

https://dspace.org/

A videofeedback parenting intervention to prevent enduring behavior problems in atrisk children aged 1236 months the Healthy Start Happy Start RCT

C O'Farrelly; B Barker; H Watt; D Babalis; E et al ...

https://hdl.handle.net/10539/35357
WOS

Downloaded from a DSpace repository. https://dspace.org/

NIHR | National Institute for Health Research

Journals Library

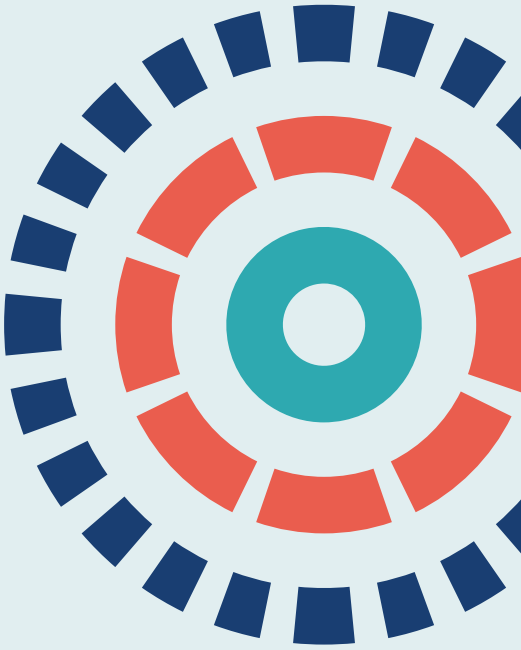
Health Technology Assessment

Volume 25 • Issue 29 • May 2021

ISSN 1366-5278

A video-feedback parenting intervention to prevent enduring behaviour problems in at-risk children aged 12–36 months: the Healthy Start, Happy Start RCT

Christine O'Farrelly, Beth Barker, Hilary Watt, Daphne Babalis, Marian Bakermans-Kranenburg, Sarah Byford, Poushali Ganguli, Ellen Grimås, Jane Iles, Holly Mattock, Julia McGinley, Charlotte Phillips, Rachael Ryan, Stephen Scott, Jessica Smith, Alan Stein, Eloise Stevens, Marinus van IJzendoorn, Jane Warwick and Paul Ramchandani



DOI 10.3310/hta25290

A video-feedback parenting intervention to prevent enduring behaviour problems in at-risk children aged 12–36 months: the Healthy Start, Happy Start RCT

Christine O'Farrelly,[1,2] Beth Barker,[1,2] Hilary Watt,[3] Daphne Babalis,[4] Marian Bakermans-Kranenburg,[5] Sarah Byford,[6] Poushali Ganguli,[6] Ellen Grimås,[1] Jane Iles,[1,7] Holly Mattock,[1] Julia McGinley,[8] Charlotte Phillips,[1] Rachael Ryan,[1] Stephen Scott,[6] Jessica Smith,[1,4] Alan Stein,[9] Eloise Stevens,[1,2] Marinus van IJzendoorn,[10] Jane Warwick[11] and Paul Ramchandani[1,2*]

[1]Division of Psychiatry, Imperial College London, London, UK
[2]Centre for Research on Play in Education, Development, and Learning, Faculty of Education, University of Cambridge, Cambridge, UK
[3]School of Public Health, Imperial College London, London, UK
[4]Imperial Clinical Trials Unit, Imperial College London, London, UK
[5]Clinical Child and Family Studies, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands
[6]Institute of Psychology, Psychiatry, and Neuroscience, King's College London, London, UK
[7]School of Psychology, University of Surrey, Guildford, UK
[8]Netmums, London, UK
[9]Department of Psychiatry, University of Oxford, Oxford, UK
[10]Department of Psychology, Education, and Child Studies, Erasmus University Rotterdam, Rotterdam, the Netherlands
[11]Warwick Clinical Trials Unit, University of Warwick, Coventry, UK

*Corresponding author

Declared competing interests of authors: Marinus van IJzendoorn and Marian Bakermans-Kranenburg declare themselves as two of the developers of the VIPP-SD intervention. Paul Ramchandani has received funding in the form of a donation for research from the LEGO Foundation (Billund, Denmark). He was a member of the Psychological and Community Therapies Panel from 2009 to 2014.

Published May 2021
DOI: 10.3310/hta25290

This report should be referenced as follows:

O'Farrelly C, Barker B, Watt H, Babalis D, Bakermans-Kranenburg M, Byford S, *et al.* A video-feedback parenting intervention to prevent enduring behaviour problems in at-risk children aged 12–36 months: the Healthy Start, Happy Start RCT. *Health Technol Assess* 2021;**25**(29).

Health Technology Assessment is indexed and abstracted in *Index Medicus*/MEDLINE, *Excerpta Medica*/EMBASE, *Science Citation Index Expanded* (SciSearch®) and *Current Contents®*/ Clinical Medicine.

Health Technology Assessment

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 3.370

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, the Cochrane Library and Clarivate Analytics Science Citation Index.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

Health Technology Assessment (HTA) research is undertaken where some evidence already exists to show that a technology can be effective and this needs to be compared to the current standard intervention to see which works best. Research can evaluate any intervention used in the treatment, prevention or diagnosis of disease, provided the study outcomes lead to findings that have the potential to be of direct benefit to NHS patients. Technologies in this context mean any method used to promote health; prevent and treat disease; and improve rehabilitation or long-term care. They are not confined to new drugs and include any intervention used in the treatment, prevention or diagnosis of disease.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 13/04/33. The contractual start date was in October 2014. The draft report began editorial review in January 2020 and was accepted for publication in October 2020. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care.

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

NIHR Journals Library Editor-in-Chief

Professor Ken Stein Professor of Public Health, University of Exeter Medical School, UK

NIHR Journals Library Editors

Professor John Powell Chair of HTA and EME Editorial Board and Editor-in-Chief of HTA and EME journals. Consultant Clinical Adviser, National Institute for Health and Care Excellence (NICE), UK, and Professor of Digital Health Care, Nuffield Department of Primary Care Health Sciences, University of Oxford, UK

Professor Andrée Le May Chair of NIHR Journals Library Editorial Group (HS&DR, PGfAR, PHR journals) and Editor-in-Chief of HS&DR, PGfAR, PHR journals

Professor Matthias Beck Professor of Management, Cork University Business School, Department of Management and Marketing, University College Cork, Ireland

Dr Tessa Crilly Director, Crystal Blue Consulting Ltd, UK

Dr Eugenia Cronin Senior Scientific Advisor, Wessex Institute, UK

Dr Peter Davidson Consultant Advisor, Wessex Institute, University of Southampton, UK

Ms Tara Lamont Senior Scientific Adviser (Evidence Use), Wessex Institute, University of Southampton, UK

Dr Catriona McDaid Senior Research Fellow, York Trials Unit, Department of Health Sciences, University of York, UK

Professor William McGuire Professor of Child Health, Hull York Medical School, University of York, UK

Professor Geoffrey Meads Emeritus Professor of Wellbeing Research, University of Winchester, UK

Professor James Raftery Professor of Health Technology Assessment, Wessex Institute, Faculty of Medicine, University of Southampton, UK

Dr Rob Riemsma Reviews Manager, Kleijnen Systematic Reviews Ltd, UK

Professor Helen Roberts Professor of Child Health Research, UCL Great Ormond Street Institute of Child Health, UK

Professor Jonathan Ross Professor of Sexual Health and HIV, University Hospital Birmingham, UK

Professor Helen Snooks Professor of Health Services Research, Institute of Life Science, College of Medicine, Swansea University, UK

Professor Ken Stein Professor of Public Health, University of Exeter Medical School, UK

Professor Jim Thornton Professor of Obstetrics and Gynaecology, Faculty of Medicine and Health Sciences, University of Nottingham, UK

Please visit the website for a list of editors: www.journalslibrary.nihr.ac.uk/about/editors

Editorial contact: journals.library@nihr.ac.uk

Abstract

A video-feedback parenting intervention to prevent enduring behaviour problems in at-risk children aged 12–36 months: the Healthy Start, Happy Start RCT

Christine O'Farrelly[1,2] Beth Barker[1,2] Hilary Watt[3] Daphne Babalis[4] Marian Bakermans-Kranenburg[5] Sarah Byford[6] Poushali Ganguli[6] Ellen Grimås[1] Jane Iles[1,7] Holly Mattock[1] Julia McGinley[8] Charlotte Phillips[1] Rachael Ryan[1] Stephen Scott[6] Jessica Smith[1,4] Alan Stein[9] Eloise Stevens[1,2] Marinus van IJzendoorn[10] Jane Warwick[11] and Paul Ramchandani[1,2*]

[1]Division of Psychiatry, Imperial College London, London, UK
[2]Centre for Research on Play in Education, Development, and Learning, Faculty of Education, University of Cambridge, Cambridge, UK
[3]School of Public Health, Imperial College London, London, UK
[4]Imperial Clinical Trials Unit, Imperial College London, London, UK
[5]Clinical Child and Family Studies, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands
[6]Institute of Psychology, Psychiatry, and Neuroscience, King's College London, London, UK
[7]School of Psychology, University of Surrey, Guildford, UK
[8]Netmums, London, UK
[9]Department of Psychiatry, University of Oxford, Oxford, UK
[10]Department of Psychology, Education, and Child Studies, Erasmus University Rotterdam, Rotterdam, the Netherlands
[11]Warwick Clinical Trials Unit, University of Warwick, Coventry, UK

*Corresponding author pr441@cam.ac.uk

Background: Behaviour problems emerge early in childhood and place children at risk for later psychopathology.

Objectives: To evaluate the clinical effectiveness and cost-effectiveness of a parenting intervention to prevent enduring behaviour problems in young children.

Design: A pragmatic, assessor-blinded, multisite, two-arm, parallel-group randomised controlled trial.

Setting: Health visiting services in six NHS trusts in England.

Participants: A total of 300 at-risk children aged 12–36 months and their parents/caregivers.

Interventions: Families were allocated in a 1 : 1 ratio to six sessions of Video-feedback Intervention to promote Positive Parenting and Sensitive Discipline (VIPP-SD) plus usual care or usual care alone.

Main outcome measures: The primary outcome was the Preschool Parental Account of Children's Symptoms, which is a structured interview of behaviour symptoms. Secondary outcomes included caregiver-reported total problems on the Child Behaviour Checklist and the Strengths and Difficulties Questionnaire. The intervention effect was estimated using linear regression. Health and social care

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

service use was recorded using the Child and Adolescent Service Use Schedule and cost-effectiveness was explored using the Preschool Parental Account of Children's Symptoms.

Results: In total, 300 families were randomised: 151 to VIPP-SD plus usual care and 149 to usual care alone. Follow-up data were available for 286 (VIPP-SD, $n = 140$; usual care, $n = 146$) participants and 282 (VIPP-SD, $n = 140$; usual care, $n = 142$) participants at 5 and 24 months, respectively. At the post-treatment (primary outcome) follow-up, a group difference of 2.03 on Preschool Parental Account of Children's Symptoms (95% confidence interval 0.06 to 4.01; $p = 0.04$) indicated a positive treatment effect on behaviour problems (Cohen's $d = 0.20$, 95% confidence interval 0.01 to 0.40). The effect was strongest for children's conduct [1.61, 95% confidence interval 0.44 to 2.78; $p = 0.007$ ($d = 0.30$, 95% confidence interval 0.08 to 0.51)] versus attention deficit hyperactivity disorder symptoms [0.29, 95% confidence interval −1.06 to 1.65; $p = 0.67$ ($d = 0.05$, 95% confidence interval −0.17 to 0.27)]. The Child Behaviour Checklist [3.24, 95% confidence interval −0.06 to 6.54; $p = 0.05$ ($d = 0.15$, 95% confidence interval 0.00 to 0.31)] and the Strengths and Difficulties Questionnaire [0.93, 95% confidence interval −0.03 to 1.9; $p = 0.06$ ($d = 0.18$, 95% confidence interval −0.01 to 0.36)] demonstrated similar positive treatment effects to those found for the Preschool Parental Account of Children's Symptoms. At 24 months, the group difference on the Preschool Parental Account of Children's Symptoms was 1.73 [95% confidence interval −0.24 to 3.71; $p = 0.08$ ($d = 0.17$, 95% confidence interval −0.02 to 0.37)]; the effect remained strongest for conduct [1.07, 95% confidence interval −0.06 to 2.20; $p = 0.06$ ($d = 0.20$, 95% confidence interval −0.01 to 0.42)] versus attention deficit hyperactivity disorder symptoms [0.62, 95% confidence interval −0.60 to 1.84; $p = 0.32$ ($d = 0.10$, 95% confidence interval −0.10 to 0.30)], with little evidence of an effect on the Child Behaviour Checklist and the Strengths and Difficulties Questionnaire. The primary economic analysis showed better outcomes in the VIPP-SD group at 24 months, but also higher costs than the usual-care group (adjusted mean difference £1450, 95% confidence interval £619 to £2281). No treatment- or trial-related adverse events were reported. The probability of VIPP-SD being cost-effective compared with usual care at the 24-month follow-up increased as willingness to pay for improvements on the Preschool Parental Account of Children's Symptoms increased, with VIPP-SD having the higher probability of being cost-effective at willingness-to-pay values above £800 per 1-point improvement on the Preschool Parental Account of Children's Symptoms.

Limitations: The proportion of participants with graduate-level qualifications was higher than among the general public.

Conclusions: VIPP-SD is effective in reducing behaviour problems in young children when delivered by health visiting teams. Most of the effect of VIPP-SD appears to be retained over 24 months. However, we can be less certain about its value for money.

Trial registration: Current Controlled Trials ISRCTN58327365.

Funding: This project was funded by the National Institute for Health Research (NIHR) Health Technology Assessment programme and will be published in full in *Health Technology Assessment*; Vol. 25, No. 29. See the NIHR Journals Library website for further project information.

Contents

List of tables xiii

List of figures xv

List of abbreviations xvii

Plain English summary xix

Scientific summary xxi

Chapter 1 Introduction 1
Behaviour problems and their impact on health and development 1
Interventions for behaviour problems 1
Rationale for the research 2
Aims and objectives 2

Chapter 2 Methods 3
Design 3
Ethics and governance 3
Participants 3
Inclusion criteria 3
Exclusion criteria 3
Setting 4
Recruitment procedure 4
Informed consent 4
Baseline assessment 5
Follow-up 5
Intervention 6
Fidelity 6
Usual care 8
Outcome measures 8
Primary outcome 8
Key secondary outcomes 8
Other secondary outcomes 9
Sample size 10
Randomisation and blinding 10
Data collection and data management 11
Quality assurance and monitoring 11
Data analysis 12
Trial organisation and management 13
Trial management 13
Patient and public involvement group 14
Amendments to protocol 15

Chapter 3 Clinical results 17
Participant flow and retention 17
Baseline participant characteristics 17

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

Delivery of intervention 19
Therapists 19
Intervention dosage 20
Intervention fidelity 20
Primary outcome: Preschool Parental Account of Children's Symptoms score at the 5- and 24-month follow-ups 20
Secondary outcomes: parent-reported measures of child behaviour at the 5- and 24-month follow-ups 22
Missing data 29
Blinding 30
Acceptability 30
Serious adverse events 31

Chapter 4 Economic evaluation **33**
Aim 33
Methods 33
Perspective 33
Method of economic evaluation 33
Costs 33
Data analysis 34
Results 36
Data completeness 36
Resource use over the 24-month follow-up 36
Costs 39
Outcomes 39
Sensitivity analyses 41
Cost–consequences analysis at the 24-month follow-up 43
Long-term cost-effectiveness 45
Discussion 45
Summary of results 45
Strengths and limitations 46

Chapter 5 Discussion **49**
Summary of findings 49
Economic evaluation 50
Strengths and limitations 51
Conclusions 52
Implications for health care 52
Implications for future research 53

Acknowledgements **55**

References **59**

Appendix 1 Primary outcome: Preschool Parental Account of Children's Symptoms **67**

Appendix 2 Additional detail regarding multiple imputation analysis **69**

Appendix 3 Participant recruitment by month from trial commencement to recruitment target ($n = 300$) **71**

Appendix 4 Secondary caregiver analysis **73**

Appendix 5 Supplementary analysis **77**

Appendix 6 Subgroup analysis **81**

Appendix 7 Breakdown of protocol deviations and violations **83**

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

List of tables

TABLE 1 Timing of data collection 5

TABLE 2 Route of recruitment for families randomised into the trial 17

TABLE 3 Participants recruited into the trial disaggregated by recruitment site 18

TABLE 4 Baseline characteristics of participating children and primary caregivers by treatment allocation 18

TABLE 5 Professional backgrounds of therapists implementing VIPP-SD 19

TABLE 6 Primary ITT analysis of adjusted mean difference on the primary outcome of children's behaviour problems at the 5- (primary time point) and 24-month follow-ups (and its subscales) 21

TABLE 7 Primary ITT analysis of adjusted mean difference on secondary outcomes of children's behaviour problems (total scores) at the 5- and 24-month follow-ups 23

TABLE 8 Intention-to-treat analysis of adjusted mean difference of change in children's behaviour problems on the secondary outcome subscales (primary caregiver-reported CBCL and SDQ) at the 5- and 24-month follow-ups 24

TABLE 9 Complier-average causal effect analysis on primary and secondary outcomes of child behaviour 26

TABLE 10 Intention-to-treat analysis of secondary outcomes on participating caregivers' parenting practices, mood, anxiety, and couple functioning, by sex of caregiver 27

TABLE 11 Participants' responses to the intervention feedback questionnaire: perceived positive impact of VIPP-SD 30

TABLE 12 Participants' responses to the intervention feedback questionnaire: perceptions of treatment setting 30

TABLE 13 Unit costs and sources used for the economic evaluation 34

TABLE 14 Costing schema for calculation of intervention costs 35

TABLE 15 Availability of resource use data 36

TABLE 16 Uptake of VIPP-SD intervention sessions by participants randomised to the intervention 37

TABLE 17 Use of health and social care services over 24 months' follow-up 37

TABLE 18 Use of medication over 24 months' follow-up 38

TABLE 19 Mean costs (£) per participant over 24 months' follow-up 39

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

TABLE 20 Mean PPACS scores per participant applied to the cost-effectiveness analysis 40

TABLE 21 Mean total cost (£) and outcomes per participant for sensitivity analyses 41

TABLE 22 Cost–consequences analysis at the 24-month follow-up 44

TABLE 23 Mean utility scores and QALYs per participant mapped from the SDQ 45

TABLE 24 Baseline characteristics of secondary caregivers by treatment allocation 73

TABLE 25 Primary ITT analysis of adjusted mean difference on secondary outcomes of children's behaviour problems (total scores) at the 5- and 24-month follow-ups reported by participating secondary caregivers 74

TABLE 26 Complier-average causal effects analysis on secondary outcomes of child behaviour reported by participating secondary caregivers 74

TABLE 27 Secondary outcomes on child behaviour according to the perception of male and female caregivers: ITT analysis 75

TABLE 28 Sensitivity analyses for lack of adjustment for length of follow-up: primary and secondary outcomes on child behaviour – ITT analysis 77

TABLE 29 Baseline characteristics of participating children and primary caregivers with missing PPACS data by treatment allocation 78

TABLE 30 Baseline and 5- and 24-month follow-up values of outcome measures in those with missing PPACS data at follow-up 79

TABLE 31 Occurrence of protocol deviations and violations 83

List of figures

FIGURE 1 The CONSORT flow diagram for the Healthy Start, Happy Start RCT 7

FIGURE 2 Forest plot for sensitivity analysis at the 5-month follow-up 21

FIGURE 3 Forest plot for sensitivity analysis at the 24-month follow-up 22

FIGURE 4 Forest plot for exploratory subgroup analyses at the 5-month follow-up 28

FIGURE 5 Forest plot for exploratory subgroup analyses at the 24-month follow-up 29

FIGURE 6 Bootstrapped mean differences in costs and PPACS scores at the 24-month follow-up 40

FIGURE 7 Cost-effectiveness acceptability curve showing the probability that VIPP-SD is cost-effective compared with usual care at different values of WTP thresholds for a 1-point improvement in PPACS score at the 24-month follow-up 41

FIGURE 8 Bootstrapped mean differences in costs and PPACS scores at the 24-month follow-up: complete cases and excluding influential outliers 42

FIGURE 9 Cost-effectiveness acceptability curve showing the probability that VIPP-SD is cost-effective compared with usual care at different values of WTP thresholds for a 1-point improvement in PPACS score at the 24-month follow-up: complete cases and excluding influential outliers 43

FIGURE 10 Bootstrapped mean differences in costs and PPACS score at the 5-month follow-up 43

FIGURE 11 Cost-effectiveness acceptability curve showing the probability that VIPP-SD is cost-effective compared with usual care at different values of WTP thresholds for a 1-point improvement in PPACS score at the 5-month follow-up 44

FIGURE 12 Participant recruitment by month from trial commencement to recruitment target 71

FIGURE 13 Subgroup analysis for baseline behaviour based on quartiles of externalising scores on the SDQ at the 5-month follow-up 81

FIGURE 14 Subgroup analysis for baseline behaviour based on quartiles of externalising scores on the SDQ at the 24-month follow-up 82

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

List of abbreviations

ADHD	attention deficit hyperactivity disorder
CACE	complier-average causal effect
CA-SUS	Child and Adolescent Service Use Schedule
CBCL	Child Behaviour Checklist
CEAC	cost-effectiveness acceptability curve
CHU-9D	Child Health Utility index 9D
CI	confidence interval
CONSORT	Consolidated Standards of Reporting Trials
CRN	Clinical Research Network
DMEC	Data Monitoring and Ethics Committee
eCRF	electronic case report form
GAD-7	Generalised Anxiety Disorder-7
GP	general practitioner
HTA	Health Technology Assessment
ICER	incremental cost-effectiveness ratios
ICTU	Imperial Clinical Trials Unit
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
ITT	intention to treat
MI	multiple imputation
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NMB	net monetary benefit
PHQ-9	Patient Health Questionnaire-9 items
PIS	participant information sheet
PMG	Project Management Group
PPACS	Preschool Parental Account of Children's Symptoms
PPI	patient and public involvement
PSS	Personal Social Services
QA	quality assurance
QALY	quality-adjusted life-year
RCT	randomised controlled trial
RDAS	Revised Dyadic Adjustment Scale
SAE	serious adverse event
SD	standard deviation
SDQ	Strengths and Difficulties Questionnaire
SOP	standard operating procedure
SWAT	study within a trial
TSC	Trial Steering Committee
VIPP-Co	Video-feedback Intervention to promote Positive Parenting for Co-parents
VIPP-SD	Video-feedback Intervention to promote Positive Parenting and Sensitive Discipline
WTP	willingness to pay

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

Plain English summary

Behaviour problems in young children are common and are linked to mental and physical health problems, and educational and social difficulties. An important factor that influences the development of behaviour problems is the quality of care that children receive from their caregivers. This study aimed to test if a six-session parenting programme [called Video-feedback Intervention to promote Positive Parenting and Sensitive Discipline (VIPP-SD)] reduced behaviour problems in children aged 1 or 2 years who were showing early signs of behaviour problems (e.g. restlessness, impulsivity, tantrums and aggression). VIPP-SD supports caregivers in responding to their child's communication and behaviour.

A total of 300 families participated. All families continued to access usual health-care services (e.g. health visitors and general practitioners), but half of the families were randomly allocated to also receive the VIPP-SD programme. We visited all families when the study started, and at 5 and 24 months to see if the children whose families received VIPP-SD showed fewer behaviour problems. We measured the children's behaviour by completing interviews and questionnaires with their caregivers. We also analysed whether or not VIPP-SD was good value for money compared with existing services. We did this by comparing the cost of all of the standard health and community services that families accessed during their time in the study, taking account of the impact that VIPP-SD had on children's behaviour.

The children in the VIPP-SD group had lower levels of behaviour problems following the programme than children whose parents did not receive the programme. On average, VIPP-SD children scored 2 points lower on the main measure of behaviour; an example difference would be tantrums being rated as mild rather than severe. By the 2-year visit, the VIPP-SD children continued to show lower levels of behaviour problems. It is less clear whether or not VIPP-SD is good value for money, as this depends on how much money policy-makers are willing to invest for reductions in behaviour problems.

Overall, there is strong evidence that the VIPP-SD programme is effective in reducing behaviour problems in the short term. Most of this benefit appears to be maintained for the following 2 years. However, we are less certain about the long-term effect and the VIPP-SD's value for money.

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

Scientific summary

Background

As many as 1 in 10 children experience behaviour problems, and enduring problems can undermine children's health, social and educational outcomes across their life course. As well as the distress that these issues can cause for children and families, there are considerable costs to society. An important factor that influences the development of behaviour problems is the quality and style of early parental care that children experience. Parenting strategies can lead to improvements in child behaviour; systematic reviews have shown that parenting programmes are effective in reducing behaviour problems in preschool- and school-aged children. However, there are very few effective early psychological interventions available. Intervening earlier in childhood before problems become established could increase the impact that parenting programmes have on children and families.

An evidence-based programme that is suitable for use with children aged $\geq$ 12 months is the Video-feedback Intervention to promote Positive Parenting and Sensitive Discipline (VIPP-SD) programme. VIPP-SD has been developed in a systematic way, has been tested in 12 randomised controlled trials and has been shown to be effective in improving parenting practices and child behaviour outcomes. However, the intervention is yet to be tested in a routine health service context in the UK.

The Healthy Start, Happy Start study was a pragmatic, assessor-blinded, multisite, two-arm, parallel-group randomised controlled trial to test the clinical effectiveness and cost-effectiveness of VIPP-SD for parents of young children (aged 12–36 months) who were at risk of developing enduring behaviour problems.

Objectives

The objectives were to:

- undertake a randomised controlled trial to evaluate whether or not a brief parenting intervention (VIPP-SD) leads to lower levels of behaviour problems in young children who are at a high risk of developing these problems, compared with usual care in the NHS
- undertake an economic evaluation to assess the cost-effectiveness of the intervention compared with usual care.

Methods

Design

The study was a pragmatic, assessor-blinded, multisite, two-arm, parallel-group randomised controlled trial.

Setting

The participants were recruited from health visiting and community services in six UK NHS trusts.

Participants

The participants were 300 children aged 12–36 months who demonstrated elevated behaviour difficulties (as measured by a parent-reported screening questionnaire) and their caregiver(s).

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

Target population

Target participants were children aged 12–36 months who scored in the top 20% for behaviour problems on the parent-reported Strengths and Difficulties Questionnaire. We excluded children or parents with a severe sensory impairment, learning disability or language limitation that was sufficient to preclude participation in the trial. We also excluded children with a participating sibling, those whose parents were actively involved in family court proceedings and those participating in a closely related research study and/or receiving an individual video-feedback-based intervention.

Randomisation

We randomly allocated participants in a 1 : 1 ratio to either VIPP-SD or usual care, stratified by recruitment site and the number of participating caregivers (one vs. two).

Intervention

All families continued to access usual care, which comprised mainly general practitioner and health visiting services. Families allocated to the VIPP-SD group were offered six home-based, fortnightly sessions of 1–2 hours' duration. Each visit was composed of two parts. In the first part, therapists recorded videos of parents during everyday interactions with their children. In the second part of the visit, therapists provided structured feedback based on the intervention manual and the contents of the interaction. This feedback aimed to promote parents' sensitivity; and their capacity to identify their child's attachment cues and exploratory behaviour, and to respond to their child appropriately, as well as providing sensitive discipline, which involves a consistent but non-harsh response to challenging behaviour. The manualised intervention was delivered predominantly by health professionals, including health visitors, community nursery nurses and psychologists, following 5 days of training and a supervised practice case. Therapists received ongoing clinical supervision throughout intervention delivery.

