-to-treat analysis.
5.3.3 Response to citalopram versus placebo

For our primary outcome (i.e., rates of response to treatment as measured with CGI-I), estimated proportions of response differed at the 8th week of treatment between citalopram and placebo groups (35% CTP vs. 6% PBO, difference of 29%, $p=0.005$; OR=11.70, 95%CI 2.00-68.16, $p=0.006$) (Table 5.2), with a number of patients needed to treat (NNT) of three (3). In addition, the group by week interaction for the estimated proportions of response across the 8 weeks of trial was also significant ($b=0.43$, 95%CI 0.12, 0.74, $p=0.006$) (Table 5.2). Figure 5.7 depicts the estimated proportion of response for all participants included in the analyses against the observed proportion of responses for individuals with available data.

**Figure 5.7.** Proportion of treatment response by week and treatment group. Both observed and estimated proportions of response are displayed. Bars represent 95% CI. Estimated proportions were extracted with command margins in Stata and are based on n=49. Observed proportions are based on n=49 at week 1; n=48 at week 2; n=47 at weeks 3, 4; n=43 at weeks 5, 6, 7; and n=41 at week 8.
The difference in irritability severity - as measured with the CGI-S - at the 8th week of the trial between groups was a non-significant trend (b=-0.62, 95%CI -1.32, 0.09, p=0.085, ES=-1.11 95%CI -2.38-0.15). However, a significant group by week interaction for the CGI-S emerged across the 8 weeks of the trial (b=-0.11, 95%CI -0.21, -0.002, p=0.046) (Table 5.2 and Figure 5.3).

Groups did not differ in functional impairment as measured with the CGAS at the 8th week of trial (b=4.72, 95%CI -1.30, 10.74 p=0.124, ES=1.36 95%CI -0.37-3.09), and the group by week interaction was not significant (b=0.75, 95%CI -0.17, 1.66, p=0.109) (Table 5.2).

Groups did not differ in severity of depressive symptoms as measured with the CDRS at the 8th week of the trial (b=0.02, 95%CI -4.76, 4.80, p=0.993, ES=0.00 95%CI -0.62-0.62), and the group by week interaction was not significant (b=-0.15, 95%CI -0.87, 0.57, p=0.680) (Table 5.2). Similarly, groups did not differ in severity of anxiety symptoms as measured with the PARS at the 8th week of the trial (b=-1.02, 95%CI -4.23, 2.19, p=0.534, ES=-0.19 95%CI -0.79-0.41), and the group by week interaction was not significant (b=0.02, 95%CI -0.40, 0.45, p=0.909)

Distribution of lowest level residuals was close to normal, so no transformations were required. Finally, the number of weeks in inpatient care did not have an impact on the outcomes.
5.3.3.1 Response to citalopram versus placebo by SMD symptom domains

Here we report the changes in CGI-I by SMD symptom domains (i.e., temper outbursts, irritable mood between outbursts and hyperarousal symptoms) in the sample included in the intent-to-treat analysis during the 8-week RCT.

For temper outbursts, estimated proportions of response differed at the 8th week of treatment between citalopram and placebo groups (36% CTP vs. 9% PBO, difference of 27%, \( p=0.009 \); OR=7.28, 95%CI 1.60-33.03, \( p=0.010 \)). In addition, the group by week interaction for the estimated proportions of response across the 8 weeks of trial was also significant (\( b=0.35, 95\%CI 0.08, 0.63, p=0.010 \)) (Figure 5.8). The difference in severity of temper outbursts - as measured with the CGI-S - at the 8th week of the trial between groups was a non-significant trend (\( b=-0.68, 95\%CI -1.38, 0.03, p=0.059, ES=-1.12 95\%CI -2.29-0.04 \)). However, a significant group by week interaction for the CGI-S emerged across the 8 weeks of the trial (\( b=-0.11, 95\%CI -0.22, -0.01, p=0.039 \)) (Figure 5.4). For irritable mood between outbursts, estimated proportions of response differed at the 8th week of treatment between citalopram and placebo groups (30% CTP vs. 10% PBO, difference of 20%, \( p=0.045 \); OR=4.22, 95%CI 1.07-16.71, \( p=0.040 \)). In addition, the group by week interaction for the estimated proportions of response across the 8 weeks of trial was also significant (\( b=0.22, 95\%CI 0.01, 0.44, p=0.040 \)) (Figure 5.9). The difference in severity of irritable mood between outbursts - as measured with the CGI-S - at the 8th week of the trial between groups was not significant (\( b=-0.59, 95\%CI -1.31, 0.13, p=0.107, ES=-0.85 95\%CI -1.87-0.18 \)). However, a significant group by week interaction for the CGI-S emerged across the 8 weeks of the trial (\( b=-0.11, 95\%CI -0.21, -0.01, p=0.035 \)) (Figure 5.5).
Figure 5.8. Proportion of treatment response by week and treatment group for *temper outbursts*. Both observed and estimated proportions of response are displayed. Bars represent 95%CI.

Figure 5.9. Proportion of treatment response by week and treatment group for *mood between outbursts*. Both observed and estimated proportions of response are displayed. Bars represent 95%CI.
For hyperarousal symptoms, estimated proportions of response did not differ at the 8th week of treatment between citalopram and placebo groups (9% CTP vs. 13% PBO, difference of -4%, p=0.642; OR=0.65, 95%CI 0.10-4.13, p=0.652). In addition, the group by week interaction for the estimated proportions of response across the 8 weeks of trial was not significant (b=-0.08, 95%CI -0.41, 0.26, p=0.652) (Figure 5.10). Similarly, the difference in severity of hyperarousal symptoms - as measured with the CGI-S - at the 8th week of the trial between groups was not significant (b=-0.11, 95%CI -0.53, 0.31, p=0.610, ES=-0.16 95%CI -0.79-0.46). However, a significant group by week interaction for the CGI-S emerged across the 8 weeks of the trial (b=-0.07, 95%CI -0.13, -0.01, p=0.030) (Figure 5.6).

![Figure 5.10](image_url)

**Figure 5.10.** Proportion of treatment response by week and treatment group for hyperarousal symptoms. Both observed and estimated proportions of response are displayed. Bars represent 95%CI.
5.3.3.2  Sensitivity analysis using Last Observation Carried Forward

Results and estimates of the sensitivity analyses, which employed the last observation carried forward method, were very similar to the results described above.

For our primary outcome (i.e., rates of response to treatment as measured with CGI-I), estimated proportions of response differed at the 8\textsuperscript{th} week of treatment between citalopram and placebo groups (33% CTP vs 5% PBO; OR=12.31, 95%CI 2.20-68.89, \(p=0.004\)), with 3 being the number of patients needed to treat (NNT). In addition, the week by group interaction for the estimated proportions of response across the 8 weeks of trial was also significant (\(b=0.48, 95\%CI 0.15, 0.80, \ p=0.004\)).

For the CGI-S, the difference in irritability severity between groups at week 8 was a non-significant trend (\(b=-0.66, 95\%CI -1.34, 0.02, \ p=0.056\), ES=-1.20 95%CI -2.43-0.03). However, a significant week by treatment group interaction emerged when taking into account the difference in irritability severity across the 8 weeks of the trial (\(b=-0.11, 95\%CI -0.21, -0.01, \ p=0.029\)).

In terms of functional impairment as measured with CGAS a non-significant trend emerged both at the 8\textsuperscript{th} week of the trial (\(b=4.82, 95\%CI -0.73, 10.36 \ p=0.089\), ES=0.63 95%CI 0.06-1.20), and in the week by group interaction (\(b=0.77, 95\%CI -0.04, 1.59, \ p=0.063\)).

Groups did not differ in severity of depressive symptoms as measured with the CDRS at the 8\textsuperscript{th} week of the trial (\(b=-0.37, 95\%CI -4.77, 4.03, \ p=0.869\), ES=-0.05 95%CI -0.62-0.52), and the week by group interaction was not significant (\(b=-0.18, 95\%CI -0.84, 0.49, \ p=0.602\)). Similarly, groups did not differ in severity of anxiety symptoms as measured...
with the PARS at the 8th week of the trial (b=-1.14, 95%CI -4.20, 1.92, p=0.465, ES=-0.21 95%CI -0.78-0.36), and the week by group interaction was not significant (b=0.00, 95%CI -0.40, 0.40, p=1.000)

5.3.4 Tolerability and adverse effects

Table 5.3 shows rates of adverse effects in each of the treatment groups. No differences were found in any adverse effect between treatment groups. In addition, there were no differences in total number of adverse effects reported between groups (CTP+MPH, Mean=14.3 SD=7.1; PBO+MPH, Mean=11.5 SD=6.1; t(47)=1.50, p=0.138). Whereas participants were asked directly about suicidal ideation during the trial, only one participant receiving CTP+MPH reported passive suicidal ideation on week 6. However, the participant indicated that this ideation was no longer present at week 7 and 8. No trial participant exhibited hypomanic or manic symptoms. The most commonly reported effects in the citalopram group were changes in appetite (100%), anger (83%), and aggression along with insomnia and intrusiveness (all three 74%). In the placebo group, the most common reported adverse effects were insomnia (92%), anger (85%), and changes in appetite (81%).
### Table 5.3. Adverse effect rates of youth randomly allocated to citalopram or placebo

<table>
<thead>
<tr>
<th></th>
<th>MPH + CTP (n=23)</th>
<th>MPH + PBO (n=26)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>N</strong>:</td>
<td>N</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td><strong>Gastrointestinal system</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dry Mouth</td>
<td>3 13</td>
<td>4 15</td>
<td>1.000</td>
</tr>
<tr>
<td>Drooling</td>
<td>2 9</td>
<td>0 0</td>
<td>0.215</td>
</tr>
<tr>
<td>Increased thirst</td>
<td>3 13</td>
<td>3 12</td>
<td>1.000</td>
</tr>
<tr>
<td>Trouble swallowing</td>
<td>2 9</td>
<td>3 12</td>
<td>1.000</td>
</tr>
<tr>
<td>Nausea</td>
<td>10 43</td>
<td>8 31</td>
<td>0.390</td>
</tr>
<tr>
<td>Vomiting</td>
<td>5 22</td>
<td>3 12</td>
<td>0.448</td>
</tr>
<tr>
<td>Stomach pains</td>
<td>13 57</td>
<td>16 52</td>
<td>0.777</td>
</tr>
<tr>
<td>Bloated abdomen</td>
<td>3 13</td>
<td>0 0</td>
<td>0.096</td>
</tr>
<tr>
<td>Changes in stool</td>
<td>8 35</td>
<td>6 23</td>
<td>0.528</td>
</tr>
<tr>
<td><strong>Cardiovascular system</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change in heart rate fast/slow</td>
<td>1 4</td>
<td>2 8</td>
<td>1.000</td>
</tr>
<tr>
<td><strong>Central nervous system</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drowsiness morning/afternoon</td>
<td>7 30</td>
<td>3 12</td>
<td>0.157</td>
</tr>
<tr>
<td>Dizziness</td>
<td>5 22</td>
<td>1 4</td>
<td>0.086</td>
</tr>
<tr>
<td><strong>General well-being</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tiredness/fatigue</td>
<td>12 52</td>
<td>9 35</td>
<td>0.257</td>
</tr>
<tr>
<td>Insomnia</td>
<td>17 74</td>
<td>24 92</td>
<td>0.125</td>
</tr>
<tr>
<td>Early morning awakening</td>
<td>3 13</td>
<td>3 12</td>
<td>1.000</td>
</tr>
<tr>
<td>Appetite changes</td>
<td>23 100</td>
<td>21 81</td>
<td>0.052</td>
</tr>
<tr>
<td>Weight changes</td>
<td>13 57</td>
<td>13 50</td>
<td>0.776</td>
</tr>
<tr>
<td>Fever</td>
<td>1 4</td>
<td>3 12</td>
<td>0.612</td>
</tr>
<tr>
<td><strong>Genitourinary system</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enuresis</td>
<td>1 4</td>
<td>2 8</td>
<td>1.000</td>
</tr>
<tr>
<td><strong>Head and neck</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eyes sensitive to light</td>
<td>1 4</td>
<td>2 8</td>
<td>1.000</td>
</tr>
<tr>
<td>Headache</td>
<td>11 48</td>
<td>12 46</td>
<td>1.000</td>
</tr>
<tr>
<td>Ringing or buzzing in ears</td>
<td>0 0</td>
<td>2 8</td>
<td>0.491</td>
</tr>
<tr>
<td>Runny nose</td>
<td>6 26</td>
<td>7 27</td>
<td>1.000</td>
</tr>
<tr>
<td><strong>Haematological changes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Easy bruising</td>
<td>2 9</td>
<td>2 8</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Note: Table only shows adverse events that were reported by more than one participant.
Table 5.3. Adverse effect rates of youth randomly allocated to citalopram or placebo (Continued)