Outcome measurements

Baseline information

We collected demographic data at baseline on parents' sex, age, ethnicity, educational attainment, employment status and relationship status. We also collected data on childrens' sex, age and ethnicity. Baseline measures of all outcome data were also collected.

Primary clinical outcome

The primary outcome was severity of behaviour problems (as measured by the Preschool Parental Account of Children's Symptoms interview) at the 5-month follow-up.

Secondary clinical outcomes

Key secondary outcomes included severity of behaviour problems, measured using the Preschool Parental Account of Children's Symptoms interview at the 24-month follow-up, as well as parent-reported child behaviour, measured at the 5- and 24-month follow-ups using the Child Behaviour Checklist and the Strengths and Difficulties Questionnaire. Additional secondary outcomes included parent-reported measures of parenting practices (Parenting Scale), parent mood (Patient Health Questionnaire-9) and anxiety (Generalised Anxiety Disorder-7) and couple functioning (Revised Dyadic Adjustment Scale) at the 5- and 24-month follow-ups.

Economic measures

Health and social care service use was recorded using the Child and Adolescent Service Use Schedule. In the short term ('within-trial'), the cost-effectiveness of the VIPP-SD was explored in terms of the primary outcome measure (Preschool Parental Account of Children's Symptoms) at the 24-month follow-up and a cost–consequences analysis was carried out. The cost–consequences analysis outlined costs alongside the key secondary outcome measures (Child Behaviour Checklist and Strengths and Difficulties Questionnaire) to explore potential economic impacts of the intervention on outcomes more broadly.

Sample size

A sample size of 300 participants was selected to provide between 80% and 90% power to detect standardised effect sizes of 0.36 and 0.42, respectively, at a 5% level of statistical significance and assuming a 20% attrition rate.

Statistical methods and analyses.

Clinical outcome analyses

We conducted the primary analysis using intention to treat for primary and secondary outcomes. We also undertook secondary analysis to estimate the effect of receiving the intervention using complier-average causal effects analysis on the primary outcome and key secondary outcomes. In the primary analysis, we estimated the effects of the VIPP-SD programme by comparing the VIPP-SD and usual-care groups using linear regression, which adjusted for baseline levels of the same outcome, treatment centre, length of follow-up, age of the child and number of participating caregivers. We undertook a sensitivity analysis to assess the impact of missing data and adjusted for the length of follow-up. For the primary outcome (measured using the Preschool Parental Account of Children's Symptoms), we used multiple imputation to fill in missing items in the Preschool Parental Account of Children's Symptoms scales (as some items were unrateable for some children) and we used multiple imputation for the families where follow-up data were not available (5% at the 5-month follow-up, 6% at the 24-month follow-up).

Economic analyses

Cost-effectiveness was assessed at the 24-month follow-up through the calculation of incremental cost-effectiveness ratios (the additional cost of one intervention compared with another divided by the additional effect) and using the net monetary benefit approach. Uncertainty around the mean estimates of cost and outcome was explored using bootstrapping and plotting the bootstrap iterations onto a cost-effectiveness plane for interpretation. Cost-effectiveness acceptability curves were constructed to examine the probability of VIPP-SD being cost-effective compared with usual care for a range of possible values of willingness to pay per unit improvement in outcome. All economic analyses were adjusted in line with the clinical analyses and based on multiply imputed data sets of total costs and outcomes using chained equations and predictive mean matching. Sensitivity analyses explored the impact of missing data, influential outliers and the selected end point, repeating the analysis for the 5-month follow-up.

Results

Between July 2015 and July 2017, we assessed 2248 families for eligibility. In total, 1430 families were ineligible, 518 families did not progress to the trial (declined/could not be contacted) and 300 families were randomised. Of the randomised families, 151 (50%) were randomly allocated to the intervention (VIPP-SD) group and 149 (50%) were randomly allocated to the usual-care group. Participant-level characteristics at baseline were well balanced between groups. Of the 151 families randomised to receive the intervention, 129 (85%) completed at least four VIPP-SD sessions (the compliance cut-off point for treatment adherence). Retention was high, with primary outcome data available for 286 (95%) participants at the 5-month follow-up and 282 (94%) participants at the 24-month follow-up.

On the primary outcome (Preschool Parental Account of Children's Symptoms at 5 months by intention-to-treat analysis), we found that VIPP-SD was superior to usual care [mean 28.80 (standard deviation 9.2) vs. 30.31 (standard deviation 9.9); adjusted mean difference 2.03 (95% confidence interval 0.06 to 4.01); $p = 0.04$], indicating a positive treatment effect (Cohen's $d = 0.20$, 95% confidence interval 0.01 to 0.40). VIPP-SD was found to be superior to usual care on the conduct problems subscale of the primary outcome (difference 1.61, 95% confidence interval 0.44 to 2.78; $p = 0.007$, $d = 0.30$, 95% confidence interval 0.08 to 0.51), but not the hyperactivity subscale (difference 0.29, 95% confidence interval −1.06 to 1.65; $p = 0.67$, $d = 0.05$, 95% confidence interval −0.17 to 0.27). The positive effect of VIPP-SD

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

on the Preschool Parental Account of Children's Symptoms total score at 5-month follow-up was robust to sensitivity analyses. The complier-average causal effects analysis on the primary outcome of child behaviour showed higher estimated treatment effects in those with acceptable treatment adherence, that is those who received at least four core VIPP-SD sessions (Preschool Parental Account of Children's Symptoms difference increased from 2.03 to 2.59, 95% confidence interval 0.24 to 4.94; $p = 0.03$, $d = 0.26$, 95% confidence interval 0.02 to 0.50). At the 24-month follow-up, there was evidence of a sustained intention-to-treat treatment effect favouring the VIPP-SD group (difference 1.73, 95% confidence interval −0.24 to 3.71; $p = 0.08$, $d = 0.17$, 95% confidence interval −0.02 to 0.37). Again, the difference was higher for the conduct subscale (difference 1.07, 95% confidence interval −0.06 to 2.2; $p = 0.06$, $d = 0.20$, 95% confidence interval −0.01 to 0.42) than for the hyperactivity scale (difference 0.62, 95% confidence interval −0.60 to 1.84; $p = 0.32$, $d = 0.10$, 95% confidence interval −0.10 to 0.30). Those who received at least four VIPP-SD sessions continued to show a greater improvement in behaviour (complier-average casual effects Preschool Parental Account of Children's Symptoms difference increased from 1.73 to 1.96, 95% confidence interval −0.30 to 4.23; $p = 0.09$, $d = 0.20$, 95% confidence interval −0.03 to 0.43).

On the main secondary outcomes (total scores of the Child Behaviour Checklist and Strengths and Difficulties Questionnaire), the results indicated a positive direction of effect favouring the VIPP-SD group at the 5-month follow-up, but less evidence of a sustained effect at 24-month follow-up. We found no appreciable evidence of differences between groups on other included secondary outcomes at the 5- or 24-month follow-up.

No treatment- or trial-related adverse events were reported. There were no group differences in the reporting of unrelated adverse events.

Mean total costs were significantly higher in the VIPP-SD group than in the usual-care group at the 24-month follow-up (adjusted mean difference £1450, 95% confidence interval £619 to £2281; $p = 0.001$) and were driven by the cost of the intervention (mean cost £1466 per family). However, VIPP-SD was also associated with Preschool Parental Account of Children's Symptoms scores that favoured the intervention, thus generating a trade-off, with VIPP-SD being more costly but also more effective than usual care. The probability of VIPP-SD being cost-effective compared with usual care increased as willingness to pay for improvements in Preschool Parental Account of Children's Symptoms score increased, with VIPP-SD having higher probability of being cost-effective at willingness-to-pay values of approximately £800 per 1-point improvement in Preschool Parental Account of Children's Symptoms score (equivalent to 0.10 standard deviation) and above. In theory, this would be equivalent to approximately £7920 for one standard deviation improvement. Because the Preschool Parental Account of Children's Symptoms is not associated with a willingness-to-pay threshold to support decision-making, it is not possible to come to any firm conclusions about the relative cost-effectiveness of VIPP-SD in the short term. These results were robust to changes in assumptions in sensitivity analyses (complete case, excluding outliers and analysis at the 5-month follow-up).

Conclusions

We found evidence that a brief, home-based intervention, VIPP-SD, was more effective than usual care in reducing behaviour problems in this group of children aged 1 or 2 years. Evidence of superiority was found for the primary outcome (the interview-based Preschool Parental Account of Children's Symptoms assessment) at the 5-month post-treatment assessment. The findings were strongest on the conduct problems scale of the Preschool Parental Account of Children's Symptoms assessment, rather than the attention deficit hyperactivity disorder/hyperkinesis scale, which is in keeping with the sensitive discipline focus of the VIPP-SD intervention, which targets conduct problems. Our results are consistent with a meta-analysis of the VIPP-SD intervention, which demonstrated similar effect sizes for child behaviour problems [Juffer F, Bakermans-Kranenburg MJ, van Ijzerdoorn MH. Video-feedback Intervention to Promote Positive Parenting and Sensitive Discipline (VIPP-SD): Development and

Meta-Analytical Evidence of Its Effectiveness. In Steele H, Steele M, editors. *Handbook of Attachment-based Interventions*. 1st edn. New York, NY: Guildford; 2017. pp. 1–26]. Thus, it is noteworthy that the present study demonstrates that this effect is robust in a routine health service context. Our best estimate is that most of the effect of VIPP-SD is retained over 24 months. However, we are less certain about its value for money.

Implications for health care and future research

The results of this research show that the VIPP-SD intervention can be delivered successfully in routine NHS care to specified groups of children with behaviour problems, and that those with particularly high levels of behaviour problems may benefit most. Furthermore, these problems can be identified using a simple, brief screening questionnaire. There is significant scope for this intervention to be incorporated in routine practice.

Key implications for future research include the following. First, further study is needed to assess the potential longer-term outcomes of early interventions such as VIPP-SD. Second, further study is needed to investigate whether or not the benefits of this early intervention can be enhanced with the addition of booster sessions or other later intervention. Third, future research is needed to elucidate the mechanisms underlying effective early interventions such as VIPP-SD and for whom the intervention may work best.

Early intervention represents a substantial opportunity for the future positive development of young children and a lack of effective interventions is a key challenge. The results of this study provide a significant step forward and represent a new opportunity for effective early childhood intervention to prevent enduring mental health problems.

Trial registration

This trial is registered as ISRCTN58327365.

Funding

This project was funded by the National Institute for Health Research (NIHR) Health Technology Assessment programme and will be published in full in *Health Technology Assessment*; Vol. 25, No. 29. See the NIHR Journals Library website for further project information.

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

Chapter 1 Introduction

Behaviour problems and their impact on health and development

Behaviour problems are among the most common mental health problems in young children, affecting an estimated 5–10% of children.[1,2] These disorders typically include attention deficit hyperactivity disorder (ADHD), conduct disorder and oppositional defiant disorder,[3] and are a key concern for the NHS. According to the National Institute for Health and Care Excellence (NICE),[4] ≈ 30% of a typical general practitioner's (GP's) child consultations are for behaviour problems and 45% of community child health referrals are for behaviour disturbances. Behaviour problems are also one of the most common reasons for children to be referred to mental health services.[4,5]

Behaviour problems are distinct in having one of the earliest onsets of mental health problems, with increasing research indicating that early symptoms can be identified in children aged 1 and 2 years.[6] Where problems endure over time, they can give rise to poorer educational attainment and physical health, as well as elevated risk of psychiatric disorders, substance misuse, antisocial behaviour and criminality.[2,5,7,8] In addition to the distress caused to individuals and families, there is a considerable cost to society through the health-care, social care and criminal justice systems.[9] Recent estimates put the lifetime costs of support for a child with conduct disorder at £280,000.[10]

Interventions for behaviour problems

The quality of the early parental care that children experience has been causally linked to the development of behaviour problems. Where children experience low levels of sensitive parenting behaviour and greater use of harsh discipline, they are at a high risk of developing behaviour problems.[11] Consequently, these parenting behaviours, sensitivity and discipline, represent the key targets of most early intervention programmes in the UK, based on attachment theory and social learning theory, respectively.[12,13]

There is an established evidence base demonstrating positive effects of parenting interventions for preschool- and school-aged children's behaviour. An umbrella meta-analysis[14] of 26 meta-analyses identified 411 studies of parenting interventions for children with externalising behaviours. The authors found that those studies that reported on externalising behaviour showed evidence of moderate positive effects. A recent meta-analysis[15] of 154 randomised controlled trials (RCTs) also found a moderate positive effect of parenting interventions on children's behaviour. The authors also reported an individual participant data meta-analysis of 13 RCTs of a specific parenting intervention, Incredible Years, which showed similar positive effects on conduct problems and ADHD symptoms.[15] Estimates of the cost savings associated with such programmes suggest that parenting interventions could provide savings of approximately £16,425 per family over 25 years.[16] As most programmes target preschool- (aged 3 or 4 years) and school-aged children, we know less about the effectiveness of programmes delivered in infancy and toddlerhood. Intervening earlier in childhood could be more effective both clinically and economically, provided that there is reliable identification of children at risk, as there is increased opportunity to intercept psychopathology before it becomes embedded and to reduce the burden of suffering experienced by families.[17]

An evidence-based programme that can be delivered from aged 12 months is the Video-feedback Intervention to promote Positive Parenting and Sensitive Discipline (VIPP-SD) programme.[18,19] VIPP-SD represents a powerful combination of social learning theory and attachment theory, and targets the two key parenting behaviours: sensitivity and discipline. The goal of the intervention is to promote parents'

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

sensitivity (i.e. their capacity to identify their child's attachment cues and exploratory behaviour, and respond to them appropriately) and sensitive discipline, which involves a consistent but non-harsh response to challenging behaviour. The VIPP-SD programme has been developed systematically and has been tested in 12 RCTs[20–31] in a range of clinical and non-clinical populations. A recent meta-analysis[32] of these trials demonstrated combined positive effects on caregivers' sensitivity and children's behaviour problems. However, VIPP-SD has yet to be tested in a routine health service context in the UK.

Rationale for the research

Early intervention has the potential to improve behaviour problems before they become established, which could yield cascading benefits for children's outcomes across their life course. Indeed, early intervention has become a key priority of global and domestic policy.[33–35] However, there are few effective early psychological interventions that target behaviour problems in very young children.[36] The literature also demonstrates that no single programme or service delivery method has been shown to be a panacea for the varied challenges that families face.[37] Rather, effective early intervention for children at risk of behaviour problems is likely to require the identification of a range of efficient and repeated interventions at different developmental stages.[38] This is in keeping with global policy recommendations on the importance of staged and developmentally informed intervention across early childhood that is grounded in nurturing caregiving.[39] Successful early intervention, therefore, first requires effective programmes that can target behaviour problems at their earliest onset in children aged 1 and 2 years. The Healthy Start, Happy Start study was designed to address this gap by testing whether or not a brief parenting intervention (VIPP-SD) could be effective in preventing enduring behaviour problems in children aged 1 and 2 years in a pragmatic health service context.

Aims and objectives

The aim of the Healthy Start, Happy Start study was to evaluate the clinical effectiveness and cost-effectiveness of a brief early parenting intervention to prevent enduring behaviour problems in young children aged 12–36 months.

The objectives were to:

- undertake a RCT to evaluate whether or not a brief parenting intervention (VIPP-SD) leads to lower levels of behaviour problems in young children who are at high risk of developing these problems, compared with usual care in the NHS
- undertake an economic evaluation to assess the cost-effectiveness of the intervention compared with usual care.

Chapter 2 Methods

Design

The Healthy Start, Happy Start study was a pragmatic, assessor-blinded, multisite, two-arm, parallel-group RCT, to test the clinical effectiveness and cost-effectiveness of a brief parenting intervention (VIPP-SD) for parents of young children (aged 12–36 months) at risk of behaviour difficulties. VIPP-SD was predominantly delivered by trained therapists through NHS health visiting teams and was compared with receiving usual care alone.

Ethics and governance

Ethics approval for the study was given by Riverside Research Ethics Committee (14/LO/2071), which is part of the NHS Research Ethics Service. The trial was registered with the Integrated Research Approval System (IRAS) under the reference number 160786 and the National Institute for Health Research (NIHR) Clinical Research Network (CRN) Portfolio under the reference number 18423. The trial was registered with the International Standard Randomised Controlled Trial Number (ISRCTN) registry as number ISRCTN58327365. Approval for the study to be conducted within participating NHS sites was obtained from the NHS Health Research Authority (HRA). Local NHS permissions were also given for each recruitment site. The trial was conducted in accordance with the Declaration of Helsinki,[40] UK Policy Framework for Health and Social Care Research[41] and Imperial Clinical Trials Unit (ICTU) standard operating procedures.

Participants

Participants were children aged 12–36 months who demonstrated behaviour problems and their caregiver(s).

Inclusion criteria

To be eligible to participate, families had to meet the following inclusion criteria:

- participating caregivers were aged $\geq$ 18 years
- child was aged between 12 and 36 months
- child scored in the top 20% of population norms for behaviour problems on the parent-reported Strengths and Difficulties Questionnaire (SDQ)
- written informed parental consent from participating caregivers.

Exclusion criteria

Participants were not eligible to participate if any of the following criteria were met:

- child or parent had severe sensory impairment, learning disability or language limitation that was sufficient to preclude participation in the trial
- a sibling was already participating in the trial
- family was participating in active family court proceedings
- parent was participating in another closely related research trial and/or was receiving an individual video-feedback-based intervention.

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

Setting

Recruitment for the study was conducted in six UK NHS trusts: Central and North West London NHS Foundation Trust, Whittington Health NHS Trust, Oxford Health NHS Foundation Trust, North East London NHS Foundation Trust, Cambridgeshire and Peterborough NHS Foundation Trust, and Hertfordshire Community NHS Trust. Recruitment efforts were predominantly targeted across seven areas within these trusts: the London Boroughs of Camden, Hillingdon, Islington, and Barking and Dagenham, as well as Oxfordshire, Peterborough and Hertfordshire. Research assessments were carried out in participants' homes or in another setting if the participant preferred (e.g. a private room in a children's centre).

Recruitment procedure

There were two stages of recruitment in the study. In stage 1, potential participants in participating NHS sites were screened for behaviour difficulties using the parent-reported SDQ.[42] The principal pathway for recruitment was through health visiting services, with health visitors screening families using a study screening pack during routine 12- and 24-month child health reviews. This was supplemented by direct screening undertaken by members of the research team and CRN support staff in the waiting room of these clinics and children's centres. Health visiting services also posted screening information to families with children within the target age range, and advertisements for the study were placed in health visiting centres and GP surgeries. Participants were also recruited through wider community services including nurseries, libraries, and through social media advertisements in local and community forums. Recruitment advertisements included the questions '[I]s your child's behaviour sometimes a challenge?' and '[A]re you interested in finding out more about your child's behaviour?'.

In stage 2 of recruitment, the research team contacted caregivers by telephone if their children met the eligibility criteria based on their scores on the SDQ questionnaire (i.e. scoring in the top 20% of population norms). During this telephone call, the family's full eligibility and interest in participating in stage 2 of the study were assessed. Researchers also enquired as to whether or not there were two parents/caregivers in the family and, if so, if both would be willing and available to participate in the study. Interested participants were sent an information pack for stage 2 of the study by e-mail or post, and a date to meet with them at their home for the baseline assessment visit was scheduled.

Informed consent

Participating parents/caregivers provided written informed consent for each stage of the study. All participant resources were ethics approved [see www.journalslibrary.nihr.ac.uk/programmes/hta/130433/#/documentation (accessed January 2021) for participant information sheets (PISs) and consent forms used through both stages of the trial, as well as template letters used to inform GPs and health visitors of families' study participation]. In stage 1, potential participants were given screening packs containing an information leaflet and the stage 1 PIS and consent form, alongside the screening questionnaire. Parents/caregivers were able to provide written or electronic consent to complete the screening stage of the study. Consent for stage 2 of the study was not provided at this point.

Participants who were eligible for and interested in the second stage of the study, the full RCT, were sent the stage 2 PIS either by e-mail or by post following a telephone call from the research team and ahead of their scheduled baseline assessment. The PIS and the clauses of the consent form were then fully explained in person to participants by a trained researcher at the beginning of the research assessment. Participants' right to withdraw from the trial at any point was explained. Participants were given the opportunity to ask any questions that they had.

Baseline assessment

Baseline assessments were carried out in participants' homes (or, in a very small number of cases, in private rooms of children's centres close to participants' homes). At baseline, trained researchers obtained written informed consent and collected family demographic information. Parent demographic information included parents' sex, age, ethnicity, education level, employment status and relationship status. Child demographic data included children's sex, age and ethnicity. Measures administered during the baseline assessment can be seen in *Table 1*.

The baseline visits took between 2 and 2.5 hours to complete (depending on the number of caregivers participating). All baseline assessments were conducted between July 2015 and July 2017.

Follow-up

Follow-up visits took place at two time points. The first was a post-treatment assessment (i.e. after the VIPP-SD intervention for those in the treatment group), approximately 5 months post randomisation, whereas the second was scheduled for 24 months post randomisation, when the children were aged 3–5 years. These assessment points were timed to allow for an initial post-intervention assessment when treatment would have usually been completed (5-month assessment) and to assess for longer-term outcomes (24 months). To assess whether or not the intervention had clinically significant effects on child behaviour problems, the measures of child behaviour collected at baseline were also administered at both follow-up visits. Similarly, measures of self-reported parenting practices, as well as parent mental health measures, and couple functioning, where appropriate, were also

TABLE 1 Timing of data collection

Data	Measure	Source	Timing of data collection		
			Baseline	5-month follow-up	24-month follow-up
Demographics[a]	Parents' and children's sex, age, and ethnicity; parent education level, employment status and relationship status	Interview	✗		✗
Primary outcome	PPACS	Interview	✗	✗	✗
Secondary outcomes	CBCL	Questionnaire	✗	✗	✗
	SDQ (parent report)	Questionnaire	✗	✗	✗
	SDQ (teacher report)	Questionnaire			✗
	PHQ-9	Questionnaire	✗	✗	✗
	GAD-7	Questionnaire	✗	✗	✗
	Parenting Scale	Questionnaire	✗	✗	✗
	RDAS	Questionnaire	✗	✗	✗
Economic data	CA-SUS	Interview	✗	✗	✗
Implementation data	Participant feedback questionnaire[a,b]	Questionnaire		✗	
	Intervention fidelity	Audio recordings		✗	

CA-SUS, Child and Adolescent Service Use Schedule; CBCL, Child Behaviour Checklist; GAD-7, Generalised Anxiety Disorder-7; PHQ-9, Patient Health Questionnaire-9; PPACS, Preschool Parental Account of Children's Symptoms; RDAS, Revised Dyadic Adjustment Scale.
a See www.journalslibrary.nihr.ac.uk/programmes/hta/130433/#/documentation (accessed January 2021).
b This measure was sent to and returned by participants by post.

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

repeated at follow-up time points. Families were also asked to complete a structured interview of service use at both follow-up time points to assess cost-effectiveness. Administration of all trial measures is outlined in *Table 1*.

The follow-up visits took between 2 and 2.5 hours to complete. All follow-up data were collected between December 2015 and July 2019. *Figure 1* shows the Consolidated Standards of Reporting Trials (CONSORT) flow diagram outlining participant recruitment and retention during the study.

Intervention

Video-feedback Intervention to promote Positive Parenting and Sensitive Discipline is a manualised, home-based intervention developed by Juffer, Bakermans-Kranenburg and van IJzendoorn and based on attachment theory and social learning theory.[18] The goal of the intervention is to promote parents' sensitivity (i.e. their capacity to identify their child's attachment cues and exploratory behaviour, and respond to them appropriately), as well as sensitive discipline, which involves a consistent but non-harsh response to challenging behaviour. VIPP-SD was delivered by trained therapists over six sessions, lasting 1–2 hours, at approximately fortnightly intervals. The role of the therapist was to develop a trusting and empathetic relationship with the parent(s) and to deliver the six sessions (four core sessions and two booster sessions) in line with the manual.

The first half of each session involved the therapist filming the parent with their child during interactions that aimed to reflect everyday situations, such as reading a book together, playing with toys and a mealtime. Following the session, the therapist used the video-recorded interactions to write a feedback script based on an appraisal of the parent's interaction profile needs, the guidelines of the protocol, and the sensitivity and discipline themes for that session. The parent and therapist spent the second half of each session reviewing the interactions filmed in the previous visit, with the therapist frequently pausing the video to deliver feedback based on the script they had prepared following the last session. The content and focus of each session were manualised, but the feedback was delivered in an individualised way, as feedback is unique to each parent and child in that moment. A modified version of VIPP-SD, Video-feedback Intervention to promote Positive Parenting for Co-parents (VIPP-Co),[43] was delivered when two caregivers in the family were participating. The content and themes of the VIPP-Co intervention broadly mirrored the VIPP-SD manual, with additional emphasis on interactions involving both caregivers together with the child and on positive co-parenting. In the first three sessions, feedback was delivered separately to each caregiver (as in VIPP-SD) and, from session 4 onwards, feedback was delivered to both caregivers together. The intervention was delivered by 40 trained therapists, the majority of whom were health visitors, community nursery nurses and clinical psychologists, as well as a small number of professionals from other backgrounds, such as family and child therapy, psychology and psychiatry.

Fidelity

Each therapist delivering the intervention undertook 4 days of VIPP-SD training, as well as an additional day of training on delivering the intervention to two caregivers and on the study protocol. Therapists undertook supervised clinical practice with a practice case before becoming a therapist on the trial and received regular clinical supervision throughout the trial. Therapists audio-recorded the feedback of sessions and reported on fidelity in terms of the delivery of key components of the treatment, as well as reporting on global adherence to the manual. We randomly selected 10% of audio-recordings to allow two assessors trained in the intervention to rate the quality of sessions and adherence to the VIPP-SD manual.