<table>
<thead>
<tr>
<th></th>
<th>MPH + CTP (n=23)</th>
<th>MPH + PBO (n=26)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
</tr>
<tr>
<td><strong>Mental status changes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anger</td>
<td>19</td>
<td>83</td>
<td>22</td>
</tr>
<tr>
<td>Aggression</td>
<td>17</td>
<td>74</td>
<td>12</td>
</tr>
<tr>
<td>Nervousness</td>
<td>15</td>
<td>65</td>
<td>13</td>
</tr>
<tr>
<td>Clingy/separation anxiety</td>
<td>13</td>
<td>57</td>
<td>8</td>
</tr>
<tr>
<td>Loss of interest/apathy</td>
<td>7</td>
<td>30</td>
<td>6</td>
</tr>
<tr>
<td>Poor concentration</td>
<td>9</td>
<td>39</td>
<td>8</td>
</tr>
<tr>
<td>Confusion</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Disorganised thinking</td>
<td>2</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Memory loss</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Speech changes</td>
<td>7</td>
<td>30</td>
<td>6</td>
</tr>
<tr>
<td>Intrusiveness</td>
<td>17</td>
<td>74</td>
<td>15</td>
</tr>
<tr>
<td>Paranoia</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td><strong>Musculoskeletal system</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Restless/Inability to sit still</td>
<td>16</td>
<td>70</td>
<td>19</td>
</tr>
<tr>
<td>Stiffness/Involuntary movements</td>
<td>2</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>Extreme stiffness/lack of movement</td>
<td>2</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>Muscle twitches</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Motor tics</td>
<td>8</td>
<td>35</td>
<td>11</td>
</tr>
<tr>
<td><strong>Pulmonary system</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficulty breathing</td>
<td>3</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td>Severe or chronic cough</td>
<td>2</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Wheezing</td>
<td>2</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td><strong>Skin changes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rash/itch</td>
<td>8</td>
<td>35</td>
<td>5</td>
</tr>
<tr>
<td>Increased sweating</td>
<td>2</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Dry skin</td>
<td>2</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Hair changes</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Acne</td>
<td>4</td>
<td>17</td>
<td>1</td>
</tr>
<tr>
<td>Picking at skin or nails</td>
<td>5</td>
<td>22</td>
<td>2</td>
</tr>
</tbody>
</table>

Note: Table only shows adverse events that were reported by more than one participant.
5.4 Discussion

5.4.1 Summary of findings

In this study, we report the findings from the first RCT with an SRI for the treatment of chronic severe irritability in youth. Specifically, we examined the outcomes of an 8-week, flexible-dose, double-blind, placebo-controlled trial of citalopram as an add-on to open-label methylphenidate. The results of the trial provide some support for the efficacy of citalopram in the treatment of irritability in children with SMD (most of them met the criteria for DMDD as well) who were already receiving methylphenidate.

Specifically, significant improvements in irritability were seen in those youth randomly assigned to citalopram prescription compared to youth assigned to placebo. However, there were no differences in secondary measures including overall functional impairment, depressive symptoms, and anxiety symptoms. Finally, no differences were found in terms of tolerability and adverse effects between citalopram and placebo.

5.4.2 Effects of adjunctive citalopram on irritability

As expected, open-label treatment with MPH led to a significant reduction in irritability with effect sizes consistent with what has previously been reported (Blader et al., 2010). This was especially true for temper outburst and mood between outburst, in contrast to hyperarousal symptoms. The latter, though, showed a large decreased during the open trial with MPH, probably explained by the overlap with ADHD symptoms.

Changes in the primary outcome of the trial suggest the efficacy of citalopram over placebo. That is, at Week 8, a significantly higher percentage of patients on citalopram improved compared to those who received placebo. This was also true when taking all time
points into account. At Week 8, the reduction in the continuous score of irritability severity did not reach significance (p=0.085), although the week by group interaction across the 8 weeks did (p=0.046). The differentiation between citalopram and placebo appeared to emerge around Week 5 (Between-group differences were: Week 5, p=0.046; Week 6, p=0.021 and Week 7, p=0.007), which was the point at which participants had returned home (while continuing in blinded outpatient treatment at our site) and was roughly two weeks after reaching 20 mg/day of drug. This time course is consistent with the timing of action of SRIs (Henssler, Kurschus, Franklin, Bschor, & Baethge, 2018). Importantly, the confidence intervals of the effect sizes vary from relatively small to large, indicating considerable uncertainty about the magnitude of the effect.

We note that the magnitude of the response to SRIs in this study was relatively low compared to that in anxiety and depression trials. In addition, the placebo response rate was also extremely low (i.e., about 5%). Together, these two observations explain the very favourable number needed to treat (NNT) of three. However, whilst the response rate to active medication may appear low in comparison to effects observed in non-add-on trials, it is consistent with what is observed in other add-on trials. For example, the remission rate in STAR*D for add-on treatment was 30% (Rush, 2007; Trivedi et al., 2006) while in CO-MED (Rush et al., 2011), it was 38%. Similarly, in a meta-analytic study, add-on treatment for ADHD showed smaller effect sizes than monotherapies (Hirota, Schwartz, & Correll, 2014). Thus, it is unclear whether the relatively low response rate to SRI treatment (and placebo) in irritability in this trial, compared to those for depression or anxiety, is due to differences in the target phenotype or in trial design.
5.4.3 Effects of adjunctive citalopram on functional impairment, anxiety and depressive symptoms

At the end of the trial, there were no differences in the CGAS scores between groups with participants still showing moderate functional impairment after completion (or severe in at least one area, range 40-50). This suggests that improvements in irritability might not have translated into reductions in overall impairment at home, at school or with peers, within the time scale of the trial. This interpretation would be in keeping with what has been described previously, in that it is common to see a dissociation between symptoms and impairment levels (Stringaris & Goodman, 2013). In addition, youth with SMD that no longer meet the criteria for the disorder in longitudinal studies often still show substantial functional impairment (Deveney et al., 2015). It is important to note, that the CGAS focuses on the lowest level of function during the period under consideration and thus may be suboptimal as an outcome measure (Shaffer et al., 1983) for disorders like DMDD. Since the functioning of a chronically irritable child during the previous week could include even short bursts of dysregulation, the CGAS might be expected to lag behind a measure of modal functioning.

We also observed no significant differences in depressive symptoms as measured with CDRS scores. We note that depressive symptoms were low – and unchanged- throughout the trial for most individuals, which is not surprising given that individuals with ongoing MDD were ineligible. In addition, no significant differences were observed for anxiety symptoms as measured with PARS either. PARS scores were in the mid-range and barely changed in either group during the trial. It is well known that irritability frequently overlaps with depression and anxiety in patients and that they share etiological underpinnings.
(Savage et al., 2015; Stringaris et al., 2012b). Our results suggest that improvements in irritability with SRI treatment may be seen in the absence of change in depression and anxiety symptoms.

5.4.4 Tolerability and side effects

The rate of side effects did not differ between citalopram and placebo; however, the absolute level of side effects was high in both groups. This may be a consequence of directly asking about side effects and the detailed observations that trained nursing staff offered during the inpatient phase of the study. In addition, nurses and parents were instructed to rate what they observed without reference to changes from baseline irritability or time of day, and thus even low levels of severity were reported in the checklist. Only one patient was withdrawn by experimenters due to adverse effects, specifically, at week 7 due to elevated liver function tests which normalised with methylphenidate discontinuation. We observed no hypomania or mania in any participant at any phase of the trial. This is consistent with previous findings from studies showing that chronic irritability is not a presage of bipolar disorder (Althoff et al., 2016; Axelson et al., 2012; Copeland et al., 2014; Stringaris et al., 2010a) and that chronically irritable youth are not at elevated risk for manic switches. However, the number of participants exposed to citalopram in this trial was too small to draw firm conclusions about the risk for medication-induced mania in this population.

5.4.5 Strengths and limitations

The current study has several strengths. First, this trial employed a rigorous double-blind, placebo-controlled approach in a well-phenotyped sample of youth to investigate the effects
of an SRI in youth with DMDD. Second, the study employed an initial open-label stimulant optimisation phase which resembles common clinical practice and reflects a rational approach to the treatment of irritability in the presence of ADHD (Fernandez de la Cruz et al., 2015; Winters et al., 2018). Third, the design maximised human subjects’ protections by using hospitalisation for safety and standardising treatment and environmental care prior to randomisation. Finally, we employed a robust statistical approach to account for unobserved data which yielded results that were replicated in the sensitivity analysis using the last observation carried forward method.

The findings of the current study should also take into account several limitations. First, the sample size was smaller than initially planned. In the 10-year duration of the study, participants were pursued through nation-wide recruitment using local resources, sending letters to every practising child and adolescent psychiatrist in the American Academy of Child and Adolescent Psychiatry, and giving local talks and presentations across the country. While the initial response to the study was positive, over time, the response declined steadily, and a corresponding increase in the use of SRIs for chronic irritability took place in clinical practice. Yet, our experience in recruitment was not unique; other studies have had the same problems (McGough, 2012; McGough, 2014). There are several potential explanations for this, including the increasing familiarisation of community practitioners with prescribing for youth with chronic irritability. Nonetheless, we note that at the current sample size, the study had a power of approximately 75% to detect a difference at $\alpha=0.05$. Clearly, further studies with a larger sample would be warranted. A second limitation is in the measures of change in irritability. Indeed, there are still no validated self- or parent-report scales for assessing change (as opposed to trait measurement
of irritability such as the ARI (Stringaris et al., 2012a). A third limitation is that the findings in this study may not be generalisable to all children with DMDD since we initially recruited a population with SMD, which requires hyper-arousal symptoms (e.g., insomnia, agitation, distractibility) and might represent a more severe subgroup of the DMDD population. Furthermore, there are limitations of generalising from subjects coming to research at NIMH to other populations because of the design of the study. Indeed, the length of the study, the compulsory inpatient interval, and the requirement for the patients to come off all psychotropic medication were all obstacles to participation which may further limit the representativeness of our participant sample. Finally, it is unclear what the effect of citalopram over placebo would be if youth with irritability had no ADHD and, consequently, were not treated with stimulant medication. Future studies employing larger samples of youth with DMDD and lower rates of ADHD should examine the independent effects of citalopram on chronic severe irritability.

5.5 Conclusion

Citalopram appears to be an efficacious treatment for chronic irritability with hyperarousal symptoms for individuals who are unresponsive to stimulant medication alone. However, the effects of add-on citalopram on functional impairment and other mood symptoms are limited.
CHAPTER 6: General Discussion

This thesis examined potential mechanisms that might explain the specific association between irritability and depression and tested a new treatment approach motivated by this association. Here I will summarise the findings of each study, discuss whether the aims and hypotheses were met, and describe methodological commonalities and disagreements between studies. Then I will discuss the limitations of the current thesis before moving on to draw a potential comprehensive model that might explain the association between chronic severe irritability and future depressive disorder. Finally, I will close this thesis by proposing some directions for future research.

6.1 Summary of findings by study

6.1.1 Longitudinal correlates of irritability

The overall aim of the study in Chapter 2 was to quantitatively test the association between chronic severe irritability, defined either as DMDD/SMD-like irritability or an irritable dimension of ODD, and future psychopathology, either in the form of psychiatric disorders or psychiatric symptoms. To do so, we systematically searched studies that examined irritability as a predictor of future psychiatric outcomes and meta-analysed the effect estimates using a random-effects model. We expected to find specific associations between irritability and future anxiety and depression disorders and symptoms based on several individual studies reporting this relation. We also examined the extent to which associations between irritability and psychiatric outcomes other than depression and anxiety had been tested. Finally, we examined the existence of publication bias, sources of variation across studies and associations with functional outcomes.
As expected, we found a significant association between chronic severe irritability and future depressive and anxiety disorders, with an odds ratio of 1.85 and 1.58 (both $p<0.001$), respectively. There was also an association with ODD (OR=1.57, $p<0.001$), probably due to the homotypic continuity of ODD irritability dimension or due to the overlap of items to measure both irritability and ODD. In contrast to the prediction of psychiatric disorders, associations between irritability and future psychiatric symptoms were significant for depressive, anxiety, CD and ODD symptoms. That is, the specific relationship between irritability and depressive and anxiety disorders was not replicated at the symptom level.

The association between irritability and internalising and externalising disorders have been examined nearly the same number of times, which makes robust the results for psychiatric disorders. However, this was not true for psychiatric symptoms, in which internalising symptoms have been examined in more analyses than externalising symptoms.

Test of publication bias suggested an absence of bias for the association between irritability and anxiety disorders and symptoms. However, the results suggested a bias toward reporting associations with larger effect sizes between irritability and depressive disorder (i.e., an association that remained significant after removing outliers), and a bias towards publishing significant associations between irritability and depressive symptoms. Finally, data for longitudinal studies also showed that chronic severe irritability is associated with functional impairment on several levels, including suicidality, even when controlling for baseline psychopathology.

The results of this meta-analysis can be taken as the starting point from which the remaining studies of this thesis depart. The following two studies examined mechanisms
that might explain the specific association between irritability and depression. The last study, motivated by this association, tested whether chronic severe irritability might be responsive to adjunctive antidepressant treatment.

6.1.2 Early stress and irritability

As found in Chapter 2, chronic severe irritability is a specific predictor of depressive disorder. Therefore, it can be said that chronic irritability is a risk factor of depression (Vidal-Ribas et al., 2016). Most cases of depression have their onset during adolescence, and rates of depression at this age are twice as much in females as in males (Angold & Rutter, 1992). Then, it is possible that females have a predisposition to chronic irritability in childhood that increases their chances of developing depression in adolescence. Indeed, some studies have found that levels of irritability are higher in girls during childhood and adolescence (Riglin et al., 2017; Stringaris et al., 2012b; Trepat & Ezpeleta, 2011).

The study in Chapter 3 aimed to test whether stress reactivity, as measured with respiratory sinus arrhythmia (RSA) in infancy, predicted distinct patterns of ODD symptoms in boys and girls in the pre-school years. Specifically, based on previous findings (Hinnant & El-Sheikh, 2013; Morales et al., 2015), we hypothesised that increased vagal reactivity in girls and decreased vagal reactivity in boys would be associated with later ODD symptoms. And most importantly, we hypothesised that in girls the association would be with the risk for later depression, namely the irritability dimension, and in boys with risk for later disruptive disorders, namely the headstrong dimension.

The results of the study supported the first hypothesis, in that higher stress reactivity predicted higher levels of ODD symptoms in girls but lower levels in boys. In contrast, the
second hypothesis was not supported by the results. That is, higher stress reactivity predicted higher levels of both ODD irritability and ODD headstrong symptoms in girls and lower levels on both ODD dimensions in boys.

6.1.3 Cognitive mechanisms in irritability

The study described in Chapter 4 of this thesis examined how deficits in facial and vocal emotion recognition were related to depressive symptoms, both cross-sectionally and longitudinally, in a clinical sample of youth with DMDD. Deficits in emotion recognition have been associated with depression (Dalili et al., 2015) and irritability (Deveney et al., 2012; Guyer et al., 2007), and might act as a risk factor for mood disorders (Beevers & Carver, 2003; Vrijen et al., 2016). Therefore, it can be assumed that deficits in emotion recognition might play a role in the transition between irritability and depression (Brotman et al., 2017; Vidal-Ribas et al., 2016) found in Chapter 2 of this thesis.

Based on previous reports, we first hypothesised that youth with DMDD and high depressive symptoms would be more likely to interpret positive stimuli (i.e. happy faces and voices) as more negative (i.e., sad, fearful and angry) than those with DMDD and low depressive symptoms, and healthy controls (HC). Second, we further hypothesised that the misidentification of positive stimuli as negative would be longitudinally associated with higher depressive symptoms independently of baseline symptoms.

As hypothesised, we found that depressive symptoms in youth with DMDD were associated with impaired recognition of happy stimuli across modalities (i.e., faces and voices), and with the misinterpretation of happy stimuli as fearful and angry, independent of irritability symptoms. Furthermore, in a subsample of DMDD youth, we found that
misinterpreting happy stimuli as negative predicted higher levels of depressive symptoms at follow-up approximately one year later, independently of baseline depressive and irritability symptoms.

In addition, we also examined the effects of comorbid anxiety disorders, age, and sex on emotion recognition. Specifically, we hypothesised that deficits in emotion recognition would be independent of comorbid anxiety disorders and increased in younger youth and males. As expected, we found that having a comorbid anxiety disorder was not associated by itself to deficits in emotion recognition and that the ability to recognise facial and vocal emotions improved with age and was better in females.

### 6.1.4 Treatment response of irritability

The study in **Chapter 5** of this thesis reports the results of the second pharmacological randomised controlled trial (RCT) in typically developed youth with chronic severe irritability. Specifically, using a double-blind design, we examined the effect of adding citalopram, an SRI, versus adding placebo on reducing irritability in youth with SMD treated with stimulant methylphenidate. The addition of citalopram to stimulant medication was motivated by the specific association between irritability and internalising disorders, commonly treated with SRI, seen in **Chapter 2**. In addition, SRIs, especially citalopram, have been shown to reduce irritability in open-label interventions (Kim & Boylan, 2016). Of note, open-label stimulant medication was prescribed prior to randomisation as an optimisation lead-in phase. This was motivated by the high rates of ADHD seen in youth with chronic severe irritability in clinical samples (Deveney et al., 2015), and the results of post-hoc analyses of data (Fernandez de la Cruz et al., 2015) and several open-label trials in
youth with ADHD and comorbid DMDD (Baweja et al., 2016; Waxmonsky et al., 2016; Winters et al., 2018). A further aim of the trial was to examine the tolerability of adjunctive citalopram versus adjunctive placebo.

Our primary hypothesis was that add-on citalopram would be associated with a greater improvement in irritability than add-on placebo, as measured by the CGI. We also hypothesised that add-on citalopram, as compared with add-on placebo, would be associated with a greater improvement in overall functional impairment, anxiety symptoms, and depressive symptoms.

Our primary hypothesis was supported by the results of the trial. Rates of response in SMD youth assigned to citalopram (35%) were significantly larger than rates of response in youths assigned to placebo (6%). Of note, overall response in both cases was low, but still in line with what is observed in add-on trials (Rush, 2007; Rush et al., 2011; Trivedi et al., 2006).

Regardless of the differences in overall improvements of irritability across groups, we did not find differences in functional impairment, depressive symptoms, and anxiety symptoms. Rates of adverse effects did not differ between intervention groups, although endorsement of adverse effects was high in both groups. However, no participant reported hypomanic or manic symptoms at any phase of the trial.

6.2 Methodological commonalities and disagreements

The studies presented in this thesis differ substantially in their methodology. This thesis includes a meta-analysis, a longitudinal modelling study in a large community-based
sample, an analysis of behavioural data in a clinical sample, and a randomised controlled trial also in a clinical sample. However, these studies share some common approaches to studying irritability. In this section, I discuss common and distinct methodological characteristics that might affect the comparability and interpretation of the studies of this thesis.

First, all four studies in this thesis used, to some extent, a longitudinal approach. The meta-analysis in Chapter 2 pooled estimates of longitudinal associations between irritability and psychiatric outcomes. Chapter 3 examined the precursors of ODD dimensions, including irritability. Chapter 4 explored how deficits in emotion recognition predicted depressive symptoms at follow up. Finally, in Chapter 5, treatment response to add-on citalopram vs. placebo was assessed in an 8-week RCT in youth with DMDD who had been prescribed stimulant medication. Nevertheless, the follow-up period varied substantially across studies. For example, follow-up periods of studies included in the meta-analysis in Chapter 2 ranged from 6 months to 6 years, with one study following participants up to 20 years (Stringaris et al., 2009). Chapter 3 had a maximum follow-up period of 4.5 years whereas Chapters 4 and 5 had average follow-up periods of 1 year and 8 weeks, respectively.

Second, whereas all studies examined irritability in young people, the ages at which irritability was examined also varied. Mean ages at baseline in the meta-analysis in Chapter 2 ranged from 3 to 15 years. Irritability in Chapter 3 was assessed at ages 2.5, 3.5 and 5 years. And ages in Chapters 4 and 5 ranged 8-20 years and 7-15 years, respectively. Whereas we did not find an effect of age in the longitudinal correlates of irritability in
Chapter 2, the interpretability of these results was limited by the few numbers of studies analysed (see section 6.3). It is still unclear whether the irritability seen in early childhood is the same as the irritability seen in adolescence or even adulthood. As mentioned in Chapter 3, some recent studies suggest that irritability in early life might be more associated, genetically and phenotypically, with developmental disorders such as ADHD, whereas adolescent irritability might be more related to affective disorders (Riglin et al., 2017; Riglin et al., Under review). These findings are supported by a genetic study that showed that the higher genetic covariance between irritability and depression peaks in early adolescence (74%) (Savage et al., 2015).

Third, all but two studies in the meta-analysis in Chapter 2 employed community-based samples, the same type of sample as in Chapter 3. In contrast, the studies in Chapters 4 and 5 were conducted in clinical samples of youth with DMDD/SMD. It is still unclear whether the correlates of irritability differ between clinical and community-based samples. However, some data have suggested that comorbidity of externalising disorders such as ODD and especially ADHD is increased in clinical samples. For example, rates of ODD and ADHD were close to 90% in a sample of 200 youth with SMD (Deveney et al., 2015). Similar rates of ADHD were found in the clinical samples of Chapters 4 and 5. However, epidemiological studies in community-based samples found ADHD rates of 6-30% in youth with DMDD (Copeland et al., 2013). The higher rates of comorbid externalising disorders in clinical samples might be explained by the major burden caused by these disorders in contrast to depressive and anxiety disorders, which might make parents be more motivated to seek clinical services and advice.
Fourth, as mentioned in Chapter 1, this thesis has focused on chronic severe irritability, which was defined either as an irritable dimension of ODD or as DMDD/SMD. The study in Chapter 2 employed both definitions, the study in Chapter 3 employed the former, and the studies in Chapter 4 and 5 employed the latter. However, to my knowledge, no study has systematically compared DMDD and the ODD irritability dimension, with the exception of the meta-analysis in Chapter 2 (Vidal-Ribas et al., 2016). Although the DSM-5 does not allow the diagnosis of ODD if DMDD criteria are met, there are some studies that have examined the overlap between these two disorders. As mentioned in section 1.3.1.5, rates of ODD are amongst the highest in youth with DMDD, even in community-based samples, with a prevalence over 60% (Althoff et al., 2016; Copeland et al., 2013). Indeed, some authors have suggested that it is extremely difficult to differentiate between DMDD and ODD in the general population given the high overlap of symptoms (Mayes et al., 2016). This would suggest that both types of chronic severe irritability are very similar, if not the same. However, this overlap also suggests that chronic severe irritability without the presence of oppositional and defiant behaviour is rare, even in DMDD. Nevertheless, most of these studies have employed the same items to measure irritability criteria to ascertain DMDD and ODD symptoms/diagnoses, or at least instruments that were not designed to capture DMDD. Therefore, the overlap between DMDD and ODD might be inflated by item overlap in community samples. In Chapter 4 and 5, where the K-SADS-PL DMDD module was used to ascertain DMDD, rates of ODD were close to 55% and 80%, respectively. However, given the small number of participants and the selection biases that occur in these samples (see section 6.3), it is unclear whether the findings of these studies would be representative of the general population. Unfortunately, the only
existing study that has used a measure specifically designed to ascertain DMDD in a community-based sample did not provide rates of comorbid ODD (Munhoz et al., 2017).