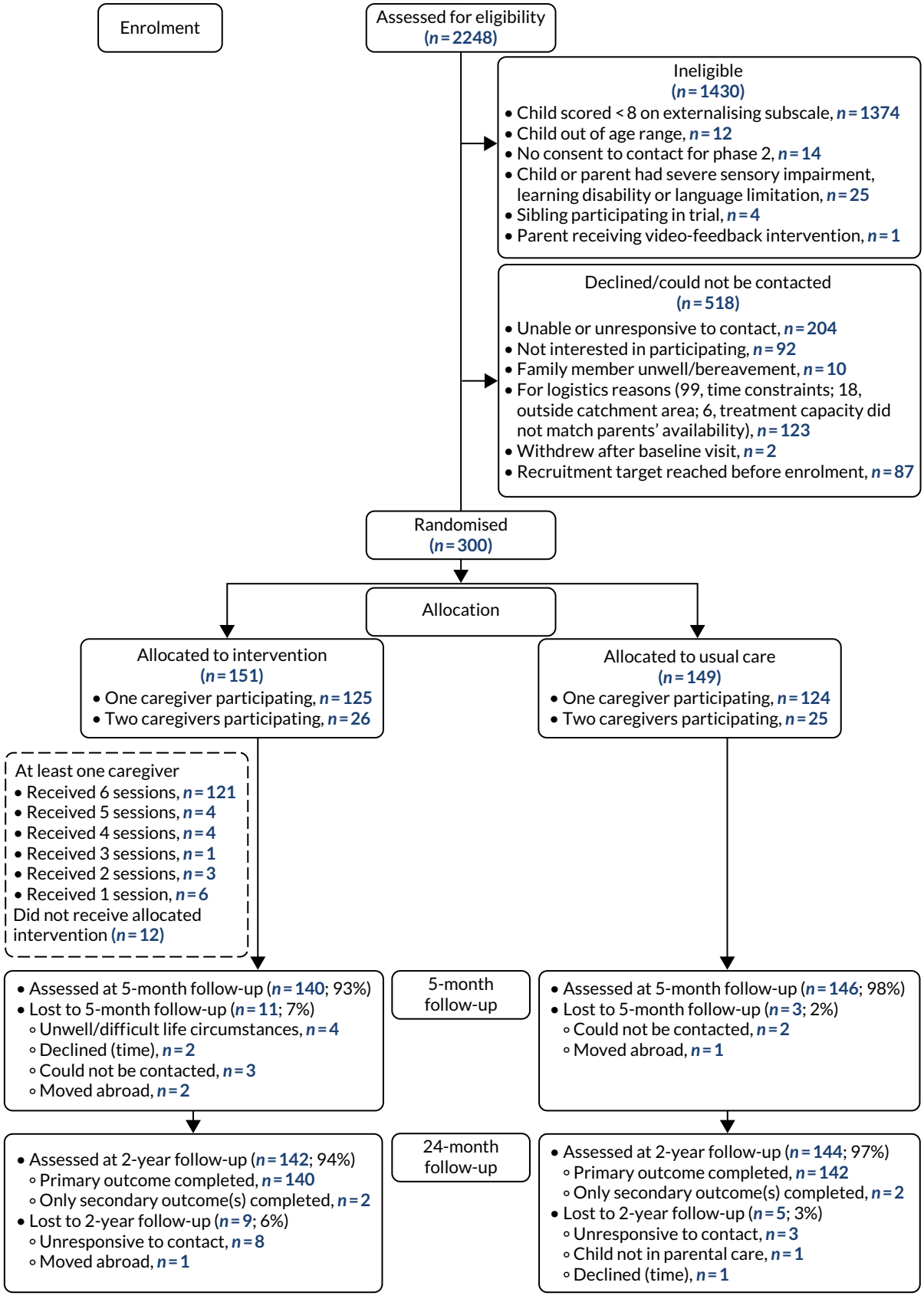


FIGURE 1 The CONSORT flow diagram for the Healthy Start, Happy Start RCT.

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

Usual care

Participants in both groups continued to receive their usual care, which was minimal in most cases (there are no standard care pathways in the NHS for early-onset behaviour problems). Some participants received support and advice from a health visitor or GP, referral to early intervention mental health services linked to a children's centre, or parenting advice and support sessions. Data were collected on concurrent use of health and social care services and are presented in *Chapter 4*.

Outcome measures

Primary outcome

The primary outcome for the trial was the assessment of the severity of behaviour problems at 5 months post randomisation, measured using a modified version of the Preschool Parental Account of Children's Symptoms (PPACS)[44,45] (see *Appendix 1*). The PPACS is a semistructured interview administered by trained researchers and conducted with the child's primary caregiver. Interview measures are considered the gold standard outcome measure as they can circumvent potential biases related to parent-reported outcomes.[46] We made minor adaptations to the measure for use in the current study to ensure its suitability for use with children in the sample's age range.

During the PPACS interview, the primary caregiver is asked to recall and describe detailed examples of their child's typical behaviour over the past week in a range of settings (e.g. in the home, with friends, in public). The objective of this approach is to allow the interviewer to rate the child's behaviour based on real examples, rather than caregivers' global impressions or judgements of whether or not the behaviour is normal. To ensure that the example given is characteristic of the child, caregivers are asked how representative the described behaviour is of the child in the last 4 months. The researcher rates the severity and frequency of the child's symptoms based on professional judgement, following training, guided by written definitions and thresholds of each of the scored behaviours. The interview is used to score the children on two subscales; the first measures ADHD/hyperkinesis, whereas the second measures conduct problems.

This measure has high inter-rater reliability and good construct validity, and has been used as an outcome measure in a number of other clinical trials assessing intervention effects on child behaviour.[47–49] Thirty interviews were randomly selected by the study statistician to be double-scored by another trained researcher to ensure inter-rater reliability in the current study. One-way, single-measurement, absolute agreement intraclass correlation coefficients were calculated for each time point and indicated high levels of inter-rater reliability among researchers (with intraclass correlation coefficients as follows: baseline – total score, 97; conduct subscale, 93; hyperactivity subscale, 97; 5-month follow-up – total score, 96; conduct subscale, 95; hyperactivity subscale, 95; and 24-month follow-up – total, 92; conduct subscale, 72; and hyperactivity subscale, 98).

Key secondary outcomes

Based on prespecified hypotheses in the trial protocol, there were two key secondary outcomes measuring child behaviour.

Strengths and Difficulties Questionnaire

The SDQ[42] was used to measure behaviour problems as it is routinely used in both research and practice settings the UK. Participating caregivers were asked to complete this questionnaire. At the 24-month follow-up, the child's nursery or school was also asked to complete the questionnaire, if parental consent was given. The measure comprises 25 items that respondents rate on a three-point scale (0 = not true, 1 = somewhat true or 2 = certainly true). The items are divided between five subscales: emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems and prosocial behaviour. Scores on the conduct problems and hyperactivity scales can be combined to

provide an overall externalising behaviour score. The preschool version of the SDQ (for children aged 2–4 years) was used. The psychometric properties and utility of this version have been established in children aged 2 and 3 years.[50,51]

Child Behaviour Checklist

Behaviour problems were also measured using the Child Behaviour Checklist (CBCL).[52] The CBCL is a 100-item questionnaire that asks parents to rate how true the behaviour is of their child over the last 2 months on a three-point scale (0 = not true, 1 = somewhat true or 2 = very true or often true). The measure gives a total score, an externalising score and an internalising score. The externalising score is made up of two syndrome subscales for attention problems and aggressive behaviours. The internalising score is made up of items from the emotionally reactive, anxious/depressed, somatic complaints and withdrawn syndrome subscales. The CBCL was selected for use as well as the SDQ because it has been validated in children as young as aged 12 months and is widely used in research studies.[53]

Other secondary outcomes

Parenting scale

Parenting practices were assessed using the self-reported Parenting Scale,[54] which is a measure of dysfunctional discipline practices in parents. The Parenting Scale is a 30-item questionnaire that asks parents to rate how they would respond to various behaviour problems, with each item receiving a score of 1–7 (effective to ineffective strategies).

Patient Health Questionnaire-9

Self-reported depression severity and symptomatology were measured using the Patient Health Questionnaire-9 (PHQ-9).[55] The PHQ-9 is a nine-item questionnaire, with each statement corresponding to one of the nine criteria for depression outlined in the *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition (DSM-IV).[56] Each statement is scored for the frequency in which the parent has experienced each problem over the last 2 weeks. Scores range from 0 (not at all) to 3 (nearly every day), with a total score obtained by summing all items of the questionnaire.

Generalised Anxiety Disorder-7

Parental anxiety was assessed using the Generalised Anxiety Disorder-7 (GAD-7).[57] The GAD-7 is a seven-item questionnaire that asks respondents how often they have experienced each problem in the last 2 weeks. Each statement is scored from 0 (not at all) to 3 (nearly every day), with a total score obtained by summing all items of the questionnaire.

Revised Dyadic Adjustment Scale

Relationship adjustment was measured using the Revised Dyadic Adjustment Scale (RDAS).[58] The RDAS is a 14-item questionnaire consisting of three subscales: dyadic consensus, dyadic satisfaction and dyadic cohesion. Scores range from 0 to 69, where higher scores indicate greater relationship satisfaction and lower scores suggest greater relationship distress.

Child and Adolescent Service-Use Schedule

Information on the use of health and social care services was recorded in an interview using a modified version of the Child and Adolescent Service Use Schedule (CA-SUS).[59] Modifications to the CA-SUS were based on a review of recent literature and clinical feedback. This is an interview measure conducted with the child's primary caregiver that asks parents to recall use (number and duration of appointments/sessions) of key services. Data were collected on the use of accommodation services (e.g. foster care, supported housing), hospital services (e.g. inpatient stays, outpatient contacts, accident and emergency attendances), community-based health and social care services (e.g. contacts with GPs, clinical psychologists) and all prescribed medication. Parents/caregivers were interviewed at baseline and asked about their child's use of services and/or their use of services in relation to their child's needs in the previous 3 months. At subsequent follow-up points (5 and 24 months), service use was

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

recorded for the period since the previous interview to ensure that the entire duration of the trial had been captured.

Feedback questionnaire

Parents who were allocated to the intervention group were also asked to complete a feedback questionnaire [see www.journalslibrary.nihr.ac.uk/programmes/hta/130433/#/documentation (accessed January 2021)] following the 5-month follow-up assessment visit to explore their satisfaction with and experiences of the VIPP-SD programme. The questions were related to their perceived impact of the programme (in terms of their relationship with their child and their child's communication and behaviour) and their satisfaction with the home-based delivery format.

Serious adverse events

Serious adverse events (SAEs) were defined as 'an adverse event that results in death, is life-threatening, requires hospitalisation or prolongation of existing inpatient's hospitalisation, results in persistent or significant disability or incapacity' [reproduced with permission from the European Medicines Association. *Guideline for Good Clinical Practice E6(R2)*[60]]. The occurrence of SAEs was collected at the 5- and 24-month follow-up assessments. Child hospitalisation data were collected using the CA-SUS. Owing to the low-risk nature of this trial, non-serious adverse events were not collected.

Sample size

A sample size of 300 children and their caregiver/parent(s) was selected to provide between 80% and 90% power to detect standardised effect sizes of 0.36 and 0.42, respectively, on the primary outcome, at a 5% level of statistical significance, and assuming a 20% attrition rate. The analysis was adjusted for baseline behaviour score, time from randomisation to follow-up, recruitment site, age of child at recruitment, and caregiver involvement (one vs. two participating caregivers) as fixed effects. This is likely to have increased power to > 90% as such adjustment, especially for baseline scores of the same variable, will have reduced the residual error variance in our model.

Randomisation and blinding

The randomisation list was prepared by an independent statistician using block sizes of two, four and six (varying at random) and 1 : 1 allocation to either VIPP-SD or usual care. The randomisation list was uploaded to the study electronic data capture system before the trial commenced. Randomisation was performed using a web-based randomisation system linked to the trial database following enrolment by the researcher who conducted the baseline assessment and who was blind to all treatment allocations. Eligible participants were allocated online to the next available treatment code in the appropriate stratum. Randomisation was stratified by recruitment site and the number of caregivers participating (one vs. two). Access to the allocation sequence was restricted to an independent statistician and appropriate members of the InForm™ (version 4.6) technical support team (Imperial College London, London, UK) to maintain allocation concealment.

Following randomisation, the trial manager informed participants of their allocated groups and matched participants to therapists for treatment according to the availability of both the therapists and the parents. All assessors remained blind to treatment allocation for the duration of the study. Participants were reminded of the importance of researchers remaining blinded to group allocation by the trial manager or administrative assistant at each contact when scheduling their follow-up assessments. Researchers also reminded parents of the importance of blinding in person at the start of the follow-up visits to ensure that researchers stayed blinded to group allocation during the assessment. In cases where unblinding did occur, the primary outcome (PPACS) was re-scored by a blinded trained assessor.

Data collection and data management

The trial's source data included paper forms, questionnaires, written interview notes, scoring from researchers' research assessments, and feedback notes and log books from trial therapists' intervention sessions. Source data were collected and stored securely in accordance with the International Council for Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) good clinical practice (GCP),[61] the Data Protection Act[62] and the UK Policy Framework for Health and Social Care Research.[41] Following the assessments/intervention visits, data were entered on an electronic Case Report Form (eCRF) developed in InForm™, an electronic data capture system built around an Oracle database. The InForm system includes automated range checks and validation rules for data entry to help ensure data accuracy. A computer-generated audit trail is in place that records the date, time, operator, operation and previous value of all manipulation of clinical data.

InForm storage and management were undertaken by the Imperial College Information and Communication Technologies (ICT) team. InForm sits on a server behind a firewall connected to the College Storage Area Network (SAN). The data are backed up regularly to removable media, which allows for disaster recovery. In addition to the College backup facility, every 20 minutes, the activity logs for the trial are moved to another server in a different location to facilitate rapid recovery of data should it become necessary (e.g. in a disaster recovery scenario). Access to the system was restricted to trained staff with unique password-protected accounts. Identifiable data were not recorded in the eCRF and participants were identified by a unique trial identifier only.

All data monitoring and cleaning were also completed on InForm. Predefined data ranges were included in the eCRF, which raised automated queries if data outside the expected range were entered. In addition to the automated queries, the trial data were reviewed on a regular basis by the trial manager and trial statistician for discrepancies and errors. Data cleaning was performed prior to each Data Monitoring and Ethics Committee (DMEC) meeting and prior to database lock. Final data checks were performed by the statistician once the database had been soft locked and before hard lock was complete. All outstanding queries were resolved prior to the database hard lock.

Quality assurance and monitoring

Quality assurance (QA) and monitoring were performed in accordance with ICTU standard operating procedures (SOPs). A risk assessment was conducted by the ICTU QA manager prior to the start of the study, which assigned a risk category of 'low risk' to the trial. A monitoring plan was developed in accordance with the risk assessment to define how monitoring procedures would be carried out and the level of source data verification required for the study.

In accordance with the monitoring plan, 100% of participant consent forms and SAEs were source-verified. Further to this, 20% of key trial data were also source-verified against the original paper records, which included the primary outcome (PPACS); two of the secondary outcomes (i.e. CBCL and SDQ); and 10% of demographics, therapist logbooks and all other outcomes (Parenting Scale, PHQ-9, GAD-7, RDAS and CA-SUS).

Discrepancies noted between the paper records and trial database were queried and corrected in the trial database. Monitoring was carried out centrally by the research team, ensuring that the member of staff carrying out source data verification was not the same person who had performed data entry in each case. Participant data to be monitored were identified by the trial statistician through random selection.

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

Data analysis

A comprehensive statistical analysis plan was produced before undertaking the analysis and agreed with the Project Management Group (PMG), Trial Steering Committee (TSC) and DMEC. Primary analyses were conducted on an intention to treat (ITT) basis.

Multiple linear regression was used to assess the primary outcome measure, the difference in PPACS score between the trial groups at 5 and 24 months post randomisation. The PPACS score at the 5-month follow-up was the dependent variable, whereas trial group, PPACS score at baseline, time since randomisation, recruitment centre, age of the child at recruitment and number of parents/caregivers participating (one or two) were included as independent variables. Separate models were fitted at 24 months (adjusting for the baseline measurements, but not the 5-month measurements). These were distinct from the models fitted at 5 months (which, again, were adjusted for baseline measurements). Similar models, with adjustment for the appropriate baseline score, were used to assess the difference in CBCL and SDQ scores between the trial groups at the 5-month follow-up.

Regression residuals were plotted to assess their fit to the normal distribution using probability plots. PPACS scores (total and subscales) were normally distributed at 5 months and close to normal at 24 months. Inspection of residuals plots suggested assumption of homoscedasticity and of linearity were also reasonably well met. Visual inspection of the residuals from the regression analyses of secondary outcomes at 5 and 24 months suggested that they were close to normally distributed. Repeating the regression models using bootstrap methods (that allow for departure from normality) produced very similar results. For consistency, the results based on linear regression only are presented.

Secondary outcomes that were completed by each participating caregiver were analysed separately by sex. Standardised effect sizes (also known as *d*) were calculated by dividing the differences in means by the standard deviation in the usual-care group at follow-up.

Missing answers to individual questions in the PPACS were imputed. Unrateable or missing items in the attention/hyperactivity scale occurred most often when children did not regularly undertake an activity required for scoring (e.g. watching television or playing with a sibling/friend on the same activity). This assumption of 'missing at random' inherent in the multiple imputation analysis seems reasonable here. This implies assessing attention based on the activities that the child regularly undertakes (e.g. playing alone and eating a meal), by imputing values for activities that the child does not regularly undertake (e.g. watching television and/or playing with a friend or sibling). There were negligible missing data in the conduct disorder scale. The level, pattern and likely causes of any missingness in the baseline variables and primary outcome were investigated. Multiple imputation was of individual missing items in the PPACS interview when it was partially completed (because this always implied that at least 50% of items in the scale had been completed). Multiple imputation of the whole score was performed if the entire scale was missing. At the 5- and 24-month follow-ups, imputation was based on randomised group, sex of child, age of child at the 5- and 24-month assessments and baseline score, utilising information from both the baseline and 5- and 24-month PPACS interviews. Imputation at baseline was based on randomised group, sex of child, age of child at baseline and information from the baseline PPACS questionnaire.

Missing items in the other secondary outcome scales (child behaviour as assessed by the CBCL and SDQ; and self-reported parental discipline behaviour, mood, anxiety and couple functioning) were dealt with by scaling up. Thus, the sum of the maximum possible scores for non-missing items in each scale was divided by the maximum possible total score for a fully completed scale to give the proportion of the scale that had been completed (inflation factor). Scaling up was achieved by dividing the observed total score for each scale by the inflation factor. The proportion of missing items was generally negligible. See *Appendix 2* for additional information on multiple imputation.

Analysis of the primary outcome was repeated without adjustment for time since randomisation and using complete cases after multiple imputation of missing items only. Analysis was also repeated assuming that those who did not have PPACS follow-up data available had better than anticipated results and then worse than anticipated results (anticipated by multiple imputation, better and worse by one standard deviation of the difference between PPACS scores at follow-up and at baseline). In further analyses, unrateable items within PPACS were filled in as the best possible scores and then as the worst possible scores. Analysis of primary and secondary child behaviour outcomes was repeated for each domain separately. Complier-average causal effects (CACE) analysis was performed, using two-stage least squares regression analysis, to determine the effect of receiving the intervention, rather than just being randomised to receive it. Compliance was defined as receipt of the four core VIPP-SD visits.

Planned subgroup analyses of the primary outcome were of the effects of child age at baseline (12–23 months compared with 24–36 months) and of the number of parents participating (one vs. two). We also undertook an ad hoc subgroup analysis to assess the effect of quartiles of severity of child behaviour problems at baseline and of fidelity of the intervention in those randomised to VIPP-SD. The magnitude of the treatment effect was estimated by repeating analysis of the primary outcome for each subgroup separately (and for each fidelity group compared with all patients allocated to the usual-care group).

All analyses were conducted using Stata® (StataCorp LP, College Station, TX, USA) software (versions 13.0 and 15.0).

Trial organisation and management

Trial management

The trial was managed by the UK Clinical Research Collaboration registered ICTU, including statistics, operations, database development and QA. The trial manager was based with the chief investigator at the Division of Psychiatry in Imperial College London, working in collaboration with the ICTU.

The trial was overseen by three main oversight committees: (1) TSC, (2) PMG and (3) DMEC. A patient and public involvement (PPI) group was also convened.

Trial Steering Committee

The TSC provided overall supervision of the trial, monitoring the progress of the study, ensuring that there were no major protocol deviations and providing advice to the trial investigators. The TSC comprised an independent chairperson, two members of the PPI group and, in general, four additional independent members, as well as non-independent members, including the chief investigator, the trial and operations managers and the senior study statistician. Membership of the TSC was approved by the NIHR Health Technology Assessment (HTA) programme prior to convening the committee. The TSC met prior to the commencement of the study, every 6 months during recruitment (2015–17) and annually during follow-up. The TSC agreed on a charter, in line with NIHR guidelines,[63] outlining its responsibilities at the first meeting.

Project Management Group

The PMG was responsible for overseeing the management of the study and operational issues. The PMG met every 2 months during the set-up phase of the trial and quarterly thereafter. Members of the PMG included the chief investigator, investigators, the trial and operations managers, and the trial statistician.

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

Data Monitoring and Ethics Committee

The independent DMEC was responsible for overseeing the safety and data quality of the trial and met annually during the trial. The committee monitored and reviewed trial recruitment, adherence to the protocol, SAEs and interim outcome data presented by the trial statisticians. The NIHR HTA programme approved and invited all members of the committee. The DMEC agreed on a charter outlining its roles and responsibilities at the first meeting. The charter was prepared in line with the Data Monitoring Committees: Lessons, Ethics, Statistics (DAMOCLES) charter[64] and NIHR guidelines.[63]

Patient and public involvement group

Our PPI group was composed of seven representatives and included members of the pilot study, as well as those identified through community and health-care services. Group members were mothers and fathers of young children and/or educational/child care professionals. The PPI group contributed to the study's progress throughout the trial and provided significant and important input on the design and management of the study. The group met annually from the start of the trial in 2015 until the end of the trial in 2019, with an average of four members attending each meeting.

Public involvement was a key element in the study's success, with the PPI group providing ideas, advice and feedback that were integral to the study's recruitment processes, participant retention, data collection procedures and dissemination of study findings. Specifically, the group was able to advise and offer feedback on the optimal settings and services needed to be targeted for trial recruitment, especially for populations, such as fathers, that previous studies have often struggled to engage and accommodate. The PPI group also helped to produce and develop a brief, informative video for use in recruitment purposes. The video outlined what participation in the trial would involve and was provided to families alongside the PIS. A study within a trial (SWAT) aimed to explore whether or not the addition of this video aided the consent process.

A key study success was sustaining high levels of participant retention across the 2-year follow-up period (95% and 94% retention at the 5- and 24-month follow-ups, respectively). The PPI group was, again, integral to this process and advised on the key engagement strategies implemented by the study team. These included the creation of a study website for caregivers to visit, a welcome pack for families, the use of certificates and visit trackers for the children following each assessment, the sending of birthday cards to the children, a thank-you toy to acknowledge children's participation, and the sharing of sample clips from the visits as a keepsake. Regular newsletters sent to participating families by e-mail were also a key initiative of the PPI process. The newsletters provided an update on the study, informed participants about what was coming next and constituted a pathway for families to easily update the study team with any changes in contact details. This meant that the study team was able to communicate more efficiently and effectively with families when it came to arranging follow-up assessments. The PPI group was also involved in decisions around how to communicate and explain the study findings to participants. The group recommended considering a range of different dissemination materials, including an animation, a brief newsletter, an infographic, and an example story to illustrate the study and its outcomes in a more personal way.

The PPI members were highly engaged and motivated throughout the project. At the start of each meeting, a study summary was provided, as well as a recap on how the group's suggestions and feedback (from the previous meeting) had been implemented in the wider project. Two members of the PPI group were also full members of the TSC and participated in the piloting of baseline and follow-up assessments. All PPI members were reimbursed for incurred travel expenses and were offered a voucher following each meeting in keeping with the INVOLVE guidelines.

Amendments to protocol

- We changed the eligibility criteria so that participants who had previously received individualised video-feedback programmes were no longer excluded. We also extended the eligibility criteria to exclude parents who had a sensory impairment or learning disability that precluded their participation in the trial.
- Timing of the primary outcome was extended from 4 to 5 months post randomisation because of scheduling difficulties in completing the treatment in the allocated time for the intervention group of the trial. This was to reduce the likelihood of systematic differences in the timing of the follow-up assessments between the trial groups.
- Reporting of outcomes in the current report is confined to those that were funded as part of this project and were included in the prespecified statistical analysis plan. Following an ethics amendment, we amended the protocol to include one secondary and five exploratory outcomes. As a secondary outcome, parent–child interaction data were collected across all three time points to allow for the measurement of parental sensitivity, which was a target of the intervention. This measure may be instructive in better understanding the mechanisms of this intervention. Exploratory outcomes were also added at the 24-month time point. These measures included direct assessment of children's executive functions, emotion regulation, delay of gratification and narrative representations during dolls house play, as well as genetic variance (through buccal swabs) and parental involvement. Analyses of these variables is yet to be undertaken and findings will be published once available.

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

Chapter 3 Clinical results

Participant flow and retention

Between 30 July 2015 and 26 July 2017, a total of 2248 families were assessed for eligibility. Of these families, 1430 were ineligible, 518 did not progress to the trial (declined/could not be contacted) and 300 were randomised to stage 2 of the trial. In total, 151 participants (50%) were randomly allocated to the VIPP-SD group and 149 participants (50%) to the usual-care group. We recruited to target ($n = 300$; see *Appendix 3* for the study's recruitment graph), with the majority of participants recruited from health visiting services (30% face to face and 25% through posting screening information to families with children in the target age range) and children's centres (30%). The remaining participants were recruited largely from other community venues and online adverts (*Table 2* shows a full outline of recruitment pathways). Participants were recruited across seven NHS sites (*Table 3*). Follow-up data were collected between December 2015 and July 2019. Follow-up data were analysed for 286 (VIPP-SD, $n = 140$; usual care, $n = 146$) participants and 282 (VIPP-SD, $n = 140$; usual care, $n = 142$) participants at 5 and 24 months, respectively. See *Figure 1* for the CONSORT flow diagram summarising the eligibility, trial group allocation and subsequent progress of randomised participants. Approximately one-third of children who were screened were eligible to take part in the study, which was higher than the anticipated 20%. It is possible that families with children with elevated problems self-selected into the study and/or that health visitors were more likely to complete screening with families who they thought may be more likely to be eligible based on the child's behaviour.

Baseline participant characteristics

Participant characteristics at baseline were well balanced between groups. *Table 4* shows the baseline characteristics of participating children and primary caregivers. The participating children had a mean age of 23 months [standard deviation (SD) 6.7 months]; 54% were male and 65% were recorded as being of white ethnicity. Participating primary caregivers were predominantly female (96%), with a mean age of 34.2 years (SD 5.8 years). Most primary caregivers were married or cohabiting (85%), were of white ethnicity (72%), had a graduate-level qualification (64%) and were in some form of employment or on paid parental leave (59%). There were more male children in the usual-care group (58%) than in the VIPP-SD group (50%), and the usual-care group was also more diverse in terms of primary caregiver and children's ethnicity. *Appendix 4*, *Table 24*, shows the baseline characteristics of participating secondary caregivers.

TABLE 2 Route of recruitment for families randomised into the trial

	Trial group, *n* (%)		
Recruitment route at screening	**VIPP-SD (*N* = 151)**	**Usual care (*N* = 149)**	**All (*N* = 300)**
Routine developmental review/clinic by health visitors	47 (31)	44 (30)	91 (30)
Health visiting service posting screening information	34 (23)	40 (27)	74 (25)
Children's centre	46 (30)	45 (30)	91 (30)
Other clinic/community venues	12 (8)	11 (7)	23 (8)
Online advert	9 (6)	5 (3)	14 (5)
Word of mouth	1 (1)	2 (1)	3 (1)
Other	2 (1)	2 (1)	4 (1)

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

TABLE 3 Participants recruited into the trial disaggregated by recruitment site

Recruitment site	Family recruited into RCT, *n* (%)	Family had two participating caregivers, *n* (%)
Barking and Dagenham	14 (5)	1 (2)
Camden	83 (28)	13 (25)
Hertfordshire	7 (2)	0 (0)
Hillingdon	22 (7)	5 (10)
Islington	105 (35)	23 (45)
Oxford	43 (14)	5 (10)
Peterborough	26 (9)	4 (8)
Total	300 (100)	51 (100)

TABLE 4 Baseline characteristics of participating children and primary caregivers by treatment allocation

	Trial group		
Characteristic	VIPP-SD (*N* = 151)	Usual care (*N* = 149)	All (*N* = 300)
Child characteristic			
Sex (male), *n* (%)	76 (50)	87 (58)	163 (54)
Age (months), mean (SD)	22.8 (6.8)	23.2 (6.5)	23 (6.7)
Ethnicity, n *(%)*			
Asian	9 (6)	8 (5)	17 (6)
Black	3 (2)	15 (10)	18 (6)
Mixed ethnicity	36 (24)	25 (17)	61 (20)
White	100 (66)	94 (63)	194 (65)
Other	3 (2)	7 (5)	10 (3)
Primary caregiver characteristic			
Sex (male), *n* (%)	8 (5)	5 (3)	13 (4)
Age (years), mean (SD)	33.7 (5.6)	34.7 (5.9)	34.2 (5.8)
Parental status, n *(%)*			
Parent (including step or adoptive)	151 (100)	149 (100)	300 (100)
Ethnicity, n *(%)*			
Asian	15 (10)	16 (11)	31 (10)
Black	3 (2)	15 (10)	18 (6)
Mixed ethnicity	11 (7)	11 (7)	22 (7)
White	114 (75)	103 (69)	217 (72)
Other	8 (5)	4 (3)	12 (4)

TABLE 4 Baseline characteristics of participating children and primary caregivers by treatment allocation (*continued*)

Characteristic	Trial group		
	VIPP-SD (*N* = 151)	Usual care (*N* = 149)	All (*N* = 300)
Relationship status, n *(%)*			
Married/civil partnership/cohabiting	128 (85)	127 (85)	255 (85)
Divorced/widowed/legally separated	1 (1)	4 (3)	5 (2)
Single and none of the above	12 (8)	17 (11)	29 (10)
In a relationship but not cohabiting	10 (7)	1 (1)	11 (4)
Employment status, n *(%)*			
Employed	66 (44)	64 (43)	130 (43)
Paid parental leave	6 (4)	10 (7)	16 (5)
Self-employed	20 (13)	12 (8)	32 (11)
Full-time student	3 (2)	7 (5)	10 (3)
Looking after home and children	56 (37)	56 (38)	112 (37)
Highest qualification, n *(%)*			
GCSE or lower	17 (11)	14 (9)	31 (10)
A level/NVQ/BTEC	42 (28)	36 (24)	78 (26)
Graduate	92 (61)	99 (66)	191 (64)

A level, Advanced level; BTEC, Business and Technology Education Council; GCSE, General Certificate of Secondary Education; NVQ, National Vocational Qualification.

Delivery of intervention

Therapists

Forty trained therapists delivered any amount of VIPP-SD (median three participants each; range 1–12 participants each). Therapists had a mean of 13 years' experience (SD 12 years' experience) post-training in NHS services. *Table 5* shows a full overview of therapist professions.

TABLE 5 Professional backgrounds of therapists implementing VIPP-SD

Profession/training background	*n* (%)
Health visitor	10 (25)
Community nursery nurse	8 (20)
Clinical psychologist	7 (17)
Trainee clinical psychologist	4 (10)
Researcher	4 (10)
Psychotherapist	3 (8)
Research nurse	2 (5)
Psychiatrist	2 (5)
Total	40 (100)

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

Intervention dosage

Of the 151 participants randomised to the VIPP-SD group, 129 participants (85%) completed at least four sessions (the compliance cut-off point for treatment adherence), with 121 of these participants (80%) receiving all six sessions. Twelve participants (8%) received no intervention, with a further 10 participants (7%) completing one, two or three visits only.

Intervention fidelity

Treatment fidelity was assessed against the VIPP-SD manual and was found to be acceptable in the majority of the sessions.[18] Two assessors trained in the intervention rated audio-recordings using a global scale of adherence to the manual on a five-point scale (1 = did not follow the manual at all; 2 = adapted most of the material, did not follow the manual closely; 3 = sometimes adapted the material, followed manual somewhat; 4 = adapted only minor elements, followed the manual quite closely; and 5 = followed the manual very closely and delivered the session as specified). A score of 3 was set as the acceptable fidelity threshold as to receive this score most core components of the intervention needed to be identified as present in the feedback. Following piloting of the scale, the assessors double-scored a subset of 10 sessions (comprising 31 scripts) and established an acceptable level of reliability [intraclass correlation coefficient (ICC) 0.69]. The fidelity assessment determined that 94% (72/77 randomly selected audio-recordings of therapy sessions) met the minimum threshold (score of ≥ 3) of adherence to the manual. The mean sum score of adherence across sessions was 3.66 (SD 0.60, range 3.4–4.0).

Primary outcome: Preschool Parental Account of Children's Symptoms score at the 5- and 24-month follow-ups

We present the primary outcome (total PPACS score) and its subscale scores at the 5-month (primary time point) and 24-month follow-ups in *Table 6*. The group mean difference in PPACS score was modelled using linear regression adjusted for baseline scores, treatment centre, randomised group, length of follow-up, age of the child and the number of participating caregivers. At the post-treatment 5-month follow-up, the mean scores of child behaviour problems on the PPACS were found to have decreased from 33.5 (SD 9.0) to 28.8 (SD 9.2) in the VIPP-SD group and from 32.4 (SD 10.6) to 30.3 (SD 9.9) in the usual-care group. This represents a significant difference in reduction of problems favouring the VIPP-SD group [adjusted mean difference, VIPP-SD vs. usual care 2.03, 95% confidence interval (CI) 0.06 to 4.01; $p = 0.04$]. The adjusted standardised effect size was Cohen's $d = 0.20$ (95% CI 0.01 to 0.40). VIPP-SD was also found to be superior to usual care, in particular, on the conduct problems subscale of the primary outcome [adjusted mean difference 1.61, 95% CI 0.44 to 2.78; $p = 0.007$ ($d = 0.30$, 95% CI 0.08 to 0.51)], but not the hyperactivity subscale [adjusted mean difference 0.29, 95% CI −1.06 to 1.65; $p = 0.67$ ($d = 0.05$, 95% CI −0.17 to 0.27)].

At the 24-month follow-up, the adjusted mean difference on the total PPACS score remained about the same, with a decrease in effect size of $d = 0.03$ [VIPP-SD vs. usual care 1.73, 95% CI −0.24 to 3.71; $p = 0.08$ ($d = 0.17$, 95% CI −0.02 to 0.37)]. The adjusted mean difference on the conduct subscale continued to favour the VIPP-SD group [VIPP-SD vs. usual care 1.07, 95% CI −0.06 to 2.2; $p = 0.06$ ($d = 0.20$, 95% CI −0.01 to 0.42)]. There was no evidence of the superiority of VIPP-SD over usual care on the hyperactivity subscale [difference 0.62, 95% CI −0.60 to 1.84; $p = 0.32$ ($d = 0.10$, 95% CI −0.10 to 0.30)].

The positive effect of VIPP-SD on the PPACS total score at the 5-month follow-up was robust to sensitivity analyses, which estimated the impact of varying the assumptions made in respect of missing data and adjustment for length of follow-up (see *Appendix 5*, *Table 28*). *Figures 2* and *3* present the forest plots for the sensitivity analyses of the 5-month and 24-month follow-up PPACS scores, respectively. *Appendix 5* details the profiles of families with and without missing PPACS data (see *Appendix 5*, *Table 29*) and reports descriptive outcome data for those with and without PPACS scores at the 5- and 24-month follow-ups (see *Appendix 5*, *Table 30*).

TABLE 6 Primary ITT analysis of adjusted mean difference on the primary outcome of children's behaviour problems at the 5- (primary time point) and 24-month follow-ups (and its subscales)

	Trial group						
	VIPP-SD		Usual care				
Outcome	*n*	Mean (SD)	*n*	Mean (SD)	Mean difference[a] (95% CI)	Effect size[b] (95% CI)	*p*-value
PPACS[c] total score – primary analysis at 5 months							
Baseline	151	33.5 (9.0)	149	32.4 (10.6)			
5 months	140	28.8 (9.2)	146	30.3 (9.9)	2.03 (0.06 to 4.01)	0.20 (0.01 to 0.40)	0.04
24 months	140	23.1 (8.2)	142	24.7 (9.9)	1.73 (–0.24 to 3.71)	0.17 (–0.02 to 0.37)	0.08
PPACS[c] conduct scale							
Baseline	151	16.0 (5.8)	149	15.5 (6.4)			
5 months	140	14.8 (5.1)	146	15.8 (5.4)	1.61 (0.44 to 2.78)	0.30 (0.08 to 0.51)	0.007
24 months	140	13.4 (4.8)	142	14.4 (5.3)	1.07 (–0.06 to 2.2)	0.20 (–0.01 to 0.42)	0.06
PPACS[c] ADHD scale							
Baseline	151	17.5 (5.8)	149	16.9 (6.6)			
5 months	140	14.0 (6.1)	146	14.5 (6.2)	0.29 (–1.06 to 1.65)	0.05 (–0.17 to 0.27)	0.67
24 months	140	9.7 (5.1)	142	10.3 (6.1)	0.62 (–0.60 to 1.84)	0.10 (–0.10 to 0.30)	0.32

a Difference in mean is the difference between treatment groups from linear regression analysis of the outcome measure on the baseline score of that same measurement, on treatment centre, randomised group, length of follow-up, age of the child and on the number of caregivers participating (all treated as fixed effects). Multiple imputation was used for any unratable items in PPACS and also for anyone without PPACS assessment at follow-up. Positive differences represent greater adjusted decreases in symptoms in the VIPP-SD group than in the usual-care group.
b Effect size is the standardised difference in mean (adjusted as above). This is the difference in mean divided by the SD of controls at follow-up.
c PPACS scores include multiple imputation for those with missing data at follow-up and for missing items within the PPACS scales.

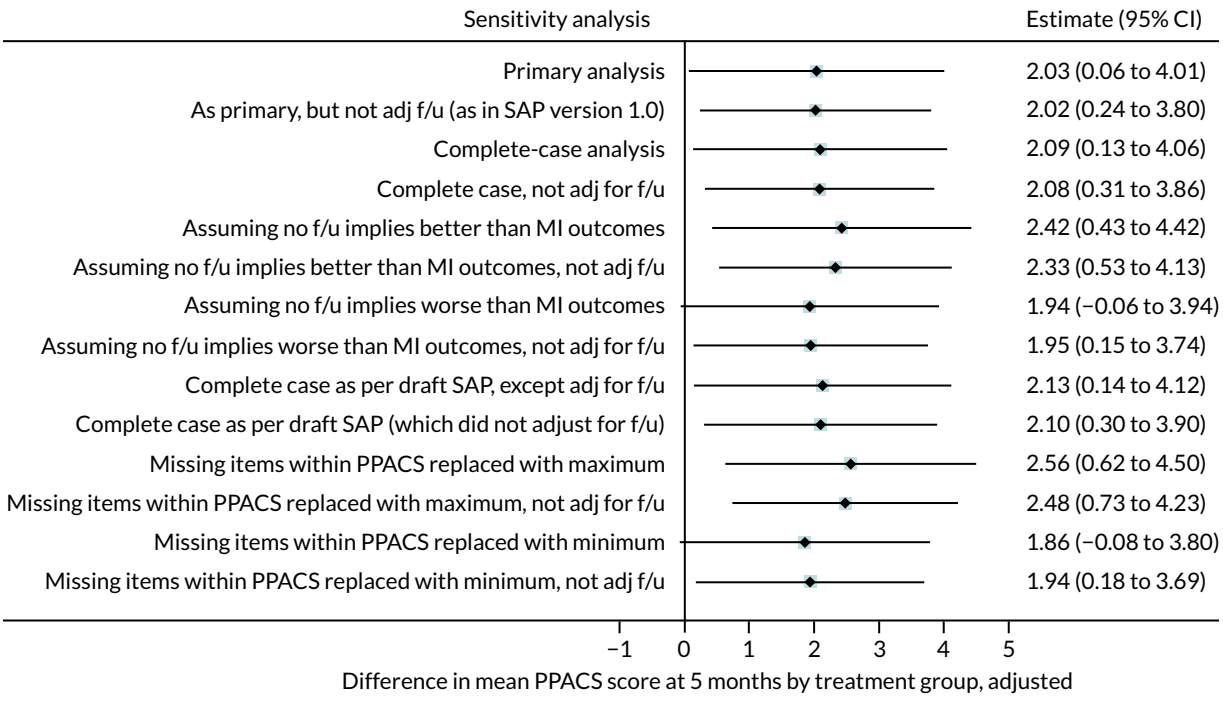


FIGURE 2 Forest plot for sensitivity analysis at the 5-month follow-up. adj, adjusted; f/u, follow-up.

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

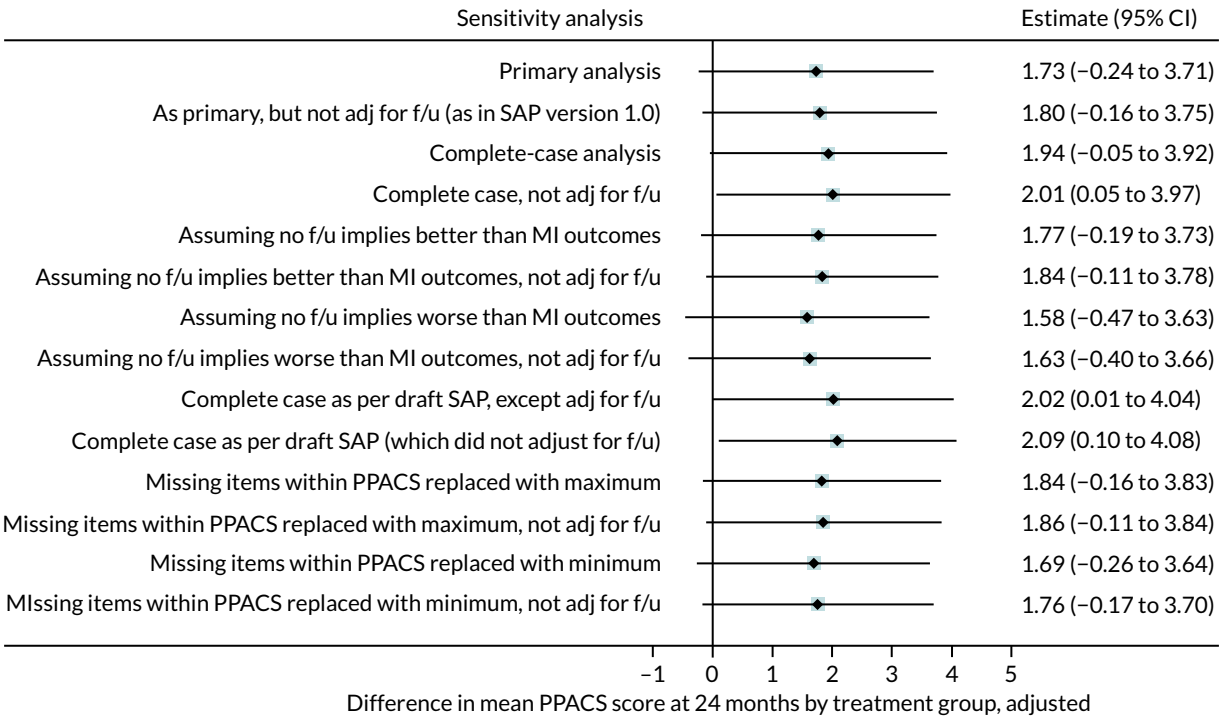


FIGURE 3 Forest plot for sensitivity analysis at the 24-month follow-up. adj, adjusted; f/u, follow-up.

The primary analyses included multiple imputation (MI) for unrateable items on the PPACS scales, MI for PPACS total scores in those who did not complete their 5-month follow-up assessment and adjustment for length of follow-up. Sensitivity analyses included lack of adjustment for length of follow-up (combined with other assumptions). Complete-case analyses are then reported (using MI solely for unrateable items on the PPACS scale and excluding those who did not complete their 5-month follow-up). Further analyses assume better PPACS scores and then worse PPACS scores (than predicted by MI by one standard deviation of the mean change in PPACS scores between the baseline and 5-month follow-up assessment) in those who did not complete their 5-month follow-up. The final analyses assume that unrateable PPACS items were replaced with the highest possible score and then the lowest possible score.

Sensitivity analyses were completed at the 24-month follow-up in the same way as in the 5-month follow-up sensitivity analyses. This included lack of adjustment for length of follow-up (combined with other assumptions). Complete-case analyses are then reported (using MI solely for unrateable items on the PPACS scale and excluding those who did not complete their 24-month follow-up). Further analyses assume better PPACS scores and then worse PPACS scores (than predicted by MI by one SD of the mean change in PPACS scores between the baseline and 24-month follow-up assessment) in those who did not complete their 24-month follow-up. The final analyses assume that unrateable PPACS items were replaced with the highest possible score and then the lowest possible score.

Secondary outcomes: parent-reported measures of child behaviour at the 5- and 24-month follow-ups

On the key secondary outcomes, parent-reported questionnaire measures of children's behaviour (total scores of the CBCL and SDQ), the results indicate a positive direction of effect favouring the VIPP-SD group at the 5-month follow-up, yet indicate little evidence of a sustained effect at the 24-month follow-up (*Table 7*). At the 5-month follow-up, the VIPP-SD group had lower CBCL scores [adjusted mean difference 3.24, 95% CI −0.06 to 6.54; $p = 0.05$ ($d = 0.15$, 95% CI 0.00 to 0.31)] and

TABLE 7 Primary ITT analysis of adjusted mean difference on secondary outcomes of children's behaviour problems (total scores) at the 5- and 24-month follow-ups

Outcome	Trial group: VIPP-SD *n*	VIPP-SD Mean (SD)	Usual care *n*	Usual care Mean (SD)	Mean difference[a] (95% CI)	Effect size[b] (95% CI)	*p*-value
CBCL[c] total score (primary caregiver report)							
Baseline	151	40.7 (21.7)	149	42.7 (21.1)			
5 months	140	32.5 (20.6)	145	37.2 (21.0)	3.24 (−0.06 to 6.54)	0.15 (0.00 to 0.31)	0.05
24 months	141	30.6 (23.4)	144	35.3 (23.7)	2.82 (−1.82 to 7.45)	0.12 (−0.08 to 0.31)	0.23
SDQ[c] total score (primary caregiver report)							
Baseline	150	13.8 (4.8)	149	14.0 (4.7)			
5 months	140	11.3 (5.1)	145	12.2 (5.2)	0.93 (−0.03 to 1.9)	0.18 (−0.01 to 0.36)	0.06
24 months	141	10.4 (5.4)	144	10.9 (5.8)	0.35 (−0.78 to 1.47)	0.06 (−0.13 to 0.25)	0.54
SDQ[c,d] total score (teacher report)							
24 months	106	7.1 (6.0)	104	7.8 (5.7)	0.54 (−1.00 to 2.08)	0.10 (−0.18 to 0.37)	0.49

a Difference in mean is the difference between treatment groups from linear regression analysis of the outcome measure on the baseline score of that same measurement, on treatment centre, on randomised group, on length of follow-up, on age of child and on number of parents/caregivers participating (all treated as fixed effects). Positive differences represent greater adjusted decreases in symptoms in the VIPP-SD group than in the usual-care group.
b Effect size is the standardised difference in mean (adjusted as above). This is the difference in mean divided by the SD of controls at follow-up.
c Higher scores represent higher levels of behaviour problems.
d Teacher-reported SDQ scores are adjusted for baseline primary caregiver-reported SDQ.

SDQ scores [adjusted mean difference 0.93, 95% CI −0.03 to 1.90; $p = 0.06$ ($d = 0.18$, 95% CI −0.01 to 0.36)] than the usual-care group. At the 24-month follow-up, although the VIPP-SD group continued to have lower behaviour scores than the usual-care group, these differences had diminished somewhat and there was little evidence of sustained effects of VIPP-SD on the CBCL (adjusted mean difference 2.82, 95% CI −1.82 to 7.45; $p = 0.23$) and SDQ (adjusted mean difference 0.35, 95% CI −0.78 to 1.47; $p = 0.54$). There was no evidence of superiority of VIPP-SD over usual care on the teacher-reported total score for the SDQ at the 24-month follow-up (difference 0.54, 95% CI −1.00 to 2.08; $p = 0.49$).

Table 8 shows the adjusted group differences on the subscale scores for the secondary behaviour outcomes (CBCL and SDQ) at the 5- and 24-month follow-ups. At the 5-month follow-up, there was evidence of an effect favouring the VIPP-SD group over the usual-care group on the SDQ externalising subscale [adjusted mean difference 0.68, 95% CI 0.00 to 1.37; $p = 0.05$ ($d = 0.18$, 95% CI 0.00 to 0.37)] as well as weaker evidence favouring the VIPP-SD group over the usual-care group on the CBCL internalising scale [adjusted mean difference 1.03, 95% CI −0.10 to 2.16; $p = 0.07$ ($d = 0.15$, 95% CI −0.01 to 0.32)]. There was no strong evidence of group differences on the remaining subscales. In general, differences on the subscales had attenuated at the 24-month follow-up and there was little evidence of sustained effects on these measures. On the teacher-reported SDQ, although teachers tend to report lower behaviour problems for the VIPP-SD group than for the usual-care group on key subscales, there was only weak evidence of group differences on these subscales: conduct [adjusted mean difference 0.38, 95% CI −0.16 to 0.92; $p = 0.17$ ($d = 0.18$, 95% CI −0.08 to 0.45)], hyperactivity [adjusted mean difference 0.46, 95% CI −0.25 to 1.16; $p = 0.20$ ($d = 0.17$, 95% CI −0.09 to 0.43)] and externalising [adjusted mean difference 0.81, 95% CI −0.29 to 1.92; $p = 0.15$ ($d = 0.19$, 95% CI −0.07 to 0.45)] subscales.

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

TABLE 8 Intention-to-treat analysis of adjusted mean difference of change in children's behaviour problems on the secondary outcome subscales (primary caregiver-reported CBCL and SDQ) at the 5- and 24-month follow-ups

Outcome	Trial group				Mean difference[a] (95% CI)	Effect size[b] (95% CI)	*p*-value
	VIPP-SD		Usual care				
	n	Mean (SD)	*n*	Mean (SD)			
CBCL[c] externalising subscale							
Baseline	151	16.5 (8.1)	149	17.5 (8.3)			
5 months	140	13.3 (8.3)	145	15.2 (8.5)	1.15 (–0.32 to 2.63)	0.14 (–0.04 to 0.31)	0.12
24 months	141	11.6 (8.8)	144	13.4 (9.6)	0.81 (–1.02 to 2.65)	0.09 (–0.11 to 0.28)	0.38
CBCL[c] internalising subscale							
Baseline	151	8.9 (7.2)	149	9.7 (7.5)			
5 months	140	7.1 (6.5)	145	8.5 (6.8)	1.03 (–0.10 to 2.16)	0.15 (–0.01 to 0.32)	0.07
24 months	141	8.1 (8.2)	144	9.0 (7.6)	0.44 (–1.16 to 2.04)	0.06 (–0.15 to 0.27)	0.59
CBCL[c] attention subscale							
Baseline	151	3.8 (2.0)	149	4.2 (2.3)			
5 months	140	2.9 (2.1)	145	3.5 (2.2)	0.25 (–0.18 to 0.68)	0.11 (–0.08 to 0.30)	0.26
24 months	141	2.2 (2.0)	144	2.7 (2.2)	0.27 (–0.18 to 0.71)	0.12 (–0.08 to 0.31)	0.24
CBCL[c] aggression subscale							
Baseline	151	12.7 (6.9)	149	13.3 (7.0)			
5 months	140	10.4 (6.8)	145	11.8 (6.9)	0.98 (–0.25 to 2.22)	0.14 (–0.04 to 0.32)	0.12
24 months	141	9.4 (7.4)	144	10.7 (7.9)	0.63 (–0.90 to 2.16)	0.08 (–0.11 to 0.27)	0.42
SDQ[d,e] externalising subscale							
Baseline	150	9.4 (3.0)	149	9.4 (3.5)			
5 months	140	7.6 (3.4)	145	8.2 (3.7)	0.68 (0.00 to 1.37)	0.18 (0.00 to 0.37)	0.05
24 months	141	7.0 (3.6)	144	7.4 (4.1)	0.39 (–0.38 to 1.16)	0.10 (–0.09 to 0.29)	0.32
24 months (teacher-report)	106	4.1 (4.2)	104	4.9 (4.3)	0.81 (–0.29 to 1.92)	0.19 (–0.07 to 0.45)	0.15
SDQ[d,e] conduct subscale							
Baseline	150	3.6 (2.0)	149	3.4 (2.1)			
5 months	140	3.1 (1.9)	145	3.1 (2.0)	0.27 (–0.14 to 0.67)	0.13 (–0.07 to 0.33)	0.20
24 months	141	2.9 (2.0)	144	3.1 (2.3)	0.34 (–0.11 to 0.79)	0.15 (–0.05 to 0.35)	0.14
24 months (teacher-report)	106	1.2 (2.0)	104	1.6 (2.1)	0.38 (–0.16 to 0.92)	0.18 (–0.08 to 0.45)	0.17
SDQ[d,e] hyperactivity subscale							
Baseline	150	5.9 (1.9)	149	6.1 (2.2)			
5 months	140	4.5 (2.1)	145	5.1 (2.3)	0.37 (–0.08 to 0.81)	0.16 (–0.03 to 0.35)	0.11
24 months	141	4.1 (2.2)	144	4.3 (2.4)	0.07 (–0.40 to 0.53)	0.03 (–0.17 to 0.22)	0.78
24 months (teacher-report)	106	2.9 (2.6)	104	3.4 (2.7)	0.46 (–0.25 to 1.16)	0.17 (–0.09 to 0.43)	0.20

TABLE 8 Intention-to-treat analysis of adjusted mean difference of change in children's behaviour problems on the secondary outcome subscales (primary caregiver-reported CBCL and SDQ) at the 5- and 24-month follow-ups (*continued*)

Outcome	Trial group				Mean difference[a] (95% CI)	Effect size[b] (95% CI)	*p*-value
	VIPP-SD		Usual care				
	n	Mean (SD)	*n*	Mean (SD)			
SDQ[d,e] emotional problems subscale							
Baseline	150	1.4 (1.5)	149	1.6 (1.5)			
5 months	140	1.4 (1.7)	145	1.4 (1.5)	0.01 (−0.32 to 0.33)	0.01 (−0.21 to 0.22)	0.96
24 months	141	1.8 (1.8)	144	1.7 (1.8)	−0.18 (−0.56 to 0.21)	−0.10 (−0.30 to 0.11)	0.37
24 months (teacher-report)	106	1.4 (1.8)	104	1.2 (1.7)	−0.21 (−0.68 to 0.27)	−0.12 (−0.40 to 0.16)	0.39
SDQ[d,e] peer problems subscale							
Baseline	150	3.0 (1.9)	149	3.0 (1.8)			
5 months	140	2.3 (1.9)	145	2.5 (2.0)	0.23 (−0.19 to 0.64)	0.12 (−0.10 to 0.33)	0.29
24 months	141	1.6 (1.6)	144	1.8 (1.9)	0.15 (−0.24 to 0.53)	0.08 (−0.12 to 0.28)	0.44
24 months (teacher-report)	106	1.6 (1.9)	104	1.6 (2.0)	−0.06 (−0.59 to 0.47)	−0.03 (−0.30 to 0.24)	0.82
SDQ[d,e] prosocial scale							
Baseline	150	5.7 (2.3)	149	5.4 (2.2)			
5 months	140	6.6 (2.0)	145	6.3 (2.2)	−0.07 (−0.47 to 0.34)	−0.03 (−0.22 to 0.16)	0.75
24 months	141	7.7 (1.8)	144	7.3 (1.9)	−0.19 (−0.59 to 0.21)	−0.10 (−0.31 to 0.11)	0.36
24 months (teacher-report)	106	7.1 (2.5)	104	6.7 (2.3)	−0.31 (−0.94 to 0.33)	−0.13 (−0.41 to 0.14)	0.34

a Difference in mean is the difference between treatment groups from linear regression analysis of the outcome measure on the baseline score of that same measurement, on treatment centre, on randomised group, on length of follow-up, on age of child and on number of parents/care-givers participating (all treated as fixed effects). Positive differences represent greater adjusted decreases in symptoms in the VIPP-SD than in the usual-care group, with the exception of the SDQ prosocial scale (see footnote d).
b Effect size is the standardised difference in mean (adjusted as above). This is the difference in mean divided by the SD of controls at follow-up.
c Higher scores on all scales represent higher levels of behaviour problems.
d Higher scores on all scales represent higher levels of behaviour problems, with the exception of the prosocial scale, where higher scores indicate lower levels of problems.
e Teacher-reported SDQ scores are adjusted for baseline primary caregiver-reported SDQ.