Fifth, related to the previous point, studies differed in the instruments employed to measure irritability. In Chapter 3 we employed items from the preschool version of the Child Behavior Checklist (CBCL)(Achenbach, 1991) to create a continuous measure of ODD irritability dimension; specifically, we employed the items “Angry moods”, “Stubborn, sullen or irritable”, and “Temper tantrums or hot temper”. In Chapters 4 and 5, instead, we employed the SMD/DMDD module of the Kiddie Schedule for Affective Disorders Present and Lifetime Version (K-SADS-PL) (Kaufman et al., 1997) to ascertain a categorical diagnosis of chronic severe irritability characterised by frequent temper outbursts and persistent irritable mood between outburst. In addition, in Chapter 4 we employed the Affective Reactivity Index (ARI), which provides a dimensional measure of irritability (Stringaris et al., 2012a), with items such as “Easily annoyed by others”, “Often loses temper”, or “Is angry most of the time”, among others. Although these measures are different, the wording used to ascertain irritability manifestations are very similar. Indeed, a recent study employing both the ARI and the CBCL irritability scale found moderate correlations between the two measures at two different time points (r’s=0.64 and 0.54, p's<0.001) (Tseng et al., 2017).

Finally, some studies in this thesis examined sex differences in their outcomes. Specifically, Chapter 3 examined how stress reactivity was differently associated with future ODD symptoms, and Chapter 4 examined sex differences in emotion recognition. Unfortunately, the number of studies meta-analysed in Chapter 2 was too small to reliably
examine sex differences in the prediction of psychiatric outcomes. However, we have tested whether the proportion of females in the sample, which ranged from 0% to 100%, was associated with the estimated effect sizes. We found no significant effect of this variable on psychiatric outcomes, but this should be interpreted with caution given the small number of studies. In Chapter 5, sex differences in treatment response could not be tested due to the small number of observations in some cells (e.g., only 6 females in the placebo group).

6.3 Limitations of the studies

The results presented in the studies of this thesis should be interpreted in light of several methodological limitations. Most of these limitations have been discussed at the end of each chapter. However, here I highlight the most common and relevant limitations across studies.

First, probably the most concerning limitation in this thesis is the limited number of observations in most of the studies, which limited substantially the analyses that could be done and the interpretation of results. Except for the study in Chapter 3, in which we employed a large community-based sample, all other studies fell short in their unit of analyses. For example, the number of studies for each psychiatric outcome in the meta-analyses in Chapter 2 was very small and we resorted to aggregate together different disorders into groups of internalising and externalising disorders to increase the power of this study. Despite this attempt, the results of subgroup analyses and meta-regressions examining sources of between-study variance were still very hard to interpret. The small number of studies in each subgroup made the results more prone to Type I and Type II
errors. Similarly, the number of observations was also very small in some groups in the secondary analyses carried out in Chapter 4 that examined the effects of sex, age, and anxiety disorders on emotion recognition. Lastly, in the trial presented in Chapter 5, the small number of participants in each group decreased the power for detecting significant differences as the wide confidence intervals in outcomes of interest were more likely to overlap between groups.

Second, related to the previous point, all studies suffered from attrition in the follow-up analyses. The initial sample of the study in Chapter 3 consisted of 1,233 children that were reduced to 770 children at the 5-years follow-up (retention of 62%). In addition, data on vagal reactivity was only collected in an intensive sample as per design. However, the sample design and attrition were accounted for in the estimates of descriptive statistics by the use of Full Information Maximum Likelihood (FIML). This enabled participants with incomplete observations to be included under the assumption of missingness being missing-at-random (MAR) (Graham, 2003). The estimates obtained with FIML account for the sample stratification and for attrition associated with both covariates and observed values of variables of interest. This method is currently regarded as the “state of the art” missing data technique (Schafer & Graham, 2002) given that it requires less strict assumptions about the mechanism that leads to missing data than listwise deletion and generally produces more accurate estimates than traditional missing data handling techniques (e.g., discarding cases).

The longitudinal analysis in Chapter 4, in which we examined how emotion recognition predicted depressive symptoms, was done in a 33% of the initial sample. However, no difference was found in baseline age, sex, race, ethnicity, ARI scores, CDI scores, or task
performance between those participants who had data at follow-up and those who were lost to follow-up. Nevertheless, other unmeasured confounders might have accounted by this attrition. For example, some participants might have taken part in other treatment trials in the NIMH, and then decided to withdraw after completing such treatments. Some others might have received treatment in the community, and perhaps stop attending follow-up assessments after improving. The high rates of attrition in this study have also prevented us from testing the moderating influences of sex on how deficits in emotion recognition predicted depressive symptoms. Of the 25 youth with DMDD who had data at follow-up, 19 were males (76%) and 6 females (24%).

In the RCT reported in Chapter 5, 8 participants of the 49 randomised into treatment groups did not complete all 8 weeks of the trial (attrition rate of 16%). However, using an intent-to-treat approach, we employed two common methods to remedy missing data in clinical trials, one of them as sensitivity analysis. Specifically, we first employed multilevel models (MLM) estimated by maximum-likelihood, which, like FIML in Chapter 3, enabled the inclusion of participants incompletely observed under the missing at random assumption. Then, we replicated the analyses employing the last observation carried forward (LOCF) method, in which missing observations were completed with the same value as the last observed value.

Third, a further methodological limitation in most studies was related to the measurement, or the lack of it, of variables of interest. In Chapter 3, for example, information on irritability was collected with an instrument (i.e., CBCL) that was not specifically developed to measure this construct. Consequently, very few items were available, thus
capturing a limited number of irritability manifestations, and some of these overlapped with the headstrong dimension of ODD (e.g., “Stubborn, sullen or irritable”). In Chapter 4, we compared three groups of participants: youth with DMDD and high depressive symptoms, youth with DMDD and low depressive symptoms, and healthy volunteers (HV). However, all dimensional analyses in both cross-sectional and longitudinal data were only conducted in youth with DMDD, because depressive symptoms as measured by the CDI were not collected in HV. It was assumed that, by definition, depressive symptoms were low in HV. However, we could have examined an interaction between depressive symptoms and group if these had been collected in HV. Finally, the study of Chapter 5 also had a limited number of instruments to collect information on primary and secondary outcomes. First, the RCT with adjunctive citalopram in youth with SMD treated with stimulant only employed a single measure to assess the treatment target of irritability (i.e., CGI). Whereas it is true that the CGI has been used as a primary outcome in other trials (Dickstein et al., 2009), it is unusual to have a single measure to assess the construct of interest. Usually, other instruments that collect information on the same construct are also employed. Indeed, the initial protocol of this trial had designated the 15-item irritability subscale of the Aberrant Behavior Checklist (ABC) (Aman et al., 1985) as primary outcome along with the CGI. However, this scale proved to be inappropriate to collect information on irritability in this population and this trial for two reasons. First, this instrument was originally developed to assess aggressive behaviours in the context of developmental disabilities. Consequently, some behaviours described in the items were not typically seen in youth with DMDD without developmental disabilities (e.g., behaviours associated with self-harm) and most items measure components of phasic irritability (e.g., “Temper tantrums”, “Yells”,
“Stamps feet”) with only two items measuring the tonic components of mood. Second, the information gathered with the ABC was hard to interpret due to the design employed during the trial. That is, both the informant of irritability (parents vs. nursing staff) and the setting (inpatient unit vs. home) under which irritability-related behaviours took place changed over time. Whereas the setting also changed in the case of the CGI, the informant always included the child, and when at home, also the parent. Information on ADHD and manic symptoms were originally planned to be collected in this trial using the Conners Teacher Rating Scale (CTRS) (Conners, 1997) and the Young Mania Rating Scale (YMRS) (Young, Biggs, Ziegler, & Meyer, 1978), respectively. However, these measures were not collected in all participants and could not be analysed. Nevertheless, in terms of response to open-label stimulant, our findings showed improvements on SMD symptoms between baseline at admission and baseline at randomisation in participants that were randomised, especially in the SMD symptom domain of hyperarousal, which overlaps substantially with ADHD symptoms. In contrast, hyperarousal symptoms did not change over the course of the RCT under citalopram or placebo. In terms of manic symptoms, although the YMRS was not collected, no participant in either group showed hypomanic or manic symptoms during the trial.

6.4 Towards a model for the relation between irritability and future depression

In this section, I will propose a model to explain the association between irritability and future depressive disorder. To do so, I will refer to both findings reported in this thesis and results from previous works mentioned in the introductory Chapter 1. The model will also include aspects that are hypothesised to play a role in this transition but still need to be
tested. Therefore, the more speculative part of the model, albeit based on theory, will serve as a basis for future research (see section 6.5).

Before defining my model, I will first describe three distinct potential explanations for the relation between irritability and future depressive disorder. I will provide evidence to refute or support each explanation and take the model with more support as a reference to build my own model.

### 6.4.1 Three potential explanations for the relation between irritability and future depression

There are three plausible but distinct explanations for the association between irritability and future depressive disorder.

First, it is possible that irritability is a developmental precursor of depression (Figure 6.1). In other words, irritability might be a sign of depression early in development that later manifests, not only in form of irritability but also in other cardinal symptoms of depression such as low mood and anhedonia.

The second possibility is that irritability causes depression. Irritability in youth could have an impact on their environment, such as school failure or social impairments with family and peers, and that might be depressogenic in its own right (Figure 6.2). This hypothesis is the basic notion of the failure model (Capaldi, 1992) which I will describe below.

Finally, the third possibility is that irritability and depression share risk factors, either genetic or environmental, that make both conditions to overlap over time (Figure 6.3).
Both irritability and depression are then affected by shared risk factors but also by their own specific risk factors. Each potential explanation is discussed separately below.

### 6.4.1.1 Model of irritability as an early manifestation of depression

In this model (Figure 6.1), irritability is a developmental precursor of depression, understanding depression as the presentation of low mood and anhedonia alongside typically associated symptoms listed in criteria B of DSM-5 (APA, 2013). That means that youth who develop depression in, for example, adolescence would have presented irritability in childhood or early adolescence.

![Figure 6.1. Model of irritability as an early manifestation of depression.](image)

If irritability were simply an early manifestation of depression one would expect that irritability decreases over time as other cardinal depressive symptoms emerge. However, evidence from longitudinal studies with multiple time points of assessment shows that levels of irritability are quite stable over time (Burke et al., 2010; Hipwell et al., 2011;
Riglin et al., 2017; Roberson-Nay et al., 2015; Savage et al., 2015). Alternatively, it is also possible that irritability, although remaining stable, transforms into depressive disorder over time, with the emergence of other depressive cardinal symptoms such as low mood and anhedonia. If that were the case, correlations between irritability and depression would also increase. However, once again, there is little evidence to suggest that this occurs; instead, correlations between irritability and emotional problems remain stable over time (Hipwell et al., 2011) or even decrease in some studies (Lavigne et al., 2014).

Finally, as discussed in section 1.4.2, irritable mood is allowed to be a cardinal symptom for depressive disorder in children and adolescents, but not in adults, according to DSM-5 criteria (APA, 2013). However, evidence shows that irritability in the absence of low mood or anhedonia is extremely rare (<6%) in young people with depression (Stringaris et al., 2013).

6.4.1.2 Model of irritability as cause of depression

Based on findings from individual studies and the meta-analysis presented in Chapter 2 of this thesis, a more plausible model is that irritability is a cause of later depressive disorder (Vidal-Ribas et al., 2016) (Figure 6.2). However, this assertion can only be made if irritability can be demonstrated as a predictor for the new onset of depressive disorders. In the meta-analysis of Chapter 2, the prediction of depressive disorder was significant for nine (82%) of the eleven studies testing this association. Among those studies, 6 (67%) adjusted their predictions by the presence of emotional disorders at baseline (Brotman et al., 2006; Dougherty et al., 2016; Dougherty et al., 2013; Leibenluft et al., 2006; Stringaris et al., 2009; Stringaris & Goodman, 2009a). This supports the idea that irritability predicts
new onset of depression. Nevertheless, it is also important to note that the strength of the association between irritability and future depressive disorder decreases substantially when adjusting for baseline emotional problems (Stringaris et al., 2009; Stringaris & Goodman, 2009a). In addition, the variance explained by baseline depressive problems is much larger than the variance explained by baseline irritability (Lavigne et al., 2014; Stringaris et al., 2012b). Of course, not all depressive disorders are preceded by irritability. Therefore, whereas irritability might cause or contribute to the development of depressive disorder, other factors, whether or not being related to irritability, can play a role in the development of depressive disorder as well.

![Figure 6.2. Model of irritability as cause of depression.](image)

If irritability causes depression, then the question that arises is how this happens. One explanation for how irritability might cause the development of depressive disorder can be found in the *failure model* (Capaldi, 1992). The failure model was formulated as an attempt...
to explain the high association seen between conduct problems and depressive disorder (Loeber et al., 2000; Zoccolillo, 1992). According to this model, the presence of conduct problems in young people creates failures in developmentally important experiences, such as school achievement and the attainment of close relationships, which in turn create vulnerability for the development of depression. Indeed, both poor academic achievement (Cole, Martin, Powers, & Truglio, 1996; McCarty, 2008) and interpersonal problems within the family and friends (Cole et al., 1996; Eley & Stevenson, 2000) have been shown to predict depressive disorder. Through a similar chain of events, irritable children could generate depressogenic environments with their behaviour that increase the risk of depression (Figure 6.2).

The patterns of associations between ODD dimensions and psychiatric outcomes, however, do not support this model. Specifically, behaviours within the headstrong and hurtful ODD dimensions such arguing and defying adults, annoying and blaming peers, and being vindictive, are also thought to be associated with academic failure and with interpersonal problems within the family and peers (Greene et al., 2002). Yet, several studies show that these behavioural dimensions of ODD are less likely to be associated with depression than the ODD dimension of irritability (Burke, 2012; Burke et al., 2010; Lavigne et al., 2014; Stringaris & Goodman, 2009a; Stringaris et al., 2012b; Whelan et al., 2013). Furthermore, as discussed in the next section, evidence coming from genetically-informative studies carried out in the last years does not support the notion that the effect of irritability on depression is explained by the generation of a depressogenic environment.
6.4.1.3 Model of shared risk factors

A third possibility that might explain the association between irritability and depression is that both constructs share the same risk factors, either genetic or environmental (Figure 6.3). As mentioned before, the association between irritability and depression is moderately strong across development (Hipwell et al., 2011), even in very young children (Lavigne et al., 2014). Therefore, it is plausible that both constructs are associated through a third variable (or a combination of variables) since early in development.

![Diagram showing model of shared risk factors between irritability and depression]

**Figure 6.3.** Model of shared risk factors between irritability and depression.

As I described in section 1.6.1, in the last years, genetically-informative studies in large twin samples have examined the causal relationships between irritability and depression. The results of these studies support the model of shared risks factors. In a sample of adolescents followed, on average, for two years, Stringaris et al. (2012b) found that the genetic association between irritability and depression at baseline accounted for the genetic
association between irritability at baseline and depression at follow up. In other words, the association between irritability and future depression was mostly explained by shared genetic factors, and to a lesser extent, by non-shared environmental factors. Similar results were found in a recent study (Mikolajewski et al., 2017).

In line with these results, Savage et al. (2015) found that shared genetic factors and non-shared environmental factors had a strong effect on the association between irritability and depression over time. Interestingly, the authors found that the impact of irritability on future emotional symptoms was larger than the impact of emotional symptoms on subsequent irritability. Moreover, this effect was especially evident from early to middle puberty (ages 13–14). Unlike other studies that were conducted in older samples, Savage et al. (2015) also found a small but significant effect of common environmental factors in the association between irritability and depression. However, the effect of these common environmental factors was only present in childhood, not in older ages.

Taken together, the finding in which irritability is linked with depression by mostly common genetic underpinnings is in line with studies showing genetic rather than environmental links between major psychiatric disorders as well as with a broad construct of negative affect and the personality trait of neuroticism (Anttila et al., 2018; Eley, 1997; Mikolajewski, Allan, Hart, Lonigan, & Taylor, 2013). The finding that this relationship is already present early in life is consistent with the links between irritability-like temperaments and later internalising psychopathology (Kiff, Lengua, & Bush, 2011a). For example, using data from the community-based sample of the Avon Longitudinal Study of Parents and Children (ALSPAC), Stringaris, Maughan, and Goodman (2010b) examined
the relation between early-temperaments of emotionality -characterised by emotion dysregulation- and activity at 38 months, and later psychiatric outcomes at 91 months. The authors found that both temperaments predicted ODD; however, whereas temperamental activity predicted comorbidity between ODD and ADHD, emotionality predicted comorbidity with internalising psychopathology.

The specific genetic variants that are common for irritability and depression are unknown. Similarly, the specific environmental factors that play a small but significant role in the association between irritability and depression are also unknown. However, it is possible that the same environmental factors discussed in section 1.6.3 also explain the association between irritability and depression. I have described in Chapter 1 how irritability is linked to negative parenting behaviours, and how negative parenting is in turn associated with later irritability (Oliver, 2015). Although it has not been tested simultaneously with irritability in genetically-informative studies, negative parenting has also been associated with an increased risk of depression in children (Kiff et al., 2011a; Oldehinkel, Veenstra, Ormel, de Winter, & Verhulst, 2006), with irritable temperament increasing the depressogenic effect of overprotection and lack of emotional warmth (Oldehinkel et al., 2006).

In summary, of the three potential models described here, the model with shared risk factors between irritability and depression is the one with more supporting evidence. Although irritability is closely linked to the development of depression, evidence from twin studies suggests that this relationship is explained mostly by shared genetic risk and, to a lesser extent, environmental factors. As I will describe in the next section, where I draw a
tentative model for the association between irritability and depression, the shared genetic mechanisms between both constructs may include affective and reward processing mechanisms.

6.4.2 Proposed model for the relationship between irritability and future depressive disorder

In this section, I will provide a hypothesised and parsimonious model to account for the relationship between irritability and future depressive disorder. Based on the evidence to date, this hypothesised model will be based on the shared factors model described in the previous section. According to the shared factors model, factors contributing to irritability are shared with those contributing to depression; therefore, I will extend the model displayed in Figure 1.7, which display the pathophysiological mechanisms of irritability, to explain the association between irritability and depression.

My model will rely on three sources of information. First, the model will include factors described in section 1.6, and of those, the ones that have been found to be more consistently associated with irritability. The second source of information will be the studies of this thesis. Finally, the third source of information will come from the literature on depression mechanisms, especially the factors related to aberrant reward processing.

The hypothesised model to explain the association between irritability and depressive disorder is displayed in Figure 6.4.
Figure 6.4. Hypothesised model of the relationship between irritability and the development of depressive disorder.

I will first briefly discuss genetic and environmental factors associated with irritability and depression. Second, I will discuss how aberrant response to emotional stimuli might play a role in the overlap between irritability and depression. Finally, based on a large body of literature on depression and some findings on irritability, I will hypothesise that aberrant response to reward is also a pivotal mechanism by which irritability might be predictive of depressive disorder. Of note, aberrant response to emotional stimuli and reward might be in turn affected by shared genetic and environmental factors with irritability and depression.

6.4.2.1 Genetic and environmental risk factors

As discussed in sections 1.6.1 and 6.4.1.3, several studies have shown that the overlap between irritability and future depressive disorder is mostly explained by common genetic variance (Mikolajewski et al., 2017; Rappaport et al., 2018a; Savage et al., 2015; Stringaris et al., 2012b) with the remaining variance explained by non-shared environmental factors. However, shared environmental factors are also thought to play a role during childhood but
not older ages (Savage et al., 2015). It is plausible to think that common environmental factors also include exposure to negative parenting styles, which have been shown to have an impact in both the development of irritability (Oliver, 2015) and depressive symptomatology (Kiff et al., 2011a). Moreover, parenting styles also interact with irritable early-temperaments (Kiff et al., 2011b) giving place to irritability (Belsky, 1984; Dougherty et al., 2014; Lengua & Kovacs, 2005; Schneider et al., 2018) and depressive symptoms (Kiff et al., 2011a; Oldehinkel et al., 2006).

Another consistent finding, related to both genetics and environment, is that children with chronic irritability are more likely to have a family history of depression (Krieger et al., 2013; Munhoz et al., 2017; Propper et al., 2017; Wiggins et al., 2014). Of course, this is also the case in youth with depressive disorder (Klein, Lewinsohn, Rohde, Seeley, & Olino, 2005; Weissman, Berry, Warner, & et al., 2016). One study even suggested that the relationship between maternal history of depression and adolescent depression is partly mediated by the presence of irritability in childhood (Whelan et al., 2015).

In Chapter 3, we tested the hypothesis that differences in vagal reactivity to a stress challenge would predict irritability symptoms in girls and headstrong symptoms in boys. This hypothesis was motivated by the different patterns of disorders and ODD symptoms, including levels of irritability, seen in males and females (Riglin et al., 2017; Rowe et al., 2010; Silva et al., 2014; Stringaris et al., 2012b; Trepat & Ezpeleta, 2011), though some studies have not found sex differences (Copeland et al., 2015; Ezpeleta & Penelo, 2015). In our study, this hypothesis was not supported. Therefore, the variable sex seems not relevant for the transition between irritability and depression, and it is not included in the model. In
any case, early stress reactivity is associated and interacts with infant temperament to predict emotion regulation strategies and depressive symptoms (Gentzler et al., 2009; Gunnar, Porter, Wolf, Rigatuso, & Larson, 1995; Santucci et al., 2008). It remains to be tested, however, whether reactivity to stress early in life might be associated with the development of irritability, and consequently, depression later in life.

### 6.4.2.2 Aberrant response to emotional stimuli

The specific pathways through which deficits in emotion processing in youth with irritability lead to depression are still unclear. Longitudinal genetically-informative studies examining the three constructs will be needed to answer this question. One possibility can be that deficits in emotion processing put children at higher risk of irritability and, consequently, of depression. The second possibility is that emotion processing deficits represent an increased risk for both irritability and depression. Since both conditions are associated with emotion processing deficits, the second possibility is thus more plausible and is included in my model (Figure 6.4).

As described in section 1.6.2.2, one of the main pathophysiological mechanisms of irritability is an aberrant response to threatening stimuli (Brotman et al., 2017). Specifically, youth with severe irritability show attentional bias towards threatening angry faces (Hommer et al., 2014; Salum et al., 2017) and are more likely to interpret ambiguous or neutral faces as more threatening (Brotman et al., 2010; Stoddard et al., 2016). Similar attentional biases towards negative stimuli are also evident in people with depression (Armstrong & Olatunji, 2012; Gotlib et al., 2004; Leppanen, 2006; Peckham et al., 2010).
In addition, as I also described in Chapter 4, youth with severe irritability also have a generalised deficit in emotion recognition, both in faces (Guyer et al., 2007; Kim et al., 2013; Rich et al., 2008) and voices (Deveney et al., 2012). The same deficits are also seen in people with depression (Dalili et al., 2015; Kan et al., 2004; Naranjo et al., 2011).

To date, there was some evidence to suggest that deficits in emotion processing might predict the development of depressive symptomatology (Beevers & Carver, 2003; Vrijen et al., 2016). In Chapter 4, we provide further evidence that suggests that this is also the case in youth with chronic severe irritability. Yet, a larger sample will be needed to replicate our findings.

In addition, Chapter 5 shows how add-on citalopram, a serotonin reuptake inhibitor (SRI), significantly reduced irritability in youth with SMD after an 8-week RCT. There is some evidence to suggest that SRIs could normalise deficits in emotion processing in depression (Harmer et al., 2003; Harmer et al., 2009; Pringle & Harmer, 2015). It is thus possible that these effects also take place in youth with irritability, and consequently decrease the likelihood of developing depression later in life, by both reducing irritability and improving emotion processing skills.

6.4.2.3 Aberrant response to reward

Another potential mechanism that might explain the overlap between irritability and depression is reward processing, which is altered in both youth with depression and youth with irritability (Blair, 2012; Forbes & Dahl, 2012).
As described in section 1.6.2.1, the other main pathophysiological mechanism of irritability is an aberrant response to frustrative non-reward (Brotman et al., 2017), which is defined as the reaction to blocked goal attainment. (Berkowitz, 1989). That is, from the RDoC framework (Insel et al., 2010), irritability and anger are conceptualised as a response to frustration, which occurs when an individual continues to do an action in the expectation of a reward but does not actually receive that reward (Berkowitz, 1989). This happens because the individual has difficulties in realising that the reinforcement contingencies have changed (i.e., the action no longer engenders reward). Overall, evidence from neuroimaging studies suggest that fronto-limbic regions are hypoactivated when rewards are omitted (Deveney et al., 2013; Perlman et al., 2015) and hyperactivated when rewards are received (Perlman et al., 2015).

There is a good theoretical background (Blair, 2012; Rolls, 2007) for investigating irritability within a reward framework. However, reward processes have been barely examined in youth with irritability as compared with those with depression (Admon & Pizzagalli, 2015; Forbes & Dahl, 2012). There is a large body of literature supporting the notion that alterations in reward processing are a central mechanism in the development and maintenance of depression.