Table 9 shows the results of the secondary CACE analysis estimating the effects of receiving the intervention (defined as at least four VIPP-SD sessions), rather than merely being randomised to it. The CACE analysis was completed on the primary and secondary behaviour outcome measures (PPACS, CBCL and SDQ). For the PPACS scores, the results showed a larger adjusted mean difference favouring the VIPP-SD group at the 5-month follow-up when accounting for treatment adherence [difference 2.59, 95% CI 0.24 to 4.94; $p = 0.03$ ($d = 0.26$, 95% CI 0.02 to 0.50)] than in the ITT analysis [difference 2.03, 95% CI 0.06 to 4.01; $p = 0.04$ ($d = 0.20$, 95% CI 0.01 to 0.40)]. At the 24-month follow-up, the difference remains in the same range, similar to the outcome in the ITT analysis, although the adjusted group difference and effect size tends to be larger in the CACE analysis [difference 1.96, 95% CI −0.30 to 4.23; $p = 0.09$ ($d = 0.20$, 95% CI −0.03 to 0.43)]. A similar pattern is observed on the CBCL [adjusted mean difference 3.56, 95% CI 0.04 to 7.09; $p = 0.05$ ($d = 0.17$, 95% CI 0.00 to 0.34)] and SDQ [adjusted mean difference 1.03, 95% CI −0.01 to 2.06; $p = 0.05$ ($d = 0.20$, 95% CI 0.00 to 0.39)] at the 5-month follow-up, such that both outcomes show slightly larger adjusted

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

TABLE 9 Complier-average causal effect analysis[a] on primary and secondary outcomes of child behaviour

Outcome	Trial group				Mean difference[b] (95% CI)	Effect size[c] (95% CI)	*p*-value
	VIPP-SD		Usual care				
	n	Mean (SD)	*n*	Mean (SD)			
PPACS[d]							
Baseline	151	33.5 (9.0)	149	32.4 (10.6)			
5 months	140	28.8 (9.2)	146	30.3 (9.9)	2.59 (0.24 to 4.94)	0.26 (0.02 to 0.5)	0.03
24 months	140	23.1 (8.2)	142	24.7 (9.9)	1.96 (–0.30 to 4.23)	0.20 (–0.03 to 0.43)	0.09
CBCL[e]							
Baseline	151	40.7 (21.7)	149	42.7 (21.1)			
5 months	140	32.5 (20.6)	145	37.2 (21.0)	3.56 (0.04 to 7.09)	0.17 (0.00 to 0.34)	0.05
24 months	141	30.6 (23.4)	144	35.3 (23.7)	2.90 (–2.19 to 7.98)	0.12 (–0.09 to 0.34)	0.26
SDQ[e] ***– primary caregiver reported***							
Baseline	150	13.8 (4.8)	149	14.0 (4.7)			
5 months	140	11.3 (5.1)	145	12.2 (5.2)	1.03 (–0.01 to 2.06)	0.20 (0.00 to 0.39)	0.05
24 months	141	10.4 (5.4)	144	10.9 (5.8)	0.32 (–0.91 to 1.56)	0.06 (–0.16 to 0.27)	0.61
SDQ[e] ***– teacher report***							
24 months	106	7.1 (6)	104	7.8 (5.7)	0.61 (–1.06 to 2.27)	0.11 (–0.19 to 0.40)	0.48

a The CACE results are based on the assumption that there is no effect of being randomised to VIPP-SD, or of receiving one, two or three VIPP-SD visits. Receiving four, five or six VIPP-SD visits counts as receiving the intervention.
b Difference in mean is the difference between treatment groups from linear regression analysis of the outcome measure on the baseline score of that same measurement, on treatment centre, on randomised group, on length of follow-up, on age of child and on number of parents/caregivers participating (all treated as fixed effects). Positive differences represent greater adjusted decreases in symptoms in the VIPP-SD group than in the usual-care group.
c Effect size is the standardised difference in mean (adjusted as above). This is the difference in mean divided by the SD of controls at follow-up.
d PPACS results include multiple imputation for those with missing data at follow-up, and for unratable items in the PPACS scales.
e Higher scores represent higher levels of behaviour problems.

mean differences and effect sizes, and stronger evidence of group differences in the CACE analysis than in the ITT analysis. As with the ITT analysis, there is little evidence of a sustained group difference on the secondary behaviour outcomes (CBCL and SDQ) at the 24-month follow-up in the CACE analysis. *Tables 25–27* in *Appendix 4* show the analysis of the secondary behaviour outcomes as reported by secondary participating caregivers.

Table 10 shows the ITT analysis for secondary outcomes relating to participating caregivers' reported parenting practices, mood, anxiety and couple functioning, disaggregated by sex of the reporting caregiver. There was no evidence of group differences on these secondary outcomes (*p*-values = 0.15–0.95).

Figures 4 and *5* show exploratory subgroup analyses that disaggregate the adjusted mean difference between VIPP-SD and usual care on the basis of the child's age at baseline, the number of participating caregivers, the severity of baseline SDQ score (by quartile) and the fidelity of the therapist. At the 5-month follow-up, the analysis suggested a greater positive response to treatment in children aged 12–23 months (adjusted mean difference 2.91, 95% CI 0.06 to 5.76), in families with one participating caregiver (adjusted mean difference 2.79, 95% CI 0.63 to 4.94), those with higher baseline behaviour

TABLE 10 Intention-to-treat analysis of secondary outcomes on participating caregivers' parenting practices, mood, anxiety, and couple functioning, by sex of caregiver

Outcome	Trial group				Mean difference[a] (95% CI)	Effect size[b] (95% CI)	*p*-value
	VIPP-SD		Usual care				
	n	Mean (SD)	*n*	Mean (SD)			
Female caregivers[c] *Parenting Scale*[d]							
Baseline	146	2.96 (0.52)	147	2.95 (0.58)			
5 months	135	2.90 (0.50)	143	2.90 (0.60)	0.06 (−0.22 to 1.23)	0.11 (−0.37 to 2.06)	0.22
24 months	136	3.02 (0.53)	142	3.02 (0.57)	0.03 (−0.07 to 0.13)	0.06 (−0.12 to 0.23)	0.51
PHQ-9[e]							
Baseline	145	4.34 (4.00)	147	4.28 (4.35)			
5 months	135	3.99 (4.49)	144	4.20 (4.71)	0.25 (−0.69 to 1.20)	0.05 (−0.15 to 0.25)	0.60
24 months	136	3.99 (4.60)	141	4.02 (4.22)	0.05 (−0.86 to 0.97)	0.01 (−0.20 to 0.23)	0.91
GAD-7[e]							
Baseline	145	4.89 (4.33)	147	4.73 (4.22)			
5 months	134	4.29 (4.46)	144	3.92 (4.00)	0.05 (−0.85 to 0.95)	0.01 (−0.21 to 0.24)	0.91
24 months	136	4.12 (4.64)	141	4.20 (4.09)	0.16 (−0.76 to 1.07)	0.04 (−0.19 to 0.26)	0.74
RDAS[f]							
Baseline	130	49.18 (8.36)	126	50.50 (9.22)			
5 months	118	49.19 (9.32)	120	49.92 (9.59)	0.20 (−1.44 to 1.83)	0.02 (−0.15 to 0.19)	0.81
24 months	120	50.01 (8.15)	115	50.59 (6.96)	−0.40 (−1.78 to 0.99)	−0.06 (−0.26 to 0.14)	0.57
Male caregivers[g] *Parenting Scale*[d]							
Baseline	31	2.98 (0.54)	25	2.78 (0.48)			
5 months	29	2.90 (0.50)	24	2.89 (0.41)	0.10 (−0.13 to 0.33)	0.24 (−0.32 to 0.8)	0.39
24 months	28	2.9 (0.49)	24	2.95 (0.45)	0.15 (−0.06 to 0.36)	0.34 (−0.13 to 0.82)	0.15
PHQ-9[e]							
Baseline	31	3.03 (2.64)	25	2.56 (2.96)			
5 months	29	2.48 (2.34)	24	3.25 (3.57)	0.63 (−0.86 to 2.12)	0.18 (−0.24 to 0.6)	0.40
24 months	28	2.64 (2.15)	24	2.54 (3.08)	0.11 (−1.21 to 1.42)	0.03 (−0.39 to 0.46)	0.87
GAD-7[e]							
Baseline	31	2.35 (2.69)	25	2.84 (2.98)			
5 months	29	2.21 (2.32)	24	3.21 (3.66)	−0.04 (−1.54 to 1.45)	−0.01 (−0.42 to 0.40)	0.95
24 months	28	2.54 (2.32)	24	2.50 (2.70)	−0.51 (−1.59 to 0.57)	−0.19 (−0.59 to 0.21)	0.35

continued

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

TABLE 10 Intention-to-treat analysis of secondary outcomes on participating caregivers' parenting practices, mood, anxiety, and couple functioning, by sex of caregiver (*continued*)

Outcome	Trial group: VIPP-SD *n*	VIPP-SD Mean (SD)	Usual care *n*	Usual care Mean (SD)	Mean difference[a] (95% CI)	Effect size[b] (95% CI)	*p*-value
RDAS[f]							
Baseline	31	50.56 (5.20)	25	50.23 (6.51)			
5 months	29	50.38 (6.42)	24	48.31 (8.80)	0.11 (−3.42 to 3.65)	0.01 (−0.39 to 0.41)	0.95
24 months	27	51.37 (4.87)	24	48.82 (6.78)	−1.27 (−3.97 to 1.42)	−0.19 (−0.59 to 0.21)	0.35

a Difference in mean is the difference between treatment groups from linear regression analysis of the outcome measure on the baseline score of that same measurement, on treatment centre, on randomised group, on length of follow-up, on age of child and on number of caregivers participating (all treated as fixed effects). Positive differences represent greater adjusted decreases in symptoms in the VIPP-SD than in the usual-care group in all cases but the RDAS (see footnote g).
b Effect size is the standardised difference in mean (adjusted as above). This is the difference in mean divided by the SD of controls at follow-up.
c Female caregivers consisted of 287 primary caregivers at baseline, 271 at the 5-month follow-up, and 272 at the 24-month follow-up, as well as eight participating female secondary caregivers at each time point. Two families had both a female primary and female secondary caregiver participating. In these cases, the scores from these caregivers were averaged prior to inclusion in the above analysis.
d Higher scores indicate more ineffective parenting strategies.
e Higher scores represent higher symptom severity.
f Higher scores indicate greater relationship satisfaction and lower scores indicate greater relationship distress.
g Male caregivers consisted of 13 primary caregivers at baseline and the 5- and 24-month follow-ups; and 43 participating male secondary caregivers at baseline, 40 at 5-month follow-up, and 39 at 24-month follow-up.

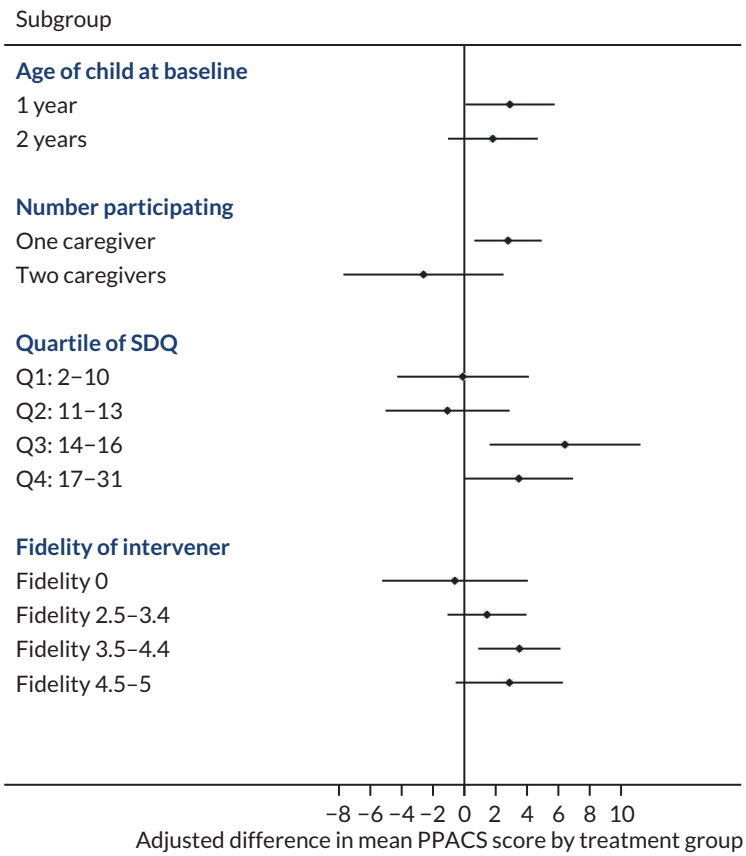


FIGURE 4 Forest plot for exploratory subgroup analyses at the 5-month follow-up.

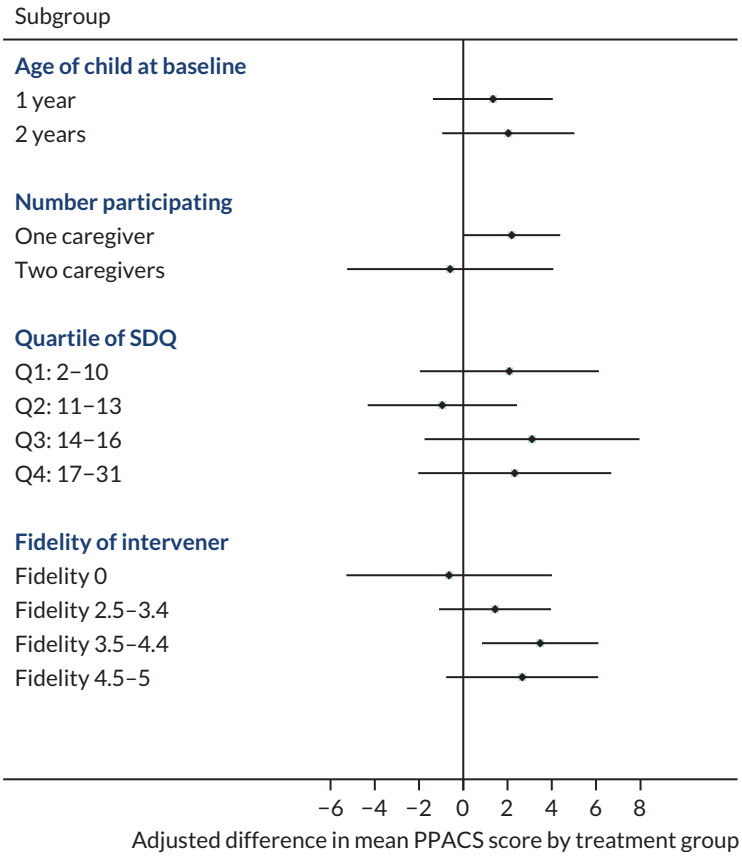


FIGURE 5 Forest plot for exploratory subgroup analyses at the 24-month follow-up.

scores (as indicated by total scores in the third quartile: adjusted mean difference 6.42, 95% CI 1.61 to 11.23; and fourth quartile: adjusted mean difference 3.48, 95% CI 0.03 to 6.93; of the baseline SDQ), and those who had received the programme from a therapist rated high (adjusted mean difference 3.47, 95% CI 0.83 to 6.12), but not the highest, on fidelity. At the 24-month follow-up, larger effects remained in the families with one participating caregiver (adjusted mean difference 2.19, 95% CI 0.00 to 4.39) and those who had received the programme from a therapist rated with high (adjusted mean difference 3.51, 95% CI 0.88 to 6.14), but not the highest, fidelity scores. *Appendix 6* shows the subgroup analysis for baseline behaviour based on quartiles of externalising scores on the SDQ at the 5- and 24-month follow-ups (see *Appendix 6*, *Figures 13* and *14*).

The age of a child at baseline and the number of participating caregivers were predefined subgroup analyses. Analyses by quartile of SDQ (at baseline) and (in those randomised to VIPP-SD) by fidelity of the intervention (compared with those randomised to usual care) are post hoc. Fidelity is independently assessed, based on the fidelity of the therapist; 'fidelity zero' represents a family who completed fewer than four VIPP-SD visits, despite being randomised to VIPP-SD.

Missing data

Eleven participants (7%) allocated to the VIPP-SD group and three participants (2%) allocated to the usual-care group were not assessed for the primary outcome measure (PPACS) at the 5-month follow-up. Eleven participants (7%) allocated to the VIPP-SD group and seven participants (5%) allocated to the usual-care group were not assessed for the primary outcome measure (PPACS) at the 24-month follow-up. Baseline characteristics of participants with and without missing primary outcome data are given in *Appendix 5*, *Table 29*.

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

Blinding

Outcome assessors reported having been unblinded for 11 participants (4%) [seven (5%) in the VIPP-SD group and four (3%) in the usual-care group] as a result of participants informing assessors of their treatment allocation at their 5-month follow-up visit. At the 24-month follow-up visit, five participants (2%) [three in the VIPP-SD group and two in the usual-care group] informed assessors of their allocation. In instances where unblinding occurred before the primary outcome had been collected, the audio-recording of the interview was double-scored by a second assessor, who remained blinded to allocation, to minimise the risk of bias. Assessors who were unblinded at the 5-month follow-up assessment did not collect primary outcome data from the same family at the 24-month follow-up visit.

Acceptability

Participants allocated to the VIPP-SD group were invited to complete a feedback questionnaire at the end of treatment. Of the 139 participants who received at least one therapy session, 39 (28%) returned the feedback questionnaire. Seven families included two participating caregivers who each completed a questionnaire, giving a total of 46 respondents. *Table 11* shows that approximately two-thirds to three-quarters of respondents endorsed the programme's positive impact on their relationship with their child; their understanding of their child's thoughts, feelings, behaviours, and likes and dislikes; as well as their reactions to their child's behaviour and communication. All respondents reported that their preferred setting for the programme was in the home. The main reasons endorsed by participants were that the home setting removed the need for travel and reflected everyday interactions (*Table 12*).

TABLE 11 Participants' responses to the intervention feedback questionnaire: perceived positive impact of VIPP-SD

How much do you think the visits have impacted on your . . .	Not at all/a little, *n* (%)	Moderately, *n* (%)	A lot/a great deal, *n* (%)
Relationship with your child	13 (28)	17 (37)	16 (35)
Understanding of your child's thoughts and feelings	9 (20)	16 (35)	21 (46)
Understanding of your child's behaviour	9 (20)	17 (37)	20 (43)
Reaction to your child's behaviour	12 (26)	12 (26)	22 (48)
Understanding of your child's likes and dislikes	17 (37)	17 (37)	12 (26)
Communication with your child	12 (26)	17 (37)	17 (37)

Note
Caregivers ($N = 46$) were from 39 participating families; of these, 32 were from families with one participating caregiver and 14 were from families with two participating caregivers.

TABLE 12 Participants' responses to the intervention feedback questionnaire: perceptions of treatment setting

I found that having the sessions delivered in my home was helpful because . . .	*n* (%)
I did not have to pay for transport	18 (39)
I did not have to arrange childcare	25 (54)
I did not spend time travelling	39 (85)
It reflected everyday interaction	37 (80)
I have not found the home-based delivery format helpful	0 (0)

Note
Caregivers ($N = 46$) were from 39 participating families; of these, 32 were from families with one participating caregiver and 14 were from families with two participating caregivers.

Serious adverse events

During the study, there were three children (2%) admitted to hospital in the VIPP-SD group by the 5-month follow-up and 10 children (7%) were admitted by the 24-month follow-up. In the usual-care group, there were four children (2%) admitted to hospital by the 5-month follow-up and eight children (5%) by the 24-month follow-up. All admissions reasons were unrelated to treatment and study procedures (e.g. accidents and respiratory infections). There was no evidence of group differences and no other adverse events were reported. Protocol deviations and violations are detailed in *Appendix 7*.

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

Chapter 4 Economic evaluation

Aim

The aim of the economic evaluation was to assess the cost-effectiveness of VIPP-SD compared with usual care at 24 months post randomisation.

Methods

Perspective

The economic evaluation took the NHS/Personal Social Services (PSS) perspective preferred by NICE,[65] and included the use of all health and social services by the child and by parent/caregivers in relation to their child's needs over the 24-month follow-up.

Method of economic evaluation

Short-term cost-effectiveness

The primary economic evaluation explored cost-effectiveness at the 24-month follow-up in terms of the primary clinical outcome measure (PPACS). In addition, a cost–consequences analysis outlining costs alongside the key secondary outcome measures (CBCL and SDQ) explored the potential economic impacts of the intervention more broadly.

Long-term cost-effectiveness

A longer-term analysis aimed to explore the cost–utility of VIPP-SD beyond the trial follow-up period by utilising data from the trial and supplementing it with available data from the literature using decision-analytic modelling techniques.

Costs

Information on the use of health and social care services was recorded in interviews using a modified version of the CA-SUS. Modifications to the CA-SUS were based on a review of recent literature and clinical feedback. Data were collected on the use of accommodation services (e.g. foster care and supported housing), hospital services (e.g. inpatient stays, outpatient contacts, accident and emergency attendances), community-based health and social care services (e.g. contacts with GPs and clinical psychologists), and all prescribed medication. Parents/caregivers were interviewed at baseline and asked about their child's use of services and/or their own use of services relating to their child's needs in the previous 3 months. At subsequent follow-up points (5 and 24 months), service use was recorded for the period since the previous interview to ensure that the entire duration of the trial had been captured.

For each item of service use reported in the CA-SUS, a nationally applicable unit cost was applied to calculate the total costs for each participant. Unit costs for hospital services were sourced from the *NHS Reference Costs 2017–18*.[66] Costs from the annually published unit costs of health and social care compendium[67] were applied to community-based health and social care, and Local Authority accommodation services. Costs applied to accommodation services were also sourced from UK government web pages.[68,69] The costs of medications were based on prices listed in the *British National Formulary*[70] for the generic drug, and were calculated based on dosages reported in the CA-SUS, or national averages for young children if left unspecified.[70] Unit costs applied were for the financial year 2017–18 (*Table 13*) and are reported in Great British pounds. Costs incurred after 12 months from the start of the trial were discounted by 3.5%, as recommended by NICE.[65]

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

TABLE 13 Unit costs and sources used for the economic evaluation

Service	Unit cost or range (£)	Source	Notes
Hospital inpatient (per night)	484.89–643.80	*NHS Reference Costs 2017–18*[66]	Weighted average of short and long stay
Hospital outpatient (per appointment)	48.81–751.12	*NHS Reference Costs 2017–18*[66]	Varied by specialty reported in the CA-SUS; unit cost for paediatrics (£198.20) applied if reason for appointment unknown, missing or reported as 'other'
Hospital accident and emergency (per attendance)	148.36–247.50	*NHS Reference Costs 2017–18*[66]	Varied by if ambulance services were used
Community-based health care, e.g. GP, district nurse (per minute of contact)	0.60–3.10	Curtis 2017[67]	Varied by specialist seen and duration of contact reported in the CA-SUS
Community-based social care, e.g. social services youth worker (per minute of contact)	0.52–0.98	Curtis 2017[67]	Varied by specialist seen and duration of contact reported in the CA-SUS
Accommodation (per day)	18.12–92.29	Curtis 2017[67] Gov.uk: foster carers[68] DirectGov: local housing allowance rates[69]	
Medication (per day)	0.02–1.37	*British National Formulary*[70]	Varied by medication type and dosage reported in the CA-SUS

The cost of delivering VIPP-SD was calculated using a standard micro-costing (bottom-up) approach (outlined for a band 6 NHS therapist in *Table 14*) and included therapist salaries plus on-costs (employers' national insurance and superannuation contributions) and appropriate capital, administrative and managerial overheads, as well as the costs of training, supervision and equipment.[65,71] Data on intervention contacts and the costs of training and materials were collected directly from trial records, and indirect time (preparation, supervision, administration, travel, etc.) was estimated using questionnaires completed by each therapist delivering the intervention on the time they spent on different activities. Intervention costs were calculated individually for each trial participant and depended on the therapist they saw and the number of VIPP-SD sessions they received.

Data analysis

Short-term analyses

The use of health and social care services is reported by trial group as the mean, SD and range, as well as by the percentage of the sample for each group that had at least one contact. Differences in resource use were not tested for statistical significance to avoid excessive significance testing and to keep the focus of the economic analysis on cost and cost-effectiveness.

For each participant, all costs were summed to calculate total costs over the 24-month follow-up period. Costs and outcomes, including costs per sector, were summarised using the mean and standard error for each trial group, and the differences between the two were compared using standard parametric *t*-tests. Despite the skewed nature of cost data, this method allows inferences to be made about the arithmetic mean, which is more meaningful from a decision-making perspective.[72]

Cost-effectiveness of VIPP-SD was explored in terms of the primary outcome measure (PPACS) and assessed (1) through the calculation of incremental cost-effectiveness ratios (ICERs) (i.e. the additional

TABLE 14 Costing schema for calculation of intervention costs

Cost and unit estimation	2017–18 value	Notes
Wages/salary[a]	£31,593	Based on a health visitor (pay band 6) delivering the intervention
Salary oncosts[a]	£7791	Includes employer's contributions to national insurance and superannuation
Overheads[a]	£24,694	Includes costs for management, administration and estates staff, plus non-staff costs
Capital overheads[a]	£5125	
Working time[a]	1599 hours per year	Based on 42.6 weeks per year, 37.5 hours per week
Face-to-face time[b]	1 : 2.25	Based on each hour of intervention delivery requiring 2.25 hours of preparation and administration time
VIPP-SD delivery[b]	£140.65 per hour	
Supervision[b]	£21.80 per hour of delivery	Based on each hour of intervention delivery requiring 0.25 hours of supervision by a senior clinician (pay band 8c)
Training and equipment[c]	£12.16 per hour of delivery	Includes costs of training sessions, manuals and therapist kits [with toys, camera, digital voice recorder; costs were annuitised over 3 years. Costs for data storage and file-sharing platform were not included
Length of sessions[c]	1.5 hours	Cost per session per participant: £261.92

a Sourced from the annual unit costs of health and social care publication.[67]
b Estimated from questionnaires completed by therapists delivering the intervention and clinicians supervising delivery for the trial.
c Estimated from trial administration data.

cost of one intervention compared with another, divided by the additional effect[73]) and (2) using the net monetary benefit (NMB) approach based on the use of a linear function of cost and effects and assuming specified values for the willingness-to-pay (WTP) for each additional unit of effect.[74] Uncertainty around the mean estimates of cost and outcome was explored by generating 1000 bootstrap iterations and plotting these on a cost-effectiveness plane for interpretation.[75] A cost-effectiveness acceptability curve (CEAC)[76] was then constructed to examine the probability of VIPP-SD being cost-effective compared with usual care for a range of possible values of WTP per unit improvement in outcome. This analysis was based on 30 multiply imputed data sets of total costs and outcomes using chained equations and predictive mean matching.