Altered neural responses to reward in the frontostriatal reward network are evident in adolescents with depressive disorder (Forbes & Dahl, 2012; Kerestes, Davey, Stephanou, Whittle, & Harrison, 2014) as well as in unaffected first degree relatives of patients with depression (Olino et al., 2014). The most replicated findings has been that depression is associated with reduced activation in striatal regions during both anticipation of reward.
(Forbes et al., 2006; Forbes et al., 2009; Olino et al., 2011; Pizzagalli et al., 2009; Smoski et al., 2009; Zhang, Chang, Guo, Zhang, & Wang, 2013) and positive feedback (i.e., receipt of reward) (Forbes et al., 2006; Forbes et al., 2009; Smoski et al., 2009; Zhang et al., 2013). These findings were summarised in a recent meta-analysis of fMRI and EEG studies. Specifically, we meta-analysed more than 50 studies and concluded that reduced striatal activation was evident in depressed participants during reward feedback and reward anticipation, and the latter was particularly strong in young people (Keren et al., 2018). Moreover, we have shown before that ventral striatum (VS) activity during reward anticipation was specifically related to anhedonia, as well as reduced in subthreshold-and clinical-depression compared to healthy adolescents in a dose-response fashion (Stringaris et al., 2015). In addition, reduced VS activity during reward anticipation predicted transition to subthreshold or clinical depression in previously healthy adolescents two years later (Stringaris et al., 2015).

Some studies show that depression is associated with lower activity in response to rewards in inferior orbitofrontal cortex (OFC) (Forbes et al., 2006; Smoski, Rittenberg, & Dichter, 2011), anterior cingulate cortex (ACC) (Knutson, Bhanji, Cooney, Atlas, & Gotlib, 2008; Smoski et al., 2009; Smoski et al., 2011) and middle frontal gyrus (Smoski et al., 2009), whilst others has reported higher activity in these regions, including the ventromedial prefrontal cortex (vmPFC) (Forbes et al., 2006; Forbes et al., 2009; Knutson et al., 2008; Zhang et al., 2013). Regarding responses to negative feedback, youth with depression seem to respond with enhanced activity in striatal regions (Stringaris et al., 2015), though some studies found the opposite effect (Eshel & Roiser, 2010; Forbes et al., 2006).
Paradoxically, though, depression is typically associated with low reward approaches (Stringaris et al., 2015; Treadway & Zald, 2011), whereas irritability is associated with behavioural approach tendencies and the seeking of rewards (Carver & Harmon-Jones, 2009; Kessel et al., 2016). That is, both irritability and depression have an impact on the motivation to obtain rewards, though in opposite directions. Figure 1.2 in the introduction of this thesis showed this particular position of irritability and depression among other emotions, in which both are negatively valenced, but associated with approach and withdrawal behaviours respectively.

Therefore, so far the pathway through which aberrant reward processing might explain the transition between irritability and depression is unclear. Here I outline two plausible and complementary explanations. First, one possibility is due to shared genes given that reward sensitivity is partly heritable (Bogdan & Pizzagalli, 2009; Menne-Lothmann et al., 2012; Yacubian et al., 2007) and shares genetic variance with depression (Bogdan & Pizzagalli, 2009). It remains to be tested whether reward processing shares genetic variance with irritability.

Second, it is possible that youth with irritability “suffer” from reward-hypersensitivity and are over-responsive to both types of reward, positive (i.e., gain) and negative (i.e., loss). Consequently, when rewards are omitted, youth with irritability might be at increased risk for depression given the high value they place on rewards. The reward-hypersensitivity model was originally conceptualised to explain reward processes in bipolar disorder that could account for both manic episodes, usually presented with irritability, and depression episodes (Alloy, Nusslock, & Boland, 2015; Depue & Iacono, 1989; Nusslock & Alloy,
Nevertheless, this model still does not explain why the development of depression is later in time, even though manifestations of irritability and frustration are frequent and persistent every time a rewarding stimulus is omitted. It is possible that depressive symptoms only emerge when the loss of reward is permanent or unavoidable, whereas irritability is the prominent presentation when the loss is perceived as something that could have been avoided. This notion is in line with the conceptualisation of anger and sadness under Roll’s theory of emotions (Rolls, 2007). This theory proposes that distinct reinforcement contingencies shape different emotions. That is, distinct emotions are elicited depending on the reward/punishment presented. However, it also considers other factors, one being the behavioural responses that are available. From this point of view anger would be elicited when an expected reward is omitted (i.e., frustrative non-reward) and an active behavioural response is possible (Rolls, 2007). By contrast, sadness would be elicited if, in the same situation, only a passive behaviour is possible. This theory is depicted in Figure 6.5 and is useful in understanding how depression and irritability can be both approached from a reward framework.
Figure 6.5. Rolls’ theory of emotions. Some of the emotions associated with different reinforcement contingencies are indicated. Intensity increases away from the centre of the diagram, on a continuous scale. The classification scheme created by the different reinforcement contingencies consists of (1) the presentation of a positive reinforcer (S+), (2) the presentation of a negative reinforcer (S−), (3) the omission of a positive reinforcer (S+ !) or the termination of a positive reinforcer (S+ !), and (4) the omission of a negative reinforcer (S−) or the termination of a negative reinforcer (S− !). It should be understood that each different reinforcer will produce different emotional states: this diagram just summarizes the types of emotion that may be elicited by different contingencies, but the actual emotions will be different for each reinforce. Adapted from Rolls (2007).

To sum up, both irritability and depression are associated with aberrant, and distinct approaches to reward. Since no study has compared reward processes between youth with irritability and youth with depression, it is still unclear how reward processing impacts on
the transition between both conditions. It is likely that common genetic variance is involved, but this and other factors need to be tested. In the next section, I discuss future research directions, including approaches to study reward processes between irritability and depression.

### 6.5 Research implications and future research directions

The first study of this thesis (Chapter 2) shows how irritability is specifically associated with the development of depressive and anxiety disorders. Whereas there has been several studies examining associations between chronic irritability and anxiety (Kircanski et al., 2018b; Stoddard et al., 2017; Wiggins et al., 2016), including the development of new treatment approaches (Kircanski et al., 2018a), research focused on pathophysiological pathways from irritability to depression is scarce (Vidal-Ribas et al., 2018a; Whelan et al., 2015; Wiggins et al., 2014). As discussed previously, some studies suggest shared genetic and environmental contributions to the overlap between irritability and depression (Rappaport et al., 2018a; Savage et al., 2015; Stringaris et al., 2012b). However, based on my own studies and some existing data, I hypothesise that deficits in emotional and reward processing might also play a role in this transition. However, several research steps need to be taken to confirm this hypothesis. In this section, I suggest future research approaches to study the overlap between irritability and depression, as well as improvement toward the study of chronic severe irritability and its treatment.

First, quantitative genetic studies need to test the genetic and environmental contributions to the relationship between irritability, depression, and deficits in emotion and reward processing. We know that irritability and depression (Stringaris et al., 2012b), as well as
depression and reward sensitivity (Bogdan & Pizzagalli, 2009), share genetic risk. However, it is unknown whether irritability and reward sensitivity share genetic and/or environmental factors. Similarly, preliminary data suggest genetic correlations between deficits in emotion recognition and irritability (Rappaport et al., 2018b). However, to my knowledge, no data from quantitative genetic studies exist that examines associations between emotion processing (not only emotion recognition deficits, but also attentional biases, for example) and depression. Longitudinal twin designs could help us to estimate the direction of effects between irritability, emotion/reward processing deficits, and depression, and the genetic contributions to these effects. For example, these studies might show that these deficits may be related to future irritability, and consequently, put the children at higher risk of developing depression. This is in line with the model of irritability as a cause of depression (section 6.4.1.2). Alternatively, and in line with my hypothesised model, emotion and reward processing deficits could account for the association between irritability and depression because they are a risk factor for both irritability and depression (see the model of shared risks in section 6.4.1.3 and the hypothesised model in section 6.4.2, and Figure 6.4).

Second, it is unclear whether behavioural and neuroimaging findings in youth with chronic irritability can be purely explained by irritability symptoms, or alternatively, whether some of the variance in associations is also explained by depressive symptoms. I have shown in Chapter 4, for example, that deficit in emotion recognition in youth with DMDD are closely associated with current depressive symptoms. However, given the high correlation between irritability and depressive symptoms (r=0.51, p<0.001, in this study), it is unclear how much of the variance in this association is explained by irritability and how much of it
is explained by depressive symptoms. Previous studies have tried to answer this question for anxiety symptoms by using a latent variable approach (Kircanski et al., 2018b). Specifically, one could fit a bifactor model in which both items of irritability and depressive symptoms would load on a general factor, but simultaneously also load on two separate specific first-order latent factors of irritability and depression. By doing so, we could examine shared and specific explained variances with behavioural or neuroimaging observations.

Third, to date, no studies have directly compared youth with chronic severe irritability (e.g., DMDD) and youth with major depression disorder (MDD). As discussed in section 6.4.2.3, both conditions present aberrant responses to reward. Whereas these responses differ overall, there might be some overlap that could explain why irritability is associated with later depressive disorder. Currently, at the NIMH, we have scanned hundreds of adolescents with MDD and some dozens of youth with DMDD using fMRI during a reward task, i.e., the Monetary Incentive Delay task (MID) (Knutson, Adams, Fong, & Hommer, 2001), which is currently the most used and validated MRI reward task. We will soon have results after comparing these youth, along with those with anxiety, that can provide us with some clues for explaining the specific association between irritability and depression. In addition, we have been also testing our participants with an Effort task. Specifically, in this task participants are presented with a series of repeated trials in which they can choose between performing a “hard-task” or an “easy-task” to earn varying amounts of monetary rewards. Evidence suggests that anger is associated with the seeking of rewards (Carver & Harmon-Jones, 2009; van Honk & Schutter, 2006) whereas depression is associated with a decreased motivation for seeking rewards (Treadway, Bossaller, Shelton, & Zald, 2012).
addition, whereas sadness seems to be associated with greater perceived task difficulty, anger is associated with perceived task ease (Gendolla & Silvestrini, 2011). We still have no results on this task, but based on basic experimental studies, I hypothesise that youth with DMDD will be more likely to choose the hard task and expend more effort than youth with MDD, who will be more likely to choose the easy task to obtain rewards. This highlights a difference in effort between irritability and depression and therefore does not explain why irritability predicts future depression. However, as depicted in Figure 6.6, I hypothesise that as soon as youth with irritability stop obtaining rewards after expending effort (i.e., by manipulating the task), they will, after an initial peak, decrease the expended effort; and this change in the amount of expended effort will be associated with changes in subjective mood. In contrast, if rewards are obtained intermittently, then irritability becomes chronic and expended effort is maintained. This experimental setup, however, is limited to explaining the longitudinal association between irritability and depression, since it would only test changes in current mood within the time of the task and not long-term changes. Therefore, follow-up studies of youth with DMDD are needed to examine what specific forms of aberrant reward processing can differentiate those DMDD youth who develop depression from those who do not.
Figure 6.6. Transition from irritability to depression. This graph shows the course of expended effort over time when expected rewards are no longer obtained (i.e., frustration condition) after a period in which these rewards have been received. Irritability is initially associated with an increased effort, but in the long term, if rewards are still unavailable, the expenditure of effort will decrease, and depressive symptoms will arise. R+ Delivery of reward; R- Omission of reward.

Fourth, one limitation of the studies of this thesis, and the irritability literature overall, is the lack of well-validated measures to assess irritability in youth (see section 1.5.2). Although the Affective Reactivity Index (ARI) (Stringaris et al., 2012a) has been shown to have good psychometric properties and discriminate well between different psychiatric conditions (Meffert et al., 2017; Mulraney, Melvin, & Tonge, 2014; Pan & Yeh, Epub), it does not capture all domains of irritability. It is a short questionnaire by design that was developed to capture irritability in busy clinical settings or large epidemiological studies, in
which participants are usually asked to complete many questionnaires. I am currently involved in the development of a new measure of irritability that is being designed at the Emotion and Development Branch at the NIMH. Specifically, we are trying to develop a comprehensive measure of irritability that captures all the domains of interest. These include phasic components of irritability (i.e., temper outbursts), tonic components of irritability (i.e., mood between outbursts), cognitive components (e.g., rumination), somatic components, information about triggers and contexts in which irritability emerges, as well as parental accommodation behaviours (e.g., parents’ change of routines or “walk on eggshells” to avoid child’s tantrums). I am also currently involved in the validation of the DMDD module of the Development and Wellbeing Assessment (DAWBA). Specifically, I am using a clinical sample of youth with DMDD that have completed both the DAWBA and the K-SADS DMDD module, which is the gold standard to date to ascertain DMDD. Once validated and a proper algorithm has been generated, I will examine the prevalence and correlates of DMDD in N~9,000 children and adolescents from the British Child and Adolescent Mental Health Survey (BCAMHS). Of note, the data from this community sample will be the first data on DMDD collected with a validated instrument specifically designed to ascertain DMDD.

Fifth, whereas several studies have examined current and future correlates of irritability, few studies have examined early predictors of chronic severe irritability (Dougherty et al., 2014; Griffith et al., 2017; Munhoz et al., 2017). In Chapter 3, I have shown how stress reactivity in infants is associated with irritability, defined as ODD dimension, at preschool age, and how this association differs by sex. Early stress is a strong predictor of depression (Monroe, Slavich, & Georgiades, 2014; Pizzagalli, 2014), and it has been hypothesised that
this is partly caused by the impact that stress has on the neural reward circuit, according to the reward mediation hypothesis (Auerbach, Admon, & Pizzagalli, 2014; Ironside et al., 2018; Stanton et al., 2018). In addition, it has also been suggested that aberrant reward processing can generate interpersonal stress in its own right, which in turn would increase the risk of depression (stress generation hypothesis) (Auerbach et al., 2014; Hammen, 1991). I am currently involved in a project where I show that both hypotheses can take place, but the reward regions involved differ in each case (Vidal-Ribas, 2018; Vidal-Ribas et al., 2018b). It would be interesting to test whether early stress also impacts on the development of chronic irritability, perhaps also by altering the reward circuit and whether chronic irritability can be a generator of interpersonal stress as well. If this was the case, that could be another pathway by which irritability increases the risk of depression.

Sixth, the findings of this thesis provide clues for the development of treatment approaches to irritability. Motivated by the relationship between irritability and emotional disorders, in Chapter 5 we showed that adjunctive citalopram, an SRI, to stimulant medication can reduce irritability in children with DMDD. It would be interesting to test whether citalopram alone (or combined with psychosocial treatment, like parent management training) could also reduce irritability in a sample of DMDD without high rates of co-occurring ADHD. Another target of intervention that needs to be addressed is anger rumination (Sukhodolsky, Golub, & Cromwell, 2001). It has been shown that anger rumination plays a role in the maintenance of angry mood, as opposed, for example, distraction or cognitive reappraisal (Denson, Moulds, & Grisham, 2012). Therefore, anger rumination can be treated in the same way as CBT targets negative and irrational thoughts in depression (Beck, Rush, Shaw, & Emery, 1979). In addition, the results from Chapter 4,
in which deficits in emotion recognition increase the risk of depressive symptoms, suggest that targeting and improving emotion processing skills can be beneficial to prevent the development of depression in youth with chronic irritability. For instance, cognitive bias modification (CBM), targeting both interpretation and attentional bias, have been found to be effective in the prevention of depression in high-risk groups (Browning et al., 2012). Recently, an open trial showed reductions in irritability by modifying interpretation biases in youth with DMDD (Stoddard et al., 2016); and an RCT employing the same method is underway at the NIMH (ClinicalTrials.gov identifier NCT02531893). It is possible that this treatment approach might also prevent the development of future depression.

Finally, there has been an increased interest in the association between irritability and suicide, with the recent publication of two systematic reviews on the topic (Benarous et al., 2018; Orri et al., 2018b) and two studies in a large epidemiological sample (Orri et al., In press; Orri et al., 2018a). There was some evidence that irritability was an independent predictor of suicidality (Pickles et al., 2010). These recent studies show that the presence of irritability increases the risk of suicide in those with depressed/anxious mood (Orri et al., 2018a) and that distinct irritability trajectories in childhood are associated with suicidality in adolescent in different ways. Specifically, a rising trajectory was directly associated with increased suicidality. However, the association with suicidality was mediated by depressive symptoms in those with a persistent trajectory (Orri et al., In press). Despite these findings, it is still unclear what are the mechanisms underlying the effects of irritability on suicide. It could be that irritability is a cause of suicide, acting as a proximal risk factor, for example, increasing the likelihood of a suicide attempt in a patient having a temper outburst. However, it is also possible that irritability and suicidality share risks, as in my
hypothesised model between irritability and depression. These risks might be also shared with depression, given that depression itself is one of the strongest factors for suicide (Nock et al., 2008). Lastly, irritability could be simply a marker associated with other risks that might be actually involved in suicide (Stringaris & Vidal-Ribas, 2018). However, research designs different from those employed so far would be needed to establish the predictive value of irritability in suicide.

6.6 Final remarks

Chronic severe irritability is associated with the development of depressive disorders. The most plausible model to account for this association seems to be one where shared risk factors, both genetic and environmental, are considered. Aberrant emotion and reward processing might play a role in this transition. However, more studies employing longitudinal designs in larger samples are needed to properly ascertain the pathways through which these mechanisms impact the developmental association between irritability and depression. Finally, the findings of these studies should be used to inform new treatment approaches for chronic severe irritability in youth.
Appendix A: Statistical Analysis Plan for study in Chapter 5

Statistical Analysis Plan for study in Chapter 5: A double-blind randomised controlled trial of adjunctive citalopram in youth with chronic severe irritability treated with stimulant
Selective serotonin reuptake inhibitor add-on to stimulant medication in youth with severe mood dysregulation

A double-blind randomised controlled trial of adjunctive citalopram in youth with chronic severe irritability treated with stimulant

Statistical Analysis Plan
Version 0.3
Version started: 21/06/2018
ClinicalTrials.gov Identifier: NCT00794040
This SAP has been written based on Protocol 09-M-0034 Amend 12

Trial Statistician: Pablo Vidal-Ribas (from December 12, 2017)

Chief Investigator: Argyris Stringaris

298
Quantitative Analysis Plan

This document details the presentation and analysis strategy for the primary paper reporting results from this trial. It is intended that the results reported in this paper will follow the strategy set out herein; subsequent papers of a more exploratory nature will not be bound by this analysis plan but will be expected to follow the broad principles laid down for the principle paper(s). The principles are not intended to curtail exploratory analysis or to prohibit sensible statistical and reporting practices but they are intended to establish the strategy that will be followed as closely as possible when analysing and reporting the trial. Reference was made to the trial protocol version 12, ICH guidelines on Statistical Principles (ICH E9 (1998)) reference and CONSORT guidelines (Moher et al., 2010).

Investigators

Ellen Leibenluft, MD
Daniel Pine, MD
Kenneth Towbin, MD
Holly Yokum, MSW
Wanda Wheeler, MSW
Gerald Overman, PharmD BCPP
Chana Engel, CRNP-PMH
Mollie Davis, MSW
Cheri McNeil, PsyD

Principal Investigator
Argyris Stringaris, MD PhD

**Lead Associate Investigator**

Kenneth Towbin, MD

**Analysis Statistician**

Pablo Vidal-Ribas (in consultation with Professor Andrew Pickles, King’s College London)
1.1. **Brief description of the trial**

Severe mood dysregulation (SMD) is a common syndrome in children and adolescents, characterized by very severe irritability - including persistent anger and frequent outbursts - as well as distractibility, hyperactivity, and other symptoms of attention deficit hyperactivity disorder (ADHD).

Many children with SMD receive the diagnosis of bipolar disorder (BD) in the community, although they do not have clear manic episodes (with symptoms such as extreme happiness and decreased need for sleep). However, today we know that SMD and BD differ in family history of psychiatric disorders, pathophysiological mechanisms and longitudinal outcomes. For example, SMD presents high comorbidity with anxiety disorders and is associated with anxiety and depression in longitudinal studies, but not with BD. Furthermore, we know that lithium, a medication typically used to treat BD, does not improve irritability in youth with SMD. To date, there are no randomised trials that have examined pharmacological treatments for SMD with irritability as a primary outcome.

ADHD is primarily treated with stimulant medication, and the first pharmacological option for anxiety and depression are selective serotonin reuptake inhibitors (SSRI). Given the presence of ADHD symptoms in SMD and the specific association with anxiety and depression, this study will evaluate the effectiveness of the stimulant medication methylphenidate when combined (or not combined) with the SSRI citalopram, in treating symptoms of SMD in children and adolescents.
1.1.1. Principal research objectives to be addressed

The hypothesis is that we will demonstrate that citalopram added to methylphenidate improves irritability symptoms to a greater extent than placebo added to methylphenidate.

**Primary objectives**

To conduct a Phase II RCT to test the efficacy of citalopram plus methylphenidate vs. placebo plus methylphenidate in decreasing irritability in youth with severe mood dysregulation.

**Secondary objectives**

To assess the effects of citalopram + methylphenidate on several secondary measures, such as depressive symptoms, anxiety symptoms, side effects, and overall functional impairment.

1.1.2. Trial design

The trial is designed as a double-blind two-arm parallel groups randomised control trial. Children and adolescents aged 7-17 years old will be randomised to either a citalopram plus methylphenidate arm or a placebo plus methylphenidate arm.

Specifically, the trial is comprised of four mandatory phases and one optional phase. During Phase I, participants will withdraw from their current medication (duration flexible, depending on the patient’s medication at admission). Phase II consists of a medication-free period (one week). Phase III involves open treatment with methylphenidate to find the
optimal dose (up to 5 weeks). Phase IV is the treatment phase, in which the randomised trial of citalopram plus methylphenidate vs. placebo plus methylphenidate is undertaken (8 weeks). After Phase IV, blindness will be broken. Finally, an optional Phase V will include open treatment as indicated in preparation for return to community care (Figures 1 and 2). Medication withdrawal, the medication-free week, the methylphenidate open trial, and initial dose stabilization using citalopram/placebo will occur while patients are either hospitalized or attending the Day Treatment Center. During the 8-week citalopram/placebo trial and subsequent open treatment, patients can be hospitalized, in day treatment, or outpatients, according to what is clinically appropriate (Figure 1). Participants will be assessed in the clinic weekly. Baseline measures will be collected just after admission into the protocol, prior Phase I (Baseline 1), before starting the open trial with methylphenidate at Phase III (Baseline 2) and just after randomisation in Phase IV (Baseline 3). Outcomes will be based on the change in reference to Baseline 1 and Baseline 3.
Figure 1. Trial design diagram
Figure 2. Trial design flow diagram
1.1.3. **Method of allocation of groups**

The study was randomised by the Pharmaceutical Development Service of the National Institutes of Health Clinical Center Pharmacy Department using a random numbers table. It was randomised in alternating blocks of six and four in a 1:1 ratio.

Participants and their parents will be blind to treatment allocation. The patient’s research physician, primary NIMH clinician who performs mood ratings, and nursing staff will also be blind to treatment allocation.

1.1.4. **Duration of the treatment period**

The duration of the study (exclusive of open treatment at the end) is approximately 12-15 weeks, depending on the time required for medication discontinuation. As mentioned in section 1.1.2 of this document, the active intervention (Phase IV) consists of 8 weeks.

1.1.5. **Frequency and duration of follow-up**

Participants will be assessed in the clinic once a week beginning at the medication withdrawal period (Phase I) until the end of the randomised trial period (Phase IV). There are no follow-up assessments.

1.1.6. **Visit windows**

The first assessment will take place before the medication withdrawn period (Baseline 1). The final assessment will take place after the randomised trial (Week 8 of Phase IV). Assessments between the first and final collection of measures will be done once a week.
Typically, there was a 2-day visit window (for the inpatients, most were done the same day every week) but it might shift slightly for outpatients—except for the end of week 8 which was exact.

1.1.7. Eligibility screening

Potential SMD participants will be screened via phone. Those who seem to meet inclusion criteria will be seen and screened in person in the clinic. If criteria are met, after this point, participants must give written consent to participate in the trial. After withdrawn medication period, free medication period, and open trial with methylphenidate, participants are rescreened for inclusion criteria prior to randomisation. Inclusion and exclusion criteria are described in section III.A and III.B of the protocol, respectively.

1.1.8. Measures

The mechanisms for which all of the following measures will be recorded are described in detail in the protocol in section IV.