A prespecified sensitivity analysis was carried out to explore cost-effectiveness in terms of PPACS score at the 24-month follow-up for complete cases and excluding influential outliers (participants with total costs in the 99th percentile[77]) to examine the sensitivity of the cost-effectiveness results to missing data and MI. A second sensitivity analysis explored cost-effectiveness using the PPACS score at the 5-month follow-up, in line with the primary clinical end point.

All economic analyses were adjusted for recruitment centre, age of child at recruitment, parental involvement (one or two participating parents/caregivers) and time to follow-up, in line with clinical analyses, plus the baseline variable of interest (cost, PPACS, CBCL, SDQ). Analyses were conducted in Stata/SE® version 15.

Long-term analyses

The economic implications of behavioural problems are long term in nature, with childhood behaviour problems being linked to later delinquency and criminality, and affecting future mental health status and education and employment outcomes.[78,79] Should the trial results suggest that VIPP-SD has an impact on outcomes at the 24-month follow-up, which is suggestive of possible future cost-effectiveness differences

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

between the two groups, longer-term cost-effectiveness will be explored using decision-analytic modelling, following methods applied in similar research.[16] As no method of direct estimation of health-related quality of life, and thus quality-adjusted life-years (QALYs), preferred for health economic evaluation,[65] currently exists for infants and preschool children, effectiveness differences at the 24-month follow-up were checked for longer-term implications using the SDQ, for which a mapping function to generate QALYs from the Child Health Utility index 9D (CHU-9D)[80] exists.[81] The five SDQ subscale scores (i.e. emotion, conduct, hyperactivity, peer and prosocial) were transformed using the published algorithm to give utility weights that were used to calculate QALYs using the area under the curve approach.[82]

Results

Data completeness

The availability of resource use data at each assessment time point is summarised in *Table 15*. Data for all periods (baseline and 5- and 24-month follow-ups) were available for 140 participants (93%) in the VIPP-SD group and 142 participants (95%) in the usual-care group.

Resource use over the 24-month follow-up

Intervention

A summary of the VIPP-SD intervention sessions completed by the trial participants is presented in *Table 16*. The majority of participants in the VIPP-SD trial group (82.86%) completed all six intervention sessions, whereas very few (5.71%) completed no sessions.

Hospital, community-based and accommodation services

Table 17 summarises the use of health and social care services over the 24-month follow-up period. Overall, all services were accessed by similar proportions of participants across the trial groups. For hospital services, mean inpatient stay was higher in the VIPP-SD group (0.78 nights, SD 6.03) than in the usual-care group (0.18 nights, SD 0.71) and was used by a slightly higher proportion of participants (14% vs. 11%). For community-based services, some were used more by VIPP-SD participants than usual care participants, including GP telephone calls (46% vs. 30%) and practice nurses (76% vs. 69%), and some were used less by VIPP-SD participants, including district nurses (39% vs. 55%) and speech and language therapists (12% vs. 23%). However, few differences were evident overall. Very few accommodation services were used and these were used by a very small proportion of the sample (1%) across the trial groups.

Medication

The use of prescribed medication is summarised in *Table 18*. Antibiotics, corticosteroids and bronchodilators were the most commonly used medications across trial groups, whereas some medication types (i.e. antiepileptic, antihypertensive and antiseptic) were used in the VIPP-SD group only and were used by a very small proportion of the sample (1% for each).

TABLE 15 Availability of resource use data

	Trial group, *n* (%)	
Assessment period	VIPP-SD	Usual care
Baseline	151 (100)	149 (100)
5 months	147 (97)	147 (99)
24 months	140 (93)	142 (95)
All periods	140 (93)	142 (95)

TABLE 16 Uptake of VIPP-SD intervention sessions by participants randomised to the intervention

Number of sessions	Attendance of participants randomised to VIPP-SD, *n* (%)[a]
0	8 (5.71)
1	5 (3.57)
2	2 (1.43)
3	1 (0.71)
4[b]	4 (2.86)
5[b]	4 (2.86)
6[b]	116 (82.86)
Total	140 (100)

a This *n* relates to the estimation sample used for the primary economic analysis (i.e. those participants with 24-month follow-up data).
b Completing four or more sessions of VIPP-SD was considered as being compliant with the intervention.

TABLE 17 Use of health and social care services over 24 months' follow-up

	Trial group					
	VIPP-SD (*n* = 140)			Usual care (*n* = 142)		
Service	Mean (SD)	Range	Percentage using	Mean (SD)	Range	Percentage using
Hospital						
Inpatient (nights)	0.78 (6.03)	0–70	14	0.18 (0.71)	0–5	11
Outpatient (appointments)	1.11 (2.61)	0–16	36	1.22 (2.43)	0–17	38
A&E (attendances)	1.19 (2.71)	0–28	51	0.92 (1.28)	0–6	49
Ambulance (attendances)	0.20 (0.95)	0–10	11	0.13 (0.43)	0–3	11
Community health and social care (contacts)						
GP home	0.01 (0.08)	0–1	1	0.01 (0.08)	0–1	1
GP surgery	3.99 (4.91)	0–50	86	4.06 (4.10)	0–25	85
GP telephone	1.21 (2.11)	0–10	46	0.98 (2.28)	0–15	30
Practice nurse	1.16 (1.01)	0–5	76	1.06 (0.97)	0–4	69
District nurse	0.99 (2.14)	0–18	39	1.14 (2.62)	0–25	55
Community paediatrician	0.06 (0.30)	0–2	5	0.09 (0.37)	0–2	6
Clinical psychologist	0.09 (0.47)	0–3	4	0.30 (2.64)	0–31	4
Speech and language therapist	0.70 (2.67)	0–23	12	1.44 (7.57)	0–87	23
Child psychiatrist	0.01 (0.08)	0–1	1	0.01 (0.08)	0–1	1
Parent training	1.04 (3.57)	0–21	12	1.34 (4.26)	0–30	15
Parenting group	0.04 (0.36)	0–3	1	0.06 (0.67)	0–8	1
Family therapist	0.14 (1.39)	0–16	1	0.16 (1.31)	0–12	2
CAMHS	0.24 (1.81)	0–20	4	0.42 (2.41)	0–20	4

continued

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

TABLE 17 Use of health and social care services over 24 months' follow-up (*continued*)

Service	Trial group					
	VIPP-SD ($n = 140$)			Usual care ($n = 142$)		
	Mean (SD)	Range	Percentage using	Mean (SD)	Range	Percentage using
Occupational therapist	0.19 (1.70)	0–20	4	0.76 (6.93)	0–81	4
Physiotherapist	0.07 (0.49)	0–4	3	0.08 (0.40)	0–3	4
Nutritionist	0.09 (0.51)	0–5	5	0.15 (0.70)	0–6	7
Genetic testing	0.01 (0.12)	0–1	1	0.01 (0.08)	0–1	1
Art/music therapist	0.29 (2.07)	0–20	2	–	–	–
Dentist	1.98 (1.47)	0–8	82	2.23 (1.65)	0–8	85
Social worker	0.14 (0.78)	0–6	4	0.33 (1.94)	0–20	6
Family support worker	0.44 (1.99)	0–16	6	0.60 (2.98)	0–25	7
Portage worker	–	–	–	0.06 (0.67)	0–8	1
Accommodation key worker	–	–	–	0.25 (2.69)	0–32	3
Social services youth worker	–	–	–	–	–	–
Accommodation (nights)						
Bed and breakfast	0.45 (5.32)	0–63	1	–	–	–
Supported housing	–	–	–	0.98 (11.66)	0–139	1
Foster care	–	–	–	–	–	–
Kinship care	–	–	–	–	–	–
Residential care	–	–	–	–	–	–
Mother and baby care	–	–	–	–	–	–
Refuge	–	–	–	–	–	–

A&E, accident and emergency; CAMHS, Child and Adolescent Mental Health Services.

TABLE 18 Use of medication over 24 months' follow-up

Medication type	Trial group, percentage using	
	VIPP-SD ($n = 140$)	Usual care ($n = 142$)
Antibiotic	41	44
Corticosteroid	29	32
Bronchodilator	19	13
Vitamin	10	13
Antihistamine	6	7
Analgesic	4	5
Antiepileptic	1	–
Antihypertensive	1	–
Antiseptic	1	–
Hormone	1	1

Costs

The total costs at the 5- and 24-month follow-up points, as well as the costs per sector and the costs at baseline, are summarised in *Table 19*. At the 24-month follow-up (primary economic end point), the mean total costs were significantly higher in the VIPP-SD group (£3131.93; SD £416.54) than in the usual-care group (£1525.38; SD £293.49). The difference in total costs between the trial groups was £1449.99 (adjusted mean difference, 95% CI £619.45 to £2280.52), which was driven mainly by the cost of the intervention (mean £1466.28, standard error £46.47). Results were similar at the 5-month follow-up (primary clinical end point), being significantly higher in the VIPP-SD group than in the usual-care group (adjusted mean difference £1281.34, 95% CI £1022.40 to £1540.29).

Outcomes

Table 20 summarises the PPACS score outcomes at baseline, and at the 5- and 24-month follow-ups used in the cost-effectiveness analysis. The PPACS scores were poorer in the VIPP-SD group than in the usual-care group at baseline (mean 33.61 for VIPP-SD vs. mean 32.27 for usual care), but improved and were better in the VIPP-SD group at 5 months (mean 28.79 for VIPP-SD vs. mean 30.28 for usual care) and 24 months (mean 23.12 for VIPP-SD vs. mean 24.68 for usual care). As outlined in the clinical results, these differences were statistically significant at 5 months (adjusted mean difference −2.55, 95% CI −4.70 to −0.39) and similar but not significant at 24 months (adjusted mean

TABLE 19 Mean costs (£) per participant over 24 months' follow-up

	Trial group, mean (SE)		VIPP-SD minus usual care[a]			
Cost category	VIPP-SD	Usual care	Unadjusted difference	Adjusted difference[b]	95% CI	*p*-value
Baseline	**n = *147***	**n = *147***				
Total	297.77 (102.56)	385.97 (72.52)	−88.87	−61.47	−276.26 to 153.33	0.574
5-month follow-up	**n = *147***	**n = *147***				
Intervention	1458.51 (46.59)	0.00 (0.00)	1423.85	1423.65	1320.09 to 1527.22	< 0.001
Hospital services	143.14 (40.35)	151.16 (28.53)	−7.14	−13.84	−101.37 to 73.69	0.759
Community-based services	221.34 (87.95)	315.63 (62.19)	−95.20	−133.65	−319.59 to 52.29	0.158
Accommodation services	0.00 (0.00)	36.92 (26.10)	−37.72	−16.86	−94.22 to 60.49	0.668
Medication	4.56 (1.63)	6.47 (1.15)	−1.90	−2.64	−6.20 to 0.93	0.147
Total	1827.55 (120.71)	510.18 (85.35)	1281.89	1281.34	1022.40 to 1540.29	< 0.001
24-month follow-up	**n = *140***	**n = *142***				
Intervention	1466.28 (46.47)	0.00 (0.00)	1423.85	1430.17	1335.55 to 1524.79	< 0.001
Hospital services	841.27 (360.78)	439.91 (254.20)	397.58	312.87	−396.63 to 1022.36	0.386
Community-based services	755.35 (194.11)	1020.52 (136.77)	−284.77	−317.21	−685.63 to 51.22	0.091
Accommodation services	32.17 (50.02)	38.22 (35.24)	−10.37	1.90	−110.25 to 114.06	0.973
Medication	36.85 (8.41)	26.74 (5.93)	9.75	10.73	−6.48 to 27.94	0.221
Total	3131.93 (416.54)	1525.38 (293.49)	1536.05	1449.99	619.45 to 2280.52	0.001

SE, standard error.
a Comparison based on multiply imputed data.
b Adjusted for recruitment centre, age of child at recruitment, number of caregivers participating, time to follow-up and baseline score.

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

TABLE 20 Mean PPACS scores per participant applied to the cost-effectiveness analysis

	Trial group							
	VIPP-SD		Usual care		VIPP-SD minus usual care[b]			
PPACS[a] time point	*n*	Mean (SE)	*n*	Mean (SE)	Unadjusted difference	Adjusted difference[c]	95% CI	*p*-value
Baseline	151	33.61 (1.15)	149	32.27 (0.81)	1.25	0.98	−1.58 to 3.55	0.450
5 months	140	28.79 (1.14)	146	30.28 (0.80)	−1.62	−2.55	−4.70 to −0.39	0.021
24 months	140	23.12 (1.08)	142	24.68 (0.76)	−1.48	−1.87	−4.10 to 0.37	0.102

SE, standard error.
a Lower scores indicate better outcomes.
b Comparison based on multiply imputed data.
c Adjusted for recruitment centre, age of child at recruitment, number of caregivers participating, time to follow-up and baseline score.

difference −1.87, 95% CI −4.10 to 0.37). Results from the clinical analysis of PPACS scores (see *Table 6*) are very similar to the economic analysis, which used a separate imputation protocol combining both outcomes and costs.

Cost-effectiveness analysis using Preschool Parental Account of Children's Symptoms at 24-month follow-up

The results of the primary cost-effectiveness analysis using PPACS as the outcome of interest suggest that VIPP-SD is more costly and more effective than usual care. *Figure 6* shows the scatterplot of bootstrapped mean differences in costs and PPACS score. The majority of scatter points (98%) lie in the north-east quadrant, where VIPP-SD is more costly and more effective, whereas the remaining scatter points (2%) lie in the north-west quadrant, where VIPP-SD is more costly and less effective.

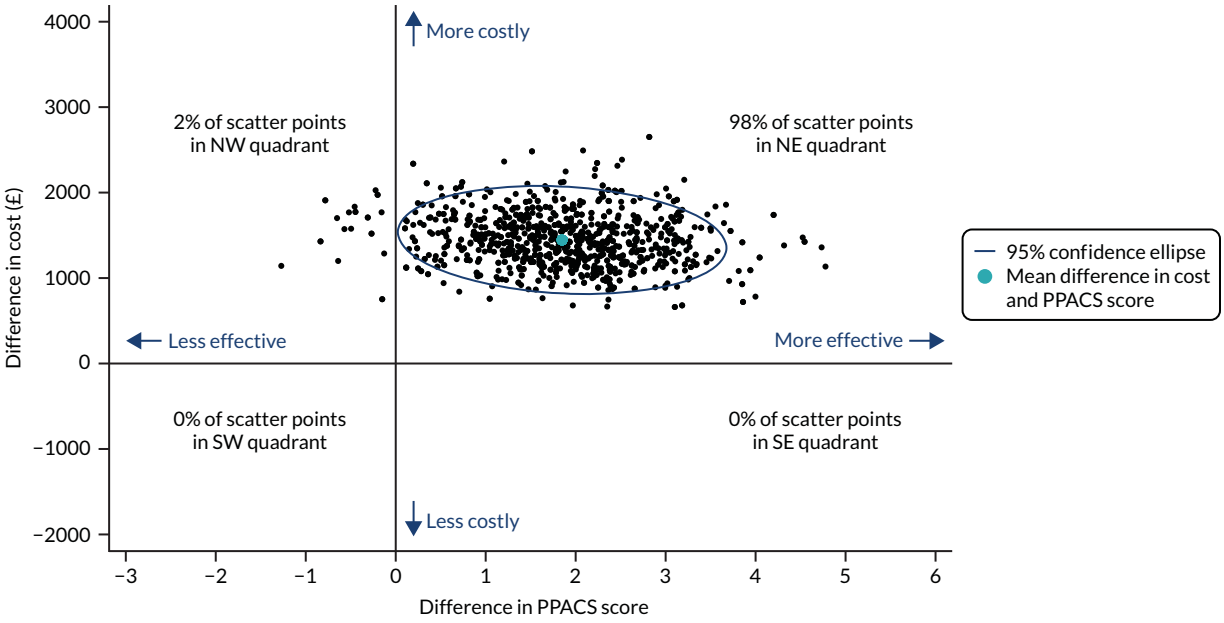


FIGURE 6 Bootstrapped mean differences in costs and PPACS scores at the 24-month follow-up. NE, north-east (more costly, more effective); NW, north-west (more costly, less effective); SE, south-east (less costly, more effective); SW, south-west (less costly, less effective).

The CEAC for the primary cost-effectiveness analysis (*Figure 7*) indicates that VIPP-SD has a higher probability of being cost-effective than usual care at WTP thresholds > £800 for a 1-point improvement in PPACS score on a scale of 0–70 points (equivalent to 0.10 SD based on usual-care group SD post treatment). However, it is not possible to make firm conclusions about the cost-effectiveness of VIPP-SD without agreed thresholds for society's WTP for improvements in PPACS score.

Sensitivity analyses

The results of the two sensitivity analyses – (1) examining the effect of multiple imputation and (2) examining the results at the 5-month follow-up – are presented in *Table 21*. The distribution of

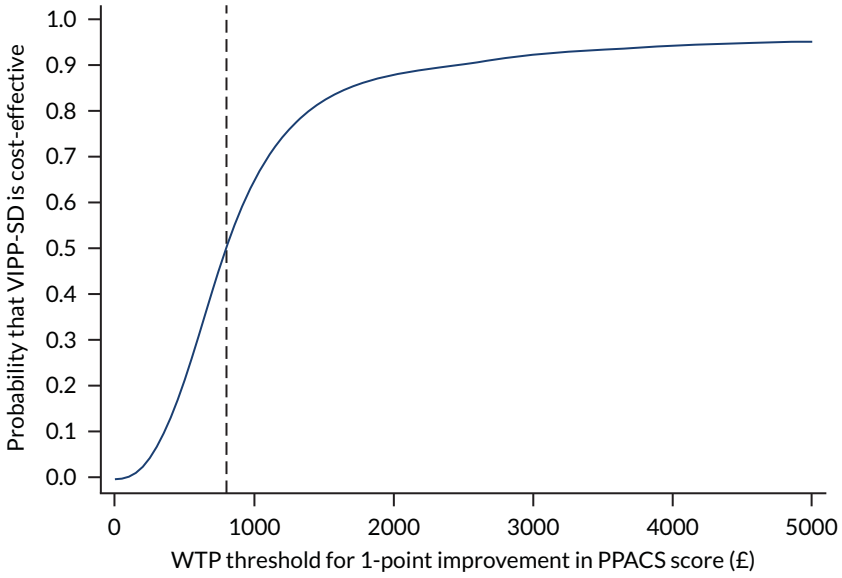


FIGURE 7 Cost-effectiveness acceptability curve showing the probability that VIPP-SD is cost-effective compared with usual care at different values of WTP thresholds for a 1-point improvement in PPACS score at the 24-month follow-up. The dashed line represents the point at which a probability of 0.5 (50%) is reached.

TABLE 21 Mean total cost (£) and outcomes per participant for sensitivity analyses

	Trial group, mean (SE)		VIPP-SD minus usual care		Distribution of bootstrap iterations (%)			
Analysis	VIPP-SD	Usual care	Adjusted difference[a]	95% CI	NW	NE	SW	SE
Base case	**n = *140***	**n = *142***						
Total cost (£)	3131.93 (416.54)	1525.38 (293.49)	1449.99[b]	619.45 to 2280.52[b]	2	98	0	0
PPACS score[c]	23.12 (1.08)	24.68 (0.76)	−1.87[b]	−4.10 to 0.37[b]				
Complete cases	**n = *137*[d]**	**n = *142***						
Total cost (£)	2638.89 (231.84)	1525.38 (162.46)	1175.17	730.42 to 1619.92	7	93	0	0
PPACS score[c]	23.14 (1.10)	24.68 (0.78)	−1.85	−3.85 to 0.15				
5-month follow-up	**n = *128***	**n = *124***						
Total cost (£)	1827.55 (120.71)	510.18 (85.35)	1281.34[b]	1022.40 to 1540.29[b]	0	100	0	0
PPACS score[c]	28.68 (1.22)	30.34 (0.87)	−2.55[b]	−4.70 to −0.39[b]				

NE, north-east (more costly, more effective); NW, north-west (more costly, less effective); SE, south-east (less costly, more effective); SW, south-west (less costly, less effective).

a Adjusted for recruitment centre, age of child at recruitment, number of caregivers participating, time to follow-up and baseline score.

b Comparison based on multiply imputed data.

c Lower scores indicate better outcomes.

d Excluding influential outliers, i.e. participants with total costs in the 99th percentile.

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

bootstrap iterations across the cost-effectiveness plane changes for both analyses compared with base case, but, in both cases, the majority of scatter points fall in the north-east quadrant, indicating that the VIPP-SD group has better outcomes but also higher costs than the usual-care group.

In the sensitivity analysis considering complete cases and excluding influential outliers, the adjusted difference in mean costs is lower (£1175.17) than in the base-case analysis (£1449.99); meanwhile, the PPACS scores remain relatively unchanged. In the sensitivity analysis examining results at the 5-month follow-up, the adjusted mean difference is greater for both the costs (£1281.34) and PPACS score (−2.55) than in the base-case analysis (£1449.99 and −1.87, respectively).

The cost-effectiveness results of the sensitivity analysis considering complete cases and excluding influential outliers were very similar to those of the base-case analysis. The scatterplot of bootstrapped mean differences and PPACS scores for this analysis (*Figure 8*) shows that the majority of scatter points (93%) lie in the north-east quadrant (i.e. VIPP-SD is more costly and more effective than usual care) whereas the remaining scatter points (7%) lie in the north-west quadrant, suggesting that VIPP-SD is more costly and less effective than usual care. The CEAC for this analysis (*Figure 9*) suggests that VIPP-SD has a higher probability of being cost-effective than usual care for WTP thresholds > £800 for a 1-point improvement in PPACS score on a scale of 0–70 points (equivalent to 0.10 SD based on usual-care group SD post treatment).

The cost-effectiveness results at the 5-month follow-up were also similar to those in the base-case analysis. The scatterplot of bootstrapped mean differences and PPACS scores for this analysis (*Figure 10*) shows that almost 100% of the scatter points lie in the north-east quadrant, suggesting that VIPP-SD is more costly and more effective than usual care. The CEAC for this analysis (*Figure 11*) suggests that VIPP-SD has a higher probability of being cost-effective than usual care for WTP thresholds > £625 for a 1-point improvement in PPACS score on a scale of 0–70 points (equivalent to 0.10 SD based on usual-care group SD at post treatment).

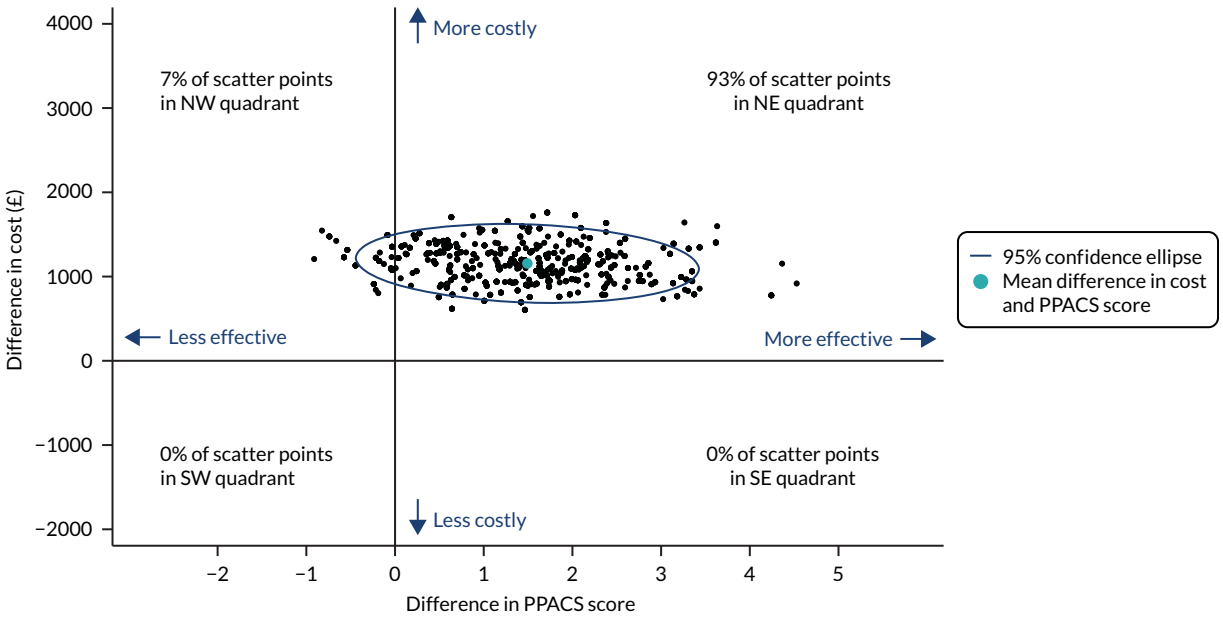


FIGURE 8 Bootstrapped mean differences in costs and PPACS scores at the 24-month follow-up: complete cases and excluding influential outliers. NE, north-east (more costly, more effective); NW, north-west (more costly, less effective); SE, south-east (less costly, more effective); SW, south-west (less costly, less effective).

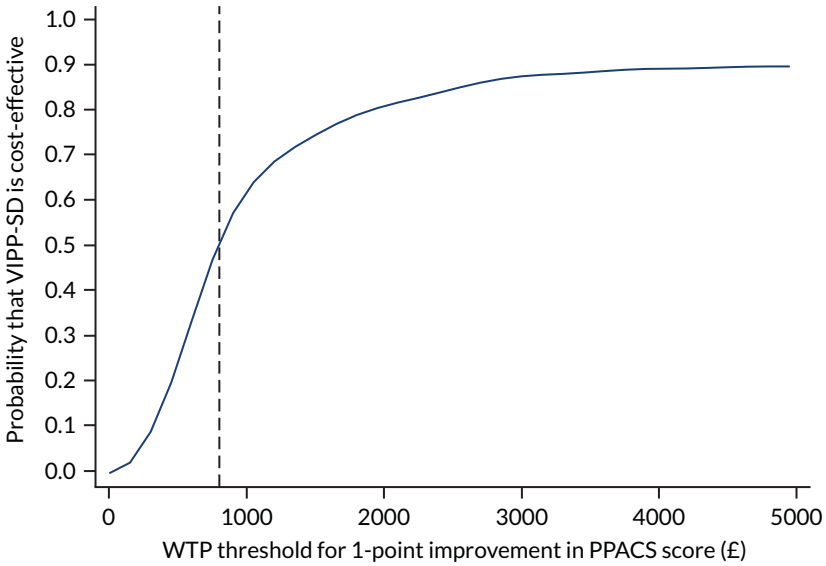


FIGURE 9 Cost-effectiveness acceptability curve showing the probability that VIPP-SD is cost-effective compared with usual care at different values of WTP thresholds for a 1-point improvement in PPACS score at the 24-month follow-up: complete cases and excluding influential outliers. The dashed line represents the point at which a probability of 0.5 (50%) is reached.

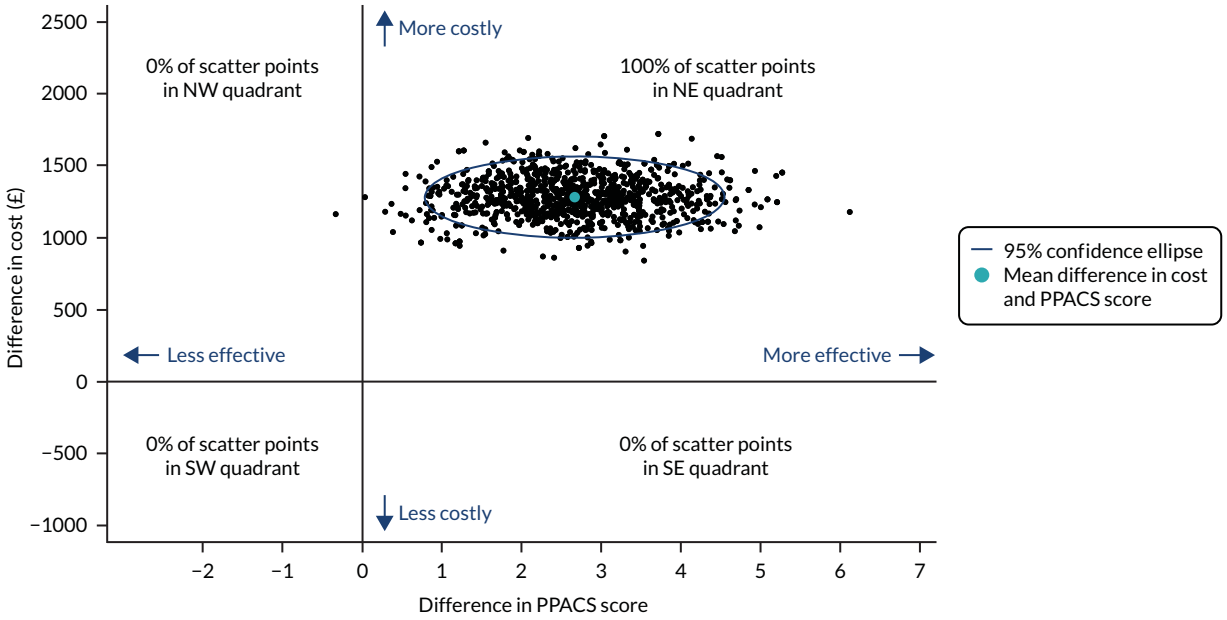


FIGURE 10 Bootstrapped mean differences in costs and PPACS scores at the 5-month follow-up. NE, north-east (more costly, more effective); NW, north-west (more costly, less effective); SE, south-east (less costly, more effective); SW, south-west (less costly, less effective).

Cost–consequences analysis at the 24-month follow-up

The 24-month costs and the key secondary outcome measures are summarised in *Table 22*. Alongside significantly higher costs, there were no significant differences in either the CBCL or the SDQ at the 24-month follow-up, although observed differences were generally in favour of the VIPP-SD group. A similar pattern was seen for the additional secondary outcome measures (Parenting Scale, PHQ-9, GAD-7, RDAS; see *Table 10*). These results suggest that the application of an alternative secondary outcome measure would produce similar results to those already demonstrated using the PPACS.

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

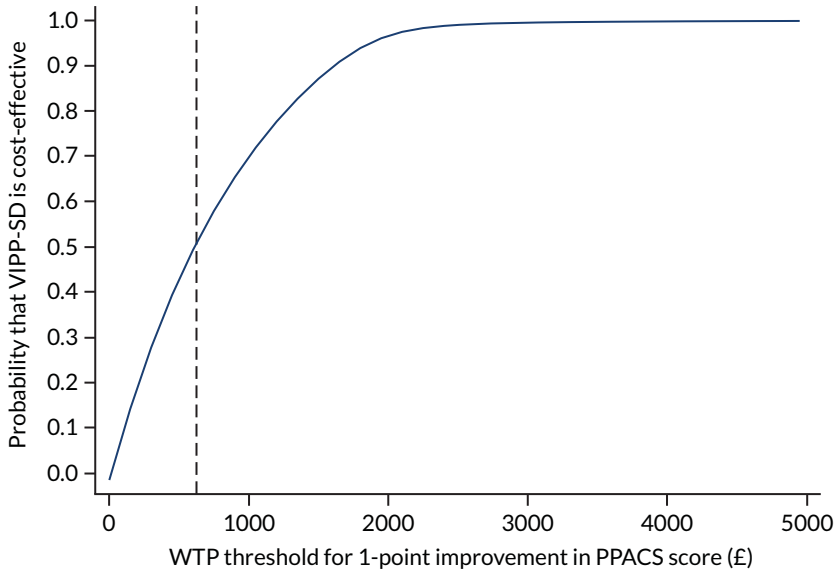


FIGURE 11 Cost-effectiveness acceptability curve showing the probability that VIPP-SD is cost-effective compared with usual care at different values of WTP thresholds for a 1-point improvement in PPACS score at the 5-month follow-up. The dashed line represents the point at which a probability of 0.5 (50%) is reached.

TABLE 22 Cost–consequences analysis at the 24-month follow-up

	Trial group						
	VIPP-SD		Usual care		VIPP-SD minus usual care[a]		
Outcome measure	*n*	Mean (SE)	*n*	Mean (SE)	Adjusted difference[b]	95% CI	*p*-value
Total costs	140	3131.93 (416.54)	142	1525.38 (293.49)	1449.99	619.45 to 2280.52	< 0.01
PPACS[c]	140	23.12 (1.08)	142	24.68 (0.76)	−1.87	−4.10 to 0.37	0.10
CBCL[c]	140	30.58 (2.79)	144	35.31 (1.96)	−2.81	−7.22 to 2.07	0.24
SDQ[c]	141	10.40 (0.66)	144	10.92 (0.47)	−0.26	−1.39 to 0.88	0.66
Female caregivers							
Parenting Scale[d]	136	3.02 (SD 0.53)	142	3.02 (SD 0.57)	0.03	−0.07 to 0.13	0.51
PHQ-9[d]	136	3.99 (SD 4.60)	141	4.02 (SD 4.22)	0.05	−0.86 to 0.97	0.91
GAD-7[d]	136	4.12 (SD 4.64)	141	4.20 (SD 4.09)	0.16	−0.76 to 1.07	0.74
RDAS[d]	120	50.01 (SD 8.15)	115	50.59 (SD 6.96)	−0.40	−1.78 to 0.99	0.57
Male caregivers							
Parenting Scale[d]	28	2.9 (SD 0.49)	24	2.95 (SD 0.45)	0.15	−0.06 to 0.36	0.15
PHQ-9[d]	28	2.64 (SD 2.15)	24	2.54 (SD 3.08)	0.11	−1.21 to 1.42	0.87
GAD-7[d]	28	2.54 (SD 2.32)	24	2.50 (SD 2.70)	−0.51	−1.59 to 0.57	0.35
RDAS[d]	27	51.37 (SD 4.87)	24	48.82 (SD 6.78)	−1.27	−3.97 to 1.42	0.35

SE, standard error.
a Comparison based on multiply imputed data.
b Adjusted for recruitment centre, age of child at recruitment, number of caregivers participating, time to follow-up and baseline score.
c Lower scores indicate better outcomes.
d Values reproduced from *Table 10*.

Long-term cost-effectiveness

Table 23 reports the utilities and QALYs generated using the mapping algorithm for the SDQ. As a result of the small difference between trial groups in SDQ scores (see *Table 22*), the utilities generated from the SDQ were similar between the two groups, which resulted in very small differences in QALYs at the 5-month follow-up (unadjusted and adjusted difference 0.010) and at the 24-month follow-up (unadjusted difference 0.007, adjusted difference 0.003). These small differences in mapped QALYs between trial groups at the 24-month follow-up, alongside no difference in costs, with the exception of the cost of the VIPP-SD intervention, are not suggestive of possible future cost-effectiveness differences between the two groups; thus, longer-term economic decision modelling was not considered appropriate.

Discussion

Summary of results

The primary cost-effectiveness analysis suggests that VIPP-SD is associated with higher health and social care costs as a result of the additional cost of the VIPP-SD intervention and non-significant observed differences in PPACS scores that favour the intervention over usual care. When combined, the probability of VIPP-SD being cost-effective compared with usual care increases as the WTP for improvements in PPACS score increases, with VIPP-SD having the higher probability of being cost-effective at WTP values of approximately £800 and above per 1-point improvement in PPACS score (equivalent to 0.10 SD based on usual-care group SD at post treatment). Because the PPACS is not associated with a WTP threshold to support decision-making (compared with QALYs with a NICE WTP threshold of £20,000–30,000 per QALY), it is not possible to come to any firm conclusion about the relative cost-effectiveness of VIPP-SD in the short term. These results were robust to changes in assumptions in sensitivity analyses (complete case, excluding outliers and analysis at the 5-month follow-up).

TABLE 23 Mean utility scores and QALYs per participant mapped from the SDQ

	Trial group							
	VIPP-SD		Usual care		VIPP-SD minus usual care[a]			
Quality of life[b,c]	*n*	Mean (SE)	*n*	Mean (SE)	Unadjusted difference	Adjusted difference[d]	95% CI	*p*-value
Baseline (utility)	150	0.820 (0.006)	149	0.817 (0.004)	0.003	0.003	−0.008 to 0.014	0.576
5 months (utility)	140	0.836 (0.006)	145	0.830 (0.004)	0.004	0.006	−0.003 to 0.015	0.209
24 months (utility)	141	0.842 (0.007)	144	0.838 (0.005)	0.003	0.001	−0.011 to 0.012	0.926
5 months (QALYs)	139	0.345 (0.002)	145	0.343 (0.002)	0.001	0.001	−0.001 to 0.003	0.206
24 months (QALYs)	140	1.632 (0.011)	144	1.620 (0.007)	0.007	0.003	−0.011 to 0.173	0.656

SE, standard error.
a Comparison based on multiply imputed data.
b Higher scores indicate better outcomes.
c Mapped from SDQ subscale scores.
d Adjusted for recruitment centre, age of child at recruitment, number of caregivers participating, time to follow-up and baseline score.

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

The cost of the VIPP-SD intervention was the key cost driver in the cost-effectiveness analyses presented. As such, it is important to note that intervention costs commonly reduce in the longer term, as therapists become more experienced, and service providers and commissioners benefit from both experience and economies of scale. In particular, the average length of sessions provided in the current trial was 1.5 hours, which may be longer than when sessions are completed by more experienced therapists, as all of the VIPP-SD trial therapists were new to the VIPP-SD intervention and had been trained only recently. This should be considered in relation to any longer-term plans to invest in VIPP-SD for this population.

The lack of differences in the SDQ at the 24-month follow-up precluded longer-term decision modelling as there was no evidence, using the prespecified measure of outcome, of a difference relating to effectiveness or, indeed, costs (excluding the cost of VIPP-SD). This conclusion may have differed if the focus had been on the PPACS, but, unlike the SDQ, the PPACS is not associated with either a mapping algorithm capable of generating QALYs or any suitable longitudinal data to identify future service use, costs or effects; thus, the use of the PPACS for modelling was not feasible.

Strengths and limitations

The economic evaluation benefited from a large sample with data collected through a robust and scientifically rigorous RCT. Although it is difficult to come to any firm conclusions regarding the cost-effectiveness of VIPP-SD compared with usual care, as this is the first cost-effectiveness analysis of VIPP-SD, the study provides valuable information that supports the methodological development of future economic evaluations.

The lack of a WTP threshold for the PPACS was a limitation. However, the key limitation around the measurement of effects is the lack of valid methods for measuring health-related quality of life, and thus QALYs, in children as young as those in the current study. At the time that the proposal and protocol were developed, no appropriate methods existed. Nearer the end of the study, the authors became aware of guidance for the proxy completion of the CHU-9D by parents of children aged < 5 years (personal communication with the developer of the measure, Dr Katherine Stevens, University of Sheffield, 25 March 2019), but this was still being tested and its validity for children as young as those aged 1 or 2 years is uncertain (given questions around worry, sadness and daily routine). Until appropriate methods for the measurement of health-related quality of life in children of this age exist, an economic evaluation will continue to prove difficult in very young populations.

To compensate for this limitation, we used a published algorithm to map health state utilities from clinical outcomes using the SDQ. Mapping is a widely used approach in economic evaluation where direct measurement of health utilities is not possible,[83–85] and is recommended by NICE and the International Society for Pharmacoeconomics and Outcomes Research (ISPOR).[65,86] However, the validity of the methods used is debated in the literature,[87–89] loss of information at each stage of transformation leads to uncertain estimates and thus potentially erroneous conclusions, and proxy reporting by parents is also associated with a number of limitations.[90,91] The mapping algorithm applied in the current study was identified in a recent systematic review as the only algorithm that converts any of the outcomes in the current study to child health utilities.[92] However, the algorithm was developed in Australia with data collected from adolescents; thus, its relevance to much younger children in the UK is uncertain. Since starting work on this analysis, a second mapping algorithm has been published that the authors argue is 'better performing' and maps from 25 individual items from the SDQ, as opposed to the five subscales.[93] However, the population remains the same (i.e. Australian adolescents).

The short-term nature of the within-trial analysis is also a limitation as participants are less likely to incur costs for resource use relating to behavioural problems during this preschool period. Service use, and thus further costs, are more likely when the child is older. It is difficult to map trajectories without suitable longitudinal data; most publicly available data sets tend to be cross-sectional, out of date,

small in terms of sample size, and thus of limited generalisability, and missing key parameters to allow for calculation of all relevant costs and health state utilities. Furthermore, combining data from multiple sources to extrapolate results to different populations and over longer periods of time would require many assumptions and huge uncertainties, and is likely to produce poor estimates for cost-effectiveness analyses. Although longer-term modelling can be a useful decision tool, it can also be a source of many biases. This is particularly true for studies in very young children, where the absence of reliable input data means that surrogate measures must often be used.

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

Chapter 5 Discussion

Summary of findings

We found clear evidence that a brief, home-based intervention, VIPP-SD, was more effective than usual care in reducing behaviour problems in this group of children aged 1 and 2 years. The effect was strongest for children's conduct problems, with evidence suggesting that most of the effect was sustained at a 24-month follow-up. This intervention uses video feedback from predominantly play-based sessions with a caregiver and their child, where the therapist shares feedback to promote sensitive responding and consistent discipline strategies with parents/caregivers of young children.

For the primary outcome (the interview-based PPACS assessment) at the 5-month post-treatment assessment, evidence of superiority was found for VIPP-SD. The secondary analysis, which tested the benefits of the intervention for those participants who received the core number of sessions (i.e. four sessions or more), showed a larger reduction in behaviour problems. This evidence, showing that increased exposure to the intervention further intensified its benefits, endorses the positive effect of the VIPP-SD programme. The VIPP-SD treatment effect was also strongest on the conduct problems scale of the PPACS assessment rather than the ADHD/hyperkinesis scale. This is particularly of note as social learning theory and the sensitive discipline component of the VIPP-SD programme specifically target conduct problems. Thus, these differential effects are in keeping with expectations and it is instructive to identify that the effects on conduct problems do not appear to extend to ADHD symptoms. Furthermore, evidence, including new evidence since this trial was designed, strongly suggests that it is conduct problems in particular that predict worse long-term outcomes for children.[94] Relatedly, early childhood conduct problems in particular, rather than hyperkinetic/ADHD problems only, are associated with greater health-care and criminal justice costs in adulthood.[9] Findings on the two secondary measures of child problems (CBCL and SDQ questionnaires) were consistent in suggesting an early benefit of VIPP-SD for children's behaviour problems.

The findings at the longer-term follow-up (24 months post randomisation) indicated a sustained effect on children's behaviour problems. The main outcome measure (PPACS) showed a small diminution of the treatment effect (approximately 15%) over the time between the 5- and 24-month follow-ups. This meant that the 95% CI did overlap with zero (with a *p*-value of 0.08), but the consistency of effect across the main outcome measures suggests that an ongoing treatment benefit is likely. The standardised effect size for the treatment difference on the total PPACS measure was 0.20 at 5 months post randomisation and 0.17 at 24 months post randomisation. Although both effect sizes are small, they are of the magnitude that may be expected to make a real difference for a brief early intervention rolled out across a large population. Similarly, the effect on the conduct disorder scale at 24 months, although somewhat reduced (0.30 standardised effect size at 5 months post randomisation and 0.20 at 24 months post randomisation), is likely to represent a meaningful effect of the intervention.

The main findings are consistent with previous, smaller studies of VIPP-SD. A recent meta-analysis[32] demonstrated a similar effect size to that seen in the present study for children's behaviour problems (0.26) and larger effect sizes for parental sensitivity (0.47). Previous VIPP-SD research also indicates the intriguing possibility that certain parents and children may be especially sensitive to the benefits of the intervention and show greater improvements in outcomes – this is known as differential susceptibility.[95–97] Data on which to base our original power calculation were limited, and in the original protocol we had anticipated a larger possible effect size than that seen in the present study. Some caution is therefore warranted in interpreting these findings. The effect size found for the VIPP-SD intervention in this study (0.20) is somewhat difficult to interpret for an individual child or family; however, it represents a mean difference of 2 points on the PPACS measure. As an example, for tantrums, this 2-point difference would

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

equate to a change from severe (breaking things) to mild (shouting) or a change in frequency from daily to once or twice per week. Although tantrums are a normative feature of early childhood, daily tantrums are not normative and occur in < 10% of preschool-aged children.[98]

As the first pragmatic trial of VIPP-SD conducted in a routine NHS health-care setting, it is striking that the effectiveness of the intervention is robust in this setting. The intervention is also highly acceptable for a psychological treatment as indicated by the high levels of adherence. There was also positive feedback from families, although this was limited to a small proportion of participating families. A substantial body of evidence for the effectiveness of several interventions already exists for childhood behaviour problems; however, this is the first study, to our knowledge, that demonstrates a beneficial treatment effect in such young children and in a pragmatic trial in routine practice. This is important as there are many examples of promising interventions for child mental health problems that do not show a clear benefit over routine care in NHS practice settings, despite some initial evidence of effectiveness in other, sometimes more tightly controlled, trial conditions.[99–101] Consequently, the findings of this study can be more reliably extrapolated to routine clinical care delivered by NHS staff, including health visitors and community nursery nurses.

The findings of this study are also able to partly address a potential concern about brief early interventions, such as VIPP-SD, for childhood behaviour problems, namely whether or not brief interventions are sufficient to be effective for children with more severe behaviour problems. The findings from this study suggest that the effect of the intervention was greater in children with more severe behaviour problems. Children with more severe behaviour problems at the beginning of the study tended to show more improvement in their behaviour, which inspired for confidence that this predominantly health visitor-led intervention can benefit families and children who may need it most. The findings suggest that interventions might be most beneficially targeted at those children with higher levels of behaviour problems. Similarly, the findings show that the intervention is effective even in children aged 1 year. The findings also give some confidence that those who may benefit from treatment can be identified using a simple screening questionnaire (SDQ).[42] The high levels of adherence and fidelity of treatment delivery seen in this pragmatic trial also suggest that this intervention can be reliably and successfully delivered in routine NHS practice.

There was no benefit of the intervention on other secondary outcome measures of parental mood, anxiety or relationship adjustment. Parent-mediated interventions often anticipate effects on parental well-being; however, this evidence suggests that these programmes are unlikely to be sufficiently powerful in and of themselves to confer benefits on global assessments of parental distress.[102]

Economic evaluation

The primary economic analysis suggests that VIPP-SD is associated with significantly higher costs than usual care as a result of the additional cost of the VIPP-SD intervention and greater improvements in behaviour. When these costs are combined with the differences in PPACS scores that favour the intervention over usual care, the probability of VIPP-SD being cost-effective compared with usual care increases as the WTP for improvements in PPACS scores increases, with VIPP-SD having a higher probability of being cost-effective at WTP values of approximately £800 and above per 1-point improvement in PPACS score (equivalent to 0.10 SD). In theory, this would equate to a cost of approximately £7920 for 1 SD improvement. As the PPACS is not associated with a WTP threshold to support decision-making, it is not possible to come to any firm conclusions about the relative cost-effectiveness of VIPP-SD in the short term. Any judgement of costs needs to be weighed against the lifetime costs of behaviour problems, estimates for which range from £90,000 per case for those with subthreshold behaviour problems to £280,000 for those with early conduct disorder.[10] Indeed, the long-term costs for conduct disorder in particular are so high that early interventions that generate even modest improvements could provide a positive return on investment in the long term.[10]

The cost of the VIPP-SD intervention was the key cost driver in the cost-effectiveness analyses. It is difficult to contextualise the cost of the intervention (mean cost £1466) because there is a lack of publicly available information on the costs of early intervention programmes,[103] differences in costing methods and few cost-effectiveness evaluations of similar early intervention programmes. However, the costs identified in the current study are lower than the unit costs (£1612–2418) reported from a group-based parenting intervention (Incredible Years) for preschool aged children delivered in the UK.[104] That study reported that a 1-point improvement on the SDQ would cost £1295 on a 40-point scale (where 1 point represents 0.14 of the SD of the control group at follow-up).[104] Overall, effective early interventions for preschool-aged children at risk of conduct disorder that cost in the region of £1500 are seen as relatively low cost.[10]

It is also noteworthy that intervention costs commonly reduce in the longer term as therapists become more experienced, and service providers and commissioners benefit from both experience and economies of scale. In particular, the length of sessions (1.5 hours) in the current trial may reflect the fact that all of the VIPP-SD trial therapists were new to the intervention and had been trained only recently. Similarly, the length of sessions in the current trial includes the longer sessions needed to deliver VIPP to two caregivers in some cases; thus, we would expect the length of sessions to be shorter if VIPP-SD was delivered to only one caregiver.

Strengths and limitations

A strength of this study is that it is a multisite, pragmatic, RCT that recruited families in routine health-care settings in urban, suburban and rural settings in the UK. The VIPP-SD intervention was acceptable to participants in this setting, with very high levels of retention (95%) achieved and very low levels of missing data. The intervention was largely delivered as intended, with the majority of assessments indicating that therapy met the necessary threshold for adherence to the manual and that the vast majority of participants received the full dosage of the programme. The intervention was delivered by NHS staff in a pragmatic context, and this is the first time, to our knowledge, that VIPP-SD has been investigated in this context. This is also the largest study of VIPP-SD to date, the first cost-effectiveness study of VIPP-SD to be attempted and one of the earliest intervention studies of children demonstrating risk for developing enduring mental health problems in the UK. The study also benefited from the use of a structured interview to assess the primary outcome, which is considered the gold standard outcome measure for mental health problems.[46] Efforts to maintain blinding were largely effective and, in the very small number of cases where research assessors were unblinded, a second researcher rescored the assessment to minimise any risk of bias.

The study had limitations. First, as with many RCTs, the participants in the final sample had a higher level of education than the general population, such that 64% of participants had a graduate-level qualification, compared with ≈ 40% of people aged 25–34 years in England (based on the 2011 Census).[105] This is notable as low parental education is a risk factor for persistent behaviour problems in young children.[10] As a large proportion of families were recruited from London (where educational attainment tends to be higher than national levels), caution should be exercised when considering the generalisability of the study findings to the rest of the UK. In most other respects, the sample was generally representative of the broad range of communities that participated in the study.

Second, approximately half of those eligible for the trial proceeded to take part. Although this is similar to the rate of progression seen in many RCTs of psychological intervention[96,97] and may relate to taking part in a clinical trial rather than to taking part in routine intervention, it does suggest that not all families will engage in this intervention in routine practice, and so alternative forms of help and provision are will also be necessary.

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

Third, we found that not all therapists were able to deliver the intervention, largely because of service changes that meant that the time planned for delivery of the therapy became restricted. This was evident in the variation in the number of VIPP-SD cases delivered by individual therapists (range 1–12 cases, median three cases). Thus, it is possible that the treatment effect observed here is an underestimation of the effect that may be seen over a longer implementation period as therapists become more experienced.

A limitation of the health economic analysis is the lack of valid methods for measuring health-related quality of life, and thus QALYs, in young children. As a result, we used a published algorithm to map health state utilities from clinical outcomes using the SDQ, as mapping is a widely used approach in economic evaluation where direct measurement of health utilities is not possible and mapping is recommended by NICE and ISPOR.[65,86] However, the validity of the methods used is debated in the literature,[87–89] loss of information at each stage of transformation leads to uncertain estimates and thus potentially erroneous conclusions, and proxy reporting by parents is also associated with a number of limitations.[90,91] The mapping algorithm applied in the current study was identified in a recent systematic review as the only algorithm available that converts any of the outcomes included in the current study to child health utilities.[92] However, the algorithm was developed in Australia with data collected from adolescents and, therefore, its relevance to much younger children in the UK is uncertain. The short-term nature of the within-trial analysis is also a limitation as participants are less likely to incur costs for resource use relating to behaviour problems during this preschool period. Service use, and thus further costs, are more likely in later years.

It is worth noting that there is a lack of evidence for effective preventative interventions for behaviour problems in very young children and a lack of services delivering interventions in the UK. The findings of this study do, therefore, offer important evidence to suggest that positive outcomes can be achieved with the VIPP-SD intervention.

Conclusions

Implications for health care

This is the largest RCT of the VIPP-SD intervention, a brief, home-based intervention for parents of young children. It is also the first trial to be conducted in a routine NHS setting. The findings demonstrate that the intervention is effective at reducing behaviour problems in very young children (aged 1 or 2 years) and there is strong potential for the intervention to be delivered successfully by NHS staff in routine practice. There was a positive endorsement of the home-based delivery format by families and the study shows that the intervention can be delivered by health visitors and community nursery nurses. The intervention could provide an important means of early detection and treatment of very early mental health problems in the community before these problems become established and more difficult to treat.

The presented study is also one of the first pragmatic trials in the UK to demonstrate an effect of an early intervention for psychological difficulties in young children. It is well established in empirical and policy forums that the scope for benefit from early preventative intervention is huge.[33–35] What is largely lacking at present is evidence of effective programmes. The findings of the present study represent a significant step towards this. It is particularly striking that the intervention is effective in children as young as 1 year of age, as well as those with more severe behaviour problems. Although the mean effect is small in magnitude, it is notable that it can be achieved this early in life, when there are few effective treatments. This is important, as early benefits could act to potentiate the effects of interventions delivered at later developmental points for those children at greatest risk of conduct disorder.[38]

The VIPP-SD intervention can be delivered successfully in routine NHS care and to specified groups of children who may benefit most, who can be identified using a simple, brief screening questionnaire.

There is significant scope for the intervention to be incorporated relatively easily in routine practice. The findings are especially notable given that there are no standard care pathways in the NHS for early-onset behaviour problems.

Implications for future research

The potential prize of successful early preventative intervention is substantial, namely the opportunity to prevent future suffering and to significantly reduce the large economic burden associated with mental health problems across many sectors of government, including health, education, social care and future employment. The findings of the present study suggest that there is hope for early interventions to be able to deliver positive outcomes in real-world pragmatic settings. Although these are findings in young children, findings from a large body of research suggest that there are significant long-term risks of untreated behaviour problems.[2,5]

We suggest four key implications for future research.

First, further study is needed to assess the potential longer-term outcomes of interventions such as VIPP-SD. Our understanding to date is largely based on modelling studies or extrapolation from findings of interventions in older children, which may or may not turn out to be accurate when applied to younger children. Longer-term follow-up of early interventions could greatly improve estimates of cost-effectiveness as costs are more likely to accrue in later years.