Baseline

The following demographics will be measured at baseline for the child participant:

• Ethnicity (White; Asian; Black or African American; Mixed; Unknown)

• Age (years)

• Sex
• Religion (Anglicanism; Baptist; Catholicism; Christian; Greek Orthodox; Judaism; Protestantism; Non-Denominational; None; Unknown)

The following measure will be collected during the first on-site screening to evaluate inclusion/exclusion criteria:

• Child Schedule for Affective Disorders Present and Lifetime Version (K-SADS-PL) with an additional supplement for SMD (Kaufman et al., 1997).

• Autism Screening Questionnaire (Berument, Rutter, Lord, Pickles, & Bailey, 1999), the Social Responsiveness Scale (SRS) (Constantino et al., 2003), and the Children’s Communication Checklist (Bishop, 1998), so that children with probable PDD can be identified and, if symptoms are severe, excluded from the protocol.

• Full-scale intelligence quotient (FSIQ) was measured by the Wechsler Abbreviated Scale of Intelligence (WASI) (Wechsler, 1999).

**Outcome measures**

All outcomes measures will be first collected before the medication withdrawn period (Baseline 1) and then every other week during this period, and the free medication period. The same measures will be collected before the open treatment with methylphenidate (Baseline 2) and then every other week until before randomisation for the trial. Then, the same measures will be collected just after randomisation (Baseline 3) and then every other week (8 weeks total) until the end of the trial.
**Primary outcome measures**

Our primary outcome measure will be the Clinical Global Impressions (CGI) for Irritability (Spearing et al., 1997). Specifically, we will measure the current severity of irritability (CGI-S) and improvement or change in relation to baseline (CGI-I), both baseline 1 and baseline 3.

The CGI is rated on a 7-point scale, with the severity of illness scale (CGI-S) using a range of responses from 1 (normal) through to 7 (amongst the most severely ill patients). CGI-I scores range from 1 (very much improved) through to 7 (very much worse).

The trial response will be defined as a CGI-I score of less than 3 at the trial’s end (i.e., children who received scores of 2 [much improved], or 1 [symptom-free]).

**Secondary outcome measures**

Several secondary outcomes will be used to assess secondary objectives, as mentioned in section 1.1.1. Namely, to assess the side effects and adverse events of prescribed medications, as well as depressive symptoms, anxiety symptoms and global functional impairment:

- Side Effects will be measured with Dosage Record & Treatment Emergent Symptom Scale (DOTES) (Garvey, Gross, & Freeman, 1991).
• Depressive symptoms will be measured with the Children's Depression Rating Scale (CDRS) (Poznanski & Mokros, 1996).

• Anxiety symptoms will be measured with the Pediatric Anxiety Rating Scale (PARS) (Riddle et al., 2002).

• Global functional impairment will be measured with the Children's Global Assessment Scale (CGAS) (Shaffer et al., 1983).

1.1.9. Sample size estimation (including clinical significance)

The calculation for the protocol assumed a response rate of 60% to citalopram+MPH and 20% to placebo+MPH. While, on the surface, this might seem like a somewhat large clinical difference, when SSRIs are highly effective in paediatric conditions, such as paediatric anxiety disorders, effect sizes of this magnitude are typically found (Ipser, Stein, Hawkridge, & Hoppe, 2009).

Since there are no treatment trials in SMD targeting irritability, the placebo response rate is based on the MTA, a large treatment trial in ADHD using stimulants (Galanter et al., 2003; Greenhill et al., 2001). Assuming power=0.8 and a two-tailed alpha= 0.05, a sample size of 56 (28 in each group) is needed. However, of course, in the event that the response to SSRIs in SMD is more similar to that in paediatric MDD than paediatric anxiety, power will be less than 0.8 (Cipriani et al., 2016; Wagner et al., 2004).

Our experience with a previous treatment trial in this population suggests that roughly 50% of participants might not reach the point of randomisation because of behavioural improvement, withdrawal from the study because of homesickness, or intolerance of
medication discontinuation. Thus, we are anticipating enrolling 112 children in order to ascertain 80 who will be randomised.

1.1.10. Brief description of proposed analyses

Analyses will be carried out by the trial statistician (PVR), following this SAP, and blind to treatment arm, once the database has been locked. Data will be analysed with an intention-to-treat approach (i.e. analyse all those with data in groups as randomised irrespective of treatment received).

There will be descriptive statistics reported on the measures mentioned in 1.1.8, with an aim of comparing the treatment arms, and to review the patient demographics.

For the primary analysis, to test the efficacy of citalopram plus methylphenidate vs. placebo plus methylphenidate in decreasing irritability in youth with SMD, a multilevel mixed-effects growth curve regression will be fitted for all continuous measures.

Further details of the analyses are given later on in this document.

Data summaries and analyses will be carried out in Stata 14.0.

1.2. Data Analysis Plan - Data Description

1.2.1. Recruitment, eligibility and representativeness of patients

A CONSORT flow chart will be constructed – see Figure 3. The number of patients will be summarized using the following categories: the total number of patients contacted; eligible; consenting; and randomised.
Then by treatment arm: patients compliant and non-compliant with intervention; continuing through the trial; withdrawing; excluded or analysed.
Figure 3. CONSORT Diagram
1.2.2. Baseline comparability of randomised groups

Table 1 will present, for all available cases, means and standard deviations (proportions and frequencies for categorical variables) disaggregated by treatment group for baseline values of variables contributing to the primary and secondary outcomes and background participant socio-demographic variables. No statistical significance tests or confidence interval will be calculated for the difference between randomised groups on any participant level baseline variables. The randomisation of intervention groups to participants should have ensured that any imbalance overall of measured and unmeasured baseline characteristics is due to chance (Altman & Dore, 1991).

1.2.3. Adherence to allocated treatment

Any departures from the intended treatment assignment will be described by the number of weeks completed. Adherence to allocated treatment will be defined as being adhered minimum 6 weeks out of 8 weeks (i.e., minimum 75%) (minimum 6).

1.2.4. Loss of cases and other missing data

Table 1 will also present the numbers with endpoint data within each randomised group for each outcome. The major known reasons for loss to follow-up will be described and any systematic differences by treatment group in the characteristics of those lost will be described.
1.2.5. Assessing the quality of outcome measures

Treatment blind analysis will be carried out to assess the quality of data collected, checking that measures conform to appropriate ranges, that scatter plots show no implausible patterns, and dates conform to expectations.

1.2.6. Adverse event reporting

Adverse side effects will be described and analysed as a secondary outcome in the primary paper, as described in section 1.1.8. Specifically, medication side effects will be measured with the Dosage Record & Treatment Emergent Symptom Scale (DOTES). Frequencies of the most common adverse events will be tabled and reported if present in more than one subject in either of the study groups.

1.2.7. Descriptive statistics for outcomes measures

The primary and secondary outcomes as listed in section 1.1.8 will be described by treatment group and time point. Means and standard deviations or medians and interquartile ranges will be used for continuous variables.

1.3. Data Analysis Plan – Inferential analysis

1.3.1. Aims of formal inferences (overview)

The study analysis and publication plan will follow CONSORT guidelines. This statistical analysis plan will be agreed before any inspection and analysis of post-randomisation assessments.
We powered our study to ask whether citalopram added to methylphenidate improves irritability symptoms and functional impairment over placebo added to methylphenidate. The formal statistical analyses will estimate the differences in relevant variables between patients randomised to citalopram added to methylphenidate compared to patients randomised to placebo added to methylphenidate, by intention to treat. That is, all randomised patients will be included in the analyses.

Group difference estimates and associated 95% confidence intervals will be reported. All data preparation and analysis for the primary paper will be blind to treatment group. If any of the data contain information that may disclose blindness, these data will be re-coded before analysis. The overall significance level will be 5% (two-sided) for the primary and secondary outcomes.

Details on the methods for handling missing data are given in sections 1.3.3.

Sensitivity analyses will be used to assess the robustness of conclusions; please refer to section 1.3.4 for details of the planned sensitivity analyses.

No interim analysis is planned.

1.3.2. Analysis of the primary outcome

The analysis population will include all randomised patients. The primary outcome is the improvement of irritability as measured with CGI, see section 1.1.8.

For our primary categorical outcome (i.e. treatment response as defined by a CGI-I score < 3), an efficient estimate was obtained by fitting a growth curve model for the repeated binary response fitted by maximum likelihood in the Stata program gsem, with binomial family and logit link
function. The model included a random intercept. The fixed part of the model included treatment
group, number of weeks as a measure of time, and the group by week interaction (a test for a
quadratic term was also carried out). The post-estimation `lincom` was used to estimate the group
difference at the 8th week and its 95% confidence interval (CI). To assist in interpretation, this
conditional or subject-specific effect estimate (and its 95% CI) was translated to an approximate,
but more easily understood, marginal mean estimate (Szmaragd et al., 2013). The estimate from the
group by week interaction in the model provided a measure of the difference in response rate
between groups across the 8 weeks of the trial. Estimated proportions of response were plotted with
estimates provided by command `margins`. We also performed this analysis for each of the SMD
symptom domains (i.e., temper outbursts, mood between outbursts and hyperarousal symptoms).

For the analysis of continuous outcomes (i.e., severity as measured by CGI-S, functionality as
measured by CGAS, anxiety symptoms as measured by PARS, and depressive symptoms as
measured by CDRS) a growth curve model was fitted using the `xtmixed` routine of the Stata
statistical program, with restricted maximum likelihood (REML) estimation and an exchangeable
covariance matrix for the covariances of the errors of the repeated measures. The fixed part of the
model included the baseline values of the outcome variables, treatment group, number of weeks as a
measure of time, and the interaction of group by week. The post-estimation `lincom` was used to
estimate the group difference at 8th week and its 95% CI. The estimate from the group by week
interaction in the model provided a measure of the difference in severity between groups across the
8 weeks of the trial. Distributional assumptions of our primary outcomes were checked by the use
of Q-Q plots of residuals.

Frequencies of the most common adverse effects are reported if present in more than one subject in
either study group. These were compared using 2-sided Fisher’s exact test.

1.3.3. Missing Data
Account for missing measures will be made under an assumption of a missing-at-random mechanism. Since all remaining variables in the analysis are acquired prior to randomisation there should be no further missing data.

1.3.4. Sensitivity analysis

Sensitivity analysis will be undertaken by replicating the main analyses using the Last Observation Carried Forward method for the treatment of missing data.
References


APA. (2013). *Diagnostic and Statistical Manual of Mental Disorders (5th ed.)* (5th ed.). Washington, DC.


North Tonawanda, NY.

Psychopharmacology and aggression. I: A meta-analysis of stimulant effects on
overt/covert aggression–related behaviors in ADHD. *J Am Acad Child Adolesc
Psychiatry, 41*(3), 253-261.

Constantino, J. N., Davis, S. A., Todd, R. D., Schindler, M. K., Gross, M. M., Brophy, S.
L., . . . Reich, W. (2003). Validation of a brief quantitative measure of autistic traits:
comparison of the social responsiveness scale with the autism diagnostic interview-

and correlates of DSM-5 proposed disruptive mood dysregulation disorder. *Am J

Developmental Findings From the Great Smoky Mountains Study. *J Am Acad Child
Adolesc Psychiatry, 54*(8), 635-642. doi: 10.1016/j.jaac.2015.05.008

diagnostic and functional outcomes of DSM-5 disruptive mood dysregulation

Dysregulation Disorder in Children With Autism Spectrum Disorder. *Journal of the
American Academy of Child & Adolescent Psychiatry, 55*(10, Supplement), S269-
S270. doi: https://doi.org/10.1016/j.jaac.2016.07.164


risk and control children. *Biol Psychol, 82*(2), 156-163. doi:
10.1016/j.biopsycho.2009.07.002


the National Comorbidity Survey Replication. Arch Gen Psychiatry, 62(6), 593-602. doi: 10.1001/archpsyc.62.6.593


McGough, J. (2012). [Characterization and Sequential Pharmacotherapy of Severe Mood Dysregulation (ClinicalTrials.gov Identifier: NCT01714310)].


Noyes Jr (Eds.), *Handbook of Anxiety. Volume 3. The Neurobiology of Anxiety* (pp. 3-58). Amsterdam, The Netherlands: Elsevier Science BV.


369


Stringaris, A., Maughan, B., Copeland, W. S., Costello, E. J., & Angold, A. (2013). Irritable mood as a symptom of depression in youth: prevalence, developmental,