Second, further study of other interventions and other time points for intervention is necessary. Research suggests that a single intervention delivered in isolation is unlikely to be a panacea for behaviour problems. Rather, repeated, efficient interventions delivered at different ages are likely to be necessary to most effectively reduce the future burden of mental health problems across the life course.[37–39] In this context, the findings of the present study are important, as a corollary of the need for staged interventions is the identification of evidence-based programmes that can be delivered to very young children at the first onset of behaviour problems. We already have some understanding of different interventions that have been shown to be effective for older children, but this is such an important area for practice that greater knowledge is needed. This further study will need to include assessment of when and if booster sessions are needed after an initial brief intervention. Similarly, it will be important to test whether or not the initial effects of early intervention such as VIPP-SD can be sustained or further potentiated when followed by later evidence-based interventions. From a cost-effectiveness point of view, the development of a 'whole disease model' that can model interventions across different ages simultaneously may prove to be of value.[106]

Third, further study is needed of how the underlying mechanisms of effective early interventions such as VIPP-SD work and for whom they may work best. The VIPP-SD intervention benefits from a clearly articulated conceptual framework that targets two key parenting behaviours: sensitivity and discipline. Identifying the contribution of these components to the programme's effects will allow better refinement of the intervention, focus on the most effective components and, therefore, make treatment more effective and potentially more efficient for delivery at scale. Future research could also adopt a differential susceptibility perspective to investigate whether or not the intervention is more effective in those parents and children who tend to benefit more from improvements in their environment. Taken together, this research may allow us to move to a position where we can offer interventions to those for whom we know they work best.

Fourth, there is a need for further methodological development that can support economic evaluations of early interventions targeting young children. Specifically, there is a need for valid tools for measuring health-related quality of life, and thus QALYs, in very young children, as well as mixed-methods approaches that can capture a range of outcomes (e.g. acceptability of interventions) that are likely to be important to decision-makers and commissioners.[84,85]

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

In conclusion, this study suggests that we have an early intervention that is effective in reducing behaviour problems in young children, is highly acceptable to parents and can be successfully delivered in routine NHS practice. There is some evidence of continuing benefit of the VIPP-SD intervention 24 months later, but we can be less certain about its value for money. Early intervention represents a huge opportunity for the future positive development of young children, and a lack of effective interventions is one of the key problems holding back the field. The results of this study provide a significant step forward and represent a new opportunity for effective early childhood intervention to prevent enduring mental health problems.

Acknowledgements

The trial was sponsored by Imperial College London.

This research was funded by the NIHR HTA programme (13/04/33). Additional funding was received from the NIHR Biomedical Research Centre based at Imperial College Healthcare NHS Trust and Imperial College London.

The authors would like to thank the families who participated in the trial, the therapists who delivered the trial intervention, the clinical supervisors, and all of the sites that facilitated intervention delivery and assisted with recruitment, as well as thanking them for their continued support. We thank Camilla Rosan, Esther Wilkinson and Mike Crawford for their assistance with study set-up. We are very grateful to the North West London NIHR Collaborations for Leadership in Applied Health Research and Care (CLAHRC) and the North West London and North Thames CRNs for their support. For their recruitment and site support, we thank Lynis Lewis, Angela Williams, Sandra O'Sullivan, Letitia Coco-Bassey, Maggie Waters, Jane McGrath, Philippa Kemsley, Kathryn Simpson, Lynda Rowlinson, Penny Kenway, Andrew Molodynski, Jennifer Potts, Sarah Mather, James Sinclair, Awulatu Sellick-Taylor, Clare Knight, Mary Cousins and Diane Hammond.

We thank Jane Iles, Eloise Stevens and Rajinder Ballman for their role in delivering training, the intervention and providing clinical supervision. We thank all those involved in the delivery of the intervention, including: Toni Adewale, Fola Akinmutande, Rupali Archaya, Kirsten Barnicot, Florence Bristow, Emma Brown, Sadie Burton, Mary Cousins, Jill Domoney, Annette Eneberi, Diane Hammond, Marina Fabrega Ribera, Janis Griffiths, Anna Hemphill, Philippa Kemsley, Leonie Lee-Carbon, Lorraine Mireku, Ainoa Mateu Mullor, Vangilista Nduku, Rachael Neville, Emily Pearson, Clare Pollard, Uma Purohit, Jacqueline Sangan, Joanna Scales, Caroline Thomas, Fatima Valencia Agudo, Verity Wilkinson and Sylvia Woolley. We thank Ainoa Mateu Mullor and Emily Pearson for their work in developing and conducting the fidelity assessments.

We thank Beth Barker, Ellen Grimås, Holly Mattock, Charlotte Phillips and Rachael Ryan for conducting the research assessments.

We are grateful for the administrative support from Nicole Hickey, Hannah Kinch, Ruby Lee and Kate Whitaker. We thank Barbara Barrett and David Daley for delivering training for key trial measures.

We thank Marinus van IJzendoorn, Femmie Juffer and Marian Bakermans-Kranenburg for making the intervention available. We acknowledge the guidance and support of the external TSC and DMEC, in particular the TSC chairperson, Paul Stallard, and the DMEC chairperson, Helen Bedford.

Participating sites:

- Central and North West London NHS Foundation Trust
- Whittington Health NHS Trust
- Oxford Health NHS Foundation Trust
- North East London NHS Foundation Trust
- Cambridgeshire and Peterborough NHS Foundation Trust
- Hertfordshire Community NHS Trust.

We are grateful to the principal investigators and local collaborators at the participating sites: Maggie Waters, Morris Zwi, Nicola Taylor and Alex Davis, Joy Coutts, Sarah Morton and Sarah Barrett.

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

Trial Steering Committee members: Jane Appleton, Erin Bibby, Richard Emsley, Sam Griffith, Bonamy Oliver, Paul Stallard (chairperson) and Essi Viding. Former members: Pasco Fearon and Jonathan Freedman.

Data Monitoring and Ethics Committee members: Helen Bedford (chairperson), Kapil Sayal and Pat Yudkin.

Project Management Group: Daphne Babalis, Marian Bakermans-Kranenburg, Sarah Byford, Poushali Ganguli, Marinus van IJzendoorn, Julia McGinley, Paul Ramchandani, Stephen Scott, Alan Stein, Jane Warwick and Hilary Watt.

Trial managers: Christine O'Farrelly and Jessica Smith.

Patient and public involvement includes Erin Bibby, Sam Griffith, Kajal Hans, Amisha Seda and Sunny Sudhan.

Contributions of authors

Christine O'Farrelly (https://orcid.org/0000-0002-9269-6564) (Trial Manager and Research Associate, Developmental Psychology) was responsible for study management and data collection, and contributed to data interpretation, drafting and revising the report.

Beth Barker (https://orcid.org/0000-0001-7811-3417) (Research Assistant) recruited participants; collected data; and contributed to study procedures and aspects of measurement selection and implementation, and the drafting and revising of the report.

Hilary Watt (https://orcid.org/0000-0001-9576-3753) (Statistician, Public Health) conducted the clinical data analysis and contributed to the writing and editing of the report.

Daphne Babalis (https://orcid.org/0000-0001-9654-4130) (Operations Manager, Clinical Trials) contributed to the oversight and conduct of the trial, and the writing and editing of the report.

Marian Bakermans-Kranenburg (https://orcid.org/0000-0001-7763-0711) (Professor in Clinical Child and Family Studies) contributed to the design of the study, provided expert advice on the parenting intervention and study methodology, and contributed to the writing and editing of the report.

Sarah Byford (https://orcid.org/0000-0001-7084-1495) (Professor of Health Economics) contributed to the design of the study, supervised and undertook data analysis, and contributed to the writing and editing of the report.

Poushali Ganguli (https://orcid.org/0000-0002-1764-5273) (Research Associate, Health Services and Population Research) conducted the health economics analysis, and contributed to the writing and editing of the report.

Ellen Grimås (https://orcid.org/0000-0002-5952-8622) (Research Assistant) recruited participants, contributed to study procedures and aspects of measurement selection and implementation, collected data, co-ordinated the study's PPI, and approved the final version of the report.

Jane Iles (https://orcid.org/0000-0002-3602-9898) (Clinical Psychologist) co-ordinated the delivery of the parenting intervention; contributed to clinical training, supervision and the development of the fidelity assessment; and approved the final version of the report.

Holly Mattock (https://orcid.org/0000-0003-4237-955X) (Research Assistant) collected data, contributed to study procedures and aspects of measurement implementation, co-ordinated the study's PPI and approved the final version of the report.

Julia McGinley (https://orcid.org/0000-0002-4468-7516) (Head of Parent Support) provided expert advice on engaging families in the study and intervention, and approved the final version of the report.

Charlotte Phillips (https://orcid.org/0000-0001-5960-1087) (Research Assistant) recruited participants, contributed to study procedures and aspects of measurement selection and implementation, collected data and approved the final version of the report.

Rachael Ryan (https://orcid.org/0000-0002-2648-8142) (Research Assistant), recruited participants, contributed to study procedures and aspects of measurement selection and implementation, collected data, co-ordinated the study's PPI and approved the final version of the report.

Stephen Scott (https://orcid.org/0000-0003-4680-6213) (Professor of Child Health and Behaviour) contributed to the design of the study, provided expert advice on the evaluation of parenting interventions and behaviour problems, and contributed to the writing and editing of the report.

Jessica Smith (https://orcid.org/0000-0001-5334-6222) (Trial Manager) was responsible for study management and data collection, and contributed to the writing and editing of the report.

Alan Stein (https://orcid.org/0000-0001-8207-2822) (Professor of Child and Adolescent Psychiatry) contributed to the design of the study, provided expert advice on early childhood intervention and psychopathology, and contributed to the writing and editing of the report.

Eloise Stevens (https://orcid.org/0000-0001-7671-3670) (Clinical Therapist) co-ordinated the delivery of the parenting intervention, contributed to clinical supervision and the development of the fidelity assessment, and approved the final version of the report.

Marinus van IJzendoorn (https://orcid.org/0000-0003-1144-454X) (Professor of Human Development) contributed to the design of the study, provided expert advice on the parenting intervention and study methodology, and contributed to the writing and editing of the report.

Jane Warwick (https://orcid.org/0000-0002-7320-6603) (Professor in Clinical Trial Statistics) contributed to the design of the study, supervised and undertook data analysis, and contributed to the writing and editing of the report.

Paul Ramchandani (https://orcid.org/0000-0003-3646-2410) (Professor of Child and Adolescent Mental Health) was chief investigator, conceived and designed the study, was responsible for its conduct, contributed to data interpretation and the drafting and revising of the report.

Publications

Ramchandani PG, O'Farrelly C, Babalis D, Bakermans-Kranenburg MJ, Byford S, Grimas ES, *et al.* Preventing enduring behavioural problems in young children through early psychological intervention (Healthy Start, Happy Start): study protocol for a randomized controlled trial. *Trials* 2017;**18**:543.

Mattock HC, Ryan R, O'Farrelly C, Babalis D, Ramchandani PG. Does a video clip enhance recruitment into a parenting trial? Learnings from a study within a trial. *Trials* 2020;**21**:856.

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

O'Farrelly C, Watt H, Babalis D, Bakermans-Kranenburg MJ, Barker B, Byford S, *et al.* A brief home-based parenting intervention (VIPP-SD) to prevent enduring behaviour problems in young children: a pragmatic randomised controlled trial [published online ahead of print March 15 2021]. *JAMA Pediatrics* 2021.

Data-sharing statement

All of the de-identified data included in the report will be available 12 months following publication and for 5 years after the date of publication. Data will be made available to researchers who provide a methodologically sound and hypothesis-driven proposal, and have the required institutional approvals in place to achieve the aims in the approved proposal. To gain access to the data, proposals should be directed to Professor Paul Ramchandani for approval by the investigator group. Requestors will be asked to sign a data access agreement. The study protocol and consent form will also be made available.

References

1. Sadler K, Vizard T, Ford T, Marchesell F, Pearce N, Mandalia D, *et al. Mental Health of Children and Young People in England, 2017*. Leeds: NHS Digital; 2018.
2. Scott S. Oppositional and Conduct Disorders. In Thapar A, Pine D, Leckman JF, Scott S, Snowling MJ, Taylor EA, editors. *Rutter's Child and Adolescent Psychiatry*. 6th edn. Hoboken, NJ: John Wiley & Sons, Inc.; 2015. pp. 966–80.
3. Wakschlag L, Bennett D, Leventhal L, Thomas J. Disruptive Behavior Disorders & ADHD In Preschool Children: Characterizing Heterotypic Continuities for a Developmentally Informed Nosology for DSM-V. In Narrow W, First M, Sirovatka P, Regier D, editors. *Age and Gender Considerations in Psychiatric Diagnosis: A Research Agenda for DSM-V*. Arlington, VA: American Psychiatric Association; 2007. pp. 243–58.
4. National Collaborating Centre for Mental Health. *Antisocial Behaviour and Conduct Disorders in Children and Young People: The NICE Guideline on Recognition, Intervention and Management*. Leicester: The British Psychological Society and The Royal College of Psychiatrists; 2013.
5. Petitclerc A, Tremblay RE. Childhood disruptive behaviour disorders: review of their origin, development, and prevention. *Can J Psychiatry* 2009;**54**:222–31. https://doi.org/10.1177/070674370905400403
6. Biedzio D, Wakschlag L. Developmental Emergence of Disruptive Behaviors Beginning in Infancy: Delineating Normal–Abnormal Boundaries to Enhance Early Identification. In Zeanah CH Jr, editor. *Handbook of Infant Mental Health*. 4th edn. New York, NY: Guilford; 2018. pp. 407–25.
7. Caspi A, Begg D, Dickson N, Harrington H, Langley J, Moffitt TE, Silva PA. Personality differences predict health-risk behaviors in young adulthood: evidence from a longitudinal study. *J Pers Soc Psychol* 1997;**73**:1052–63. https://doi.org/10.1037//0022-3514.73.5.1052
8. Fergusson DM, Horwood LJ, Ridder EM. Show me the child at seven: the consequences of conduct problems in childhood for psychosocial functioning in adulthood. *J Child Psychol Psychiatry* 2005;**46**:837–49. https://doi.org/10.1111/j.1469-7610.2004.00387.x
9. D'Amico F, Knapp M, Beecham J, Sandberg S, Taylor E, Sayal K. Use of services and associated costs for young adults with childhood hyperactivity/conduct problems: 20-year follow-up. *Br J Psychiatry* 2014;**204**:441–7. https://doi.org/10.1192/bjp.bp.113.131367
10. Gutman LM, Joshi H, Khan L, Schoon I. *Children of the Millennium: Understanding the Course of Conduct Problems During Childhood*. London: Centre for Mental Health; 2018.
11. Miner JL, Clarke-Stewart KA. Trajectories of externalizing behavior from age 2 to age 9: relations with gender, temperament, ethnicity, parenting, and rater. *Dev Psychol* 2008;**44**:771–86. https://doi.org/10.1037/0012-1649.44.3.771
12. Ryan R, O'Farrelly C, Ramchandani P. Parenting and child mental health. *London J Prim Care* 2017;**9**:86–94. https://doi.org/10.1080/17571472.2017.1361630
13. Scott S, Gardner F. Parenting Programs. In Thapar A, Pine D, Leckman JF, Scott S, Snowling MJ, Taylor EA, editors. *Rutter's Child and Adolescent Psychiatry*. 6th edn. Hoboken, NJ: John Wiley & Sons, Inc.; 2015. pp. 1046–61.
14. Mingebach T, Kamp-Becker I, Christiansen H, Weber L. Meta-meta-analysis on the effectiveness of parent-based interventions for the treatment of child externalizing behavior problems. *PLOS ONE* 2018;**13**:e0202855. https://doi.org/10.1371/journal.pone.0202855

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

15. Gardner F, Leijten P, Melendez-Torres GJ, Landau S, Harris V, Mann J, *et al.* The earlier the better? Individual participant data and traditional meta-analysis of age effects of parenting interventions. *Child Dev* 2019;**90**:7–19. https://doi.org/10.1111/cdev.13138

16. Bonin EM, Stevens M, Beecham J, Byford S, Parsonage M. Costs and longer-term savings of parenting programmes for the prevention of persistent conduct disorder: a modelling study. *BMC Public Health* 2011;**11**:803. https://doi.org/10.1186/1471-2458-11-803

17. Doyle O, Harmon CP, Heckman JJ, Tremblay RE. Investing in early human development: timing and economic efficiency. *Econ Hum Biol* 2009;**7**:1–6. https://doi.org/10.1016/j.ehb.2009.01.002

18. Juffer F, Bakermans-Kranenburg M, van IJzendoorn M. *Manual VIPP-SD: Video-Feedback Intervention to Promote Positive Parenting and Sensitive Discipline*. Leiden: Centre for Child and Family Studies, Leiden University; 2008.

19. Juffer F, Bakermans-Kranenburg M, van IJzendoorn M. *Manual Video-feedback Intervention to promote Positive Parenting and Sensitive Discipline (VIPP-SD) (version 3.0)*. Leiden: Centre for Child and Family Studies, Leiden University; 2015.

20. Juffer F, Bakermans-Kranenburg MJ, van IJzendoorn MH. The importance of parenting in the development of disorganized attachment: evidence from a preventive intervention study in adoptive families. *J Child Psychol Psychiatry* 2005;**46**:263–74. https://doi.org/10.1111/j.1469-7610.2004.00353.x

21. Bakermans-Kranenburg MJ, Breddels-van Baardewijk P, Juffer F, Velderman MK, van IJzendoorn MH. Insecure Mothers with Temperamentally Reactive Infants: A Chance for Intervention. In Juffer F, Bakermans-Kranenburg MJ, van IJzendoorn MH, editors. *Promoting Positive Parenting: An Attachment-based Intervention*. New York, NY: Taylor & Francis; 2018. pp. 75–90.

22. Woolley H, Hertzmann L, Stein A. Video-feedback Intervention with Mothers with Postnatal Eating Disorders and Their Infants. In Juffer F, Bakermans-Kranenburg MJ, van IJzendoorn MH, editors. *Promoting Positive Parenting: An Attachment-based Intervention*. New York, NY: Taylor & Francis; 2018. pp. 111–138.

23. van Zeijl J, Mesman J, van IJzendoorn MH, Bakermans-Kranenburg MJ, Juffer F, Stolk MN, *et al.* Attachment-based intervention for enhancing sensitive discipline in mothers of one- to three-year-old children at risk for externalizing behaviour problems. *J Consult Clinical Psychol* 2006;**74**:994–1005. https://doi.org/10.1037/0022-006X.74.6.994

24. Kalinauskiene L, Cekuoliene D, van IJzendoorn MH, Bakermans-Kranenburg MJ, Juffer F, Kusakovskaja I. Supportive insensitive mothers: the Vilnius randomized control trial of video feedback intervention to promote maternal sensitivity and infant attachment. *Child Care Health Dev* 2009;**35**:613–23. https://doi.org/10.1111/j.1365-2214.2009.00962.x

25. Groeneveld MG, Vermeer HJ, van IJzendoorn MH, Linting M. Enhancing home-based child care quality through video-feedback intervention: a randomized controlled trial. *J Fam Psychol* 2011;**25**:86–96. https://doi.org/10.1037/a0022451

26. Moss E, Dubois-Comtois K, Cyr C, Tarabulsy GM, St-Laurent D, Bernier A. Efficacy of a home-visiting intervention aimed at improving maternal sensitivity, child attachment, and behavioural outcomes for maltreated children: a randomized control trial. *Dev Psychopathol* 2011;**23**:195–210. https://doi.org/10.1017/S0954579410000738

27. Pereira M, Negrão M, Soares I, Mesman J. Decreasing harsh discipline in mothers at risk for maltreatment: a randomized control trial. *Infant Ment Health J* 2014;**35**:604–13. https://doi.org/10.1002/imhj.21464

28. Poslawsky IE, Naber FBA, Bakermans-Kranenburg MJ, van Daalen E, van Engeland H, van IJzendoorn MH. Video-feedback Intervention to promote Positive Parenting adapted to Autism (VIPP-AUTI): a randomized controlled trial. *Autism* 2015;**19**:588–603. https://doi.org/10.1177/1362361314537124

29. Yagmur S, Mesman J, Malda M, Bakermans-Kranenburg MJ, Ekmekci H. Video-feedback intervention increases sensitive parenting in ethnic minority mothers: a randomized control trial. *Attach Human Dev* 2014;**16**:371–86. https://doi.org/10.1080/14616734.2014.912489

30. Green J, Charman T, Pickles A, Wan MW, Elsabbagh M, Slonims V, *et al.* Parent-mediated intervention versus no intervention for infants at high risk of autism: a parallel, single-blind, randomised trial. *Lancet* 2015;**2**:133–40. https://doi.org/10.1016/S2215-0366(14)00091-1

31. Werner C, Vermeer HJ, Linting M, van IJzendoorn MH. Video-feedback intervention in center-based child care: a randomized controlled trial. *Early Child Res Q* 2016;**42**:93–104. https://doi.org/10.1016/j.ecresq.2017.07.005

32. Juffer F, Bakermans-Kranenburg MJ, van IJzendoorn MH. Video-feedback Intervention to Promote Positive Parenting and Sensitive Discipline (VIPP-SD): Development and Meta-Analytical Evidence of Its Effectiveness. In Steele H, Steele M, editors. *Handbook of Attachment-based Interventions*. 1st edn. New York, NY: Guildford; 2017. pp. 1–26.

33. Allen G. *Early Intervention: The Next Steps*. London: Cabinet Office and Department for Work and Pensions; 2011.

34. Black MM, Walker SP, Fernald LCH, Andersen CT, DiGirolamo AM, Lu C, *et al.* Early childhood development coming of age: science through the life course. *Lancet* 2017;**389**:77–90. https://doi.org/10.1016/S0140-6736(16)31389-7

35. Department of Health and Social Care (DHSC) and Department for Education. *Transforming Children and Young People's Mental Health Provision: A Green Paper*. London: DHSC and Department for Education; 2017.

36. Maughan B, Barker ED. The earlier the better? Pausing for thought . . . *Child Dev* 2019;**90**:20–4. https://doi.org/10.1111/cdev.13168

37. Centre on the Developing Child. *A Science-Based Framework for Early Childhood Policy: Using Evidence to Improve Outcomes in Learning, Behavior, and Health for Vulnerable Children*. National Forum on Early Childhood Program Evaluation and National Scientific Council on the Developing Child; 2007. URL: www.developingchild.harvard.edu (accessed 1 January 2020).

38. Pingault JB, Rijsdijk F, Zheng Y, Plomin R, Viding E. Developmentally dynamic genome: evidence of genetic influences on increases and decreases in conduct problems from early childhood to adolescence. *Sci Rep* 2015;**5**:10053. https://doi.org/10.1038/srep10053

39. Britto PR, Lye SJ, Proulx K, Yousafzai AK, Matthews SG, Vaivada T, *et al.* Nurturing care: promoting early childhood development. *Lancet* 2017;**389**:91–102. https://doi.org/10.1016/S0140-6736(16)31390-3

40. World Medical Association. *WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects*. 2013. URL: https://wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/ (accessed April 2021).

41. NHS Health Research Authority. *UK Policy Framework for Health and Social Care Research*. 2021. URL: https://hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/uk-policy-framework-health-social-care-research/uk-policy-framework-health-and-social-care-research/#context (accessed April 2021).

42. Goodman R. The Strengths and Difficulties Questionnaire: a research note. *J Child Psychol Psychiatry* 1997;**38**:581–6. https://doi.org/10.1111/j.1469-7610.1997.tb01545.x

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

43. Iles JE, Rosan C, Wilkinson E, Ramchandani PG. Adapting and developing a video-feedback intervention for co-parents of infants at risk of externalising behaviour problems (VIPP-Co): a feasibility study. *Clin Child Psychol Psychiatry* 2017;**22**:483–99. https://doi.org/10.1177/1359104517704025

44. Sonuga-Barke EJ, Lamparelli M, Stevenson J, Thompson M, Henry A. Behaviour problems and pre-school intellectual attainment: the associations of hyperactivity and conduct problems. *J Child Psychol Psychiatry* 1994;**35**:949–60. https://doi.org/10.1111/j.1469-7610.1994.tb02304.x

45. Taylor E, Schachar R, Thorley G, Wieselberg M. Conduct disorder and hyperactivity: I. Separation of hyperactivity and antisocial conduct in British child psychiatric patients. *Br J Psychiatry* 1986;**149**:760–7. https://doi.org/10.1192/bjp.149.6.760

46. Costello EJ, Egger H, Angold A. 10-year research update review: the epidemiology of child and adolescent psychiatric disorders: I. Methods and public health burden. *J Am Acad Child Adolesc Psychiatry* 2005;**44**:972–86. https://doi.org/10.1097/01.chi.0000172552.41596.6f

47. Daley D, O'Brien M. A small-scale randomized controlled trial of the self-help version of the New Forest Parent Training Programme for children with ADHD symptoms. *Eur Child Adolesc Psychiatry* 2013;**22**:543–52. https://doi.org/10.1007/s00787-013-0396-8

48. Sonuga-Barke EJ, Daley D, Thompson M, Laver-Bradbury C, Weeks A. Parent-based therapies for preschool attention-deficit/hyperactivity disorder: a randomized, controlled trial with a community sample. *J Am Acad Child Adolesc Psychiatry* 2001;**40**:402–8. https://doi.org/10.1097/00004583-200104000-00008

49. Sonuga-Barke EJ, Thompson M, Daley D, Laver-Bradbury C. Parent training for attention deficit/hyperactivity disorder: is it as effective when delivered as routine rather than as specialist care? *Br J Clin Psychol* 2004;**43**:449–57. https://doi.org/10.1348/0144665042388973

50. Croft S, Stride C, Maughan B, Rowe R. Validity of the strengths and difficulties questionnaire in preschool-aged children. *Pediatrics* 2015;**135**:e1210–9. https://doi.org/10.1542/peds.2014-2920

51. D'Souza S, Waldie KE, Peterson ER, Underwood L, Morton SM. Psychometric properties and normative data for the preschool Strengths and Difficulties Questionnaire in two-year-old children. *J Abnorm Child Psychol* 2017;**45**:345–57. https://doi.org/10.1007/s10802-016-0176-2

52. Achenbach TM, Rescorla L. *Manual for the ASEBA School-Age Forms & Profiles: An Integrated System of Multi-Informant Assessment*. Burlington, VT: University of Vermont Research Center for Children, Youth, and Families; 2001.

53. van Zeijl J, Mesman J, Stolk MN, Alink LR, van Ijzendoorn MH, Bakermans-Kranenburg MJ, *et al.* Terrible ones? Assessment of externalizing behaviors in infancy with the Child Behavior Checklist. *J Child Psychol Psychiatry* 2006;**47**:801–10. https://doi.org/10.1111/j.1469-7610.2006.01616.x

54. Arnold DS, O'Leary SG, Wolff LS, Acker MM. The Parenting Scale: a measure of dysfunctional parenting in discipline situations. *Psychol Assess* 1993;**5**:137. https://doi.org/10.1037/1040-3590.5.2.137

55. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001;**16**:606–13. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

56. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV*. 4th edn. Washington, DC: American Psychiatric Association; 1994.

57. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med* 2006;**166**:1092–7. https://doi.org/10.1001/archinte.166.10.1092

58. Busby DM, Christensen C, Crane DR, Larson JH. A revision of the Dyadic Adjustment Scale for use with distressed and nondistressed couples: construct hierarchy and multidimensional scales. *J Marital Fam Ther* 1995;**21**:289–308. https://doi.org/10.1111/j.1752-0606.1995.tb00163.x

59. Byford S, Cary M, Barrett B, Aldred CR, Charman T, Howlin P, *et al.* Cost-effectiveness analysis of a communication-focused therapy for pre-school children with autism: results from a randomised controlled trial. *BMC Psychiatry* 2015;**15**:316. https://doi.org/10.1186/s12888-015-0700-x

60. European Medicines Association. *Guideline for Good Clinical Practice E6(R2)*. 2016. URL: www.ema.europa.eu/en/documents/scientific-guideline/ich-e-6-r2-guideline-good-clinical-practice-step-5_en.pdf (accessed February 2021).

61. European Medicines Agency. *Guideline for Good Clinical Practice*. 2016. URL: https://ema.europa.eu/en/documents/scientific-guideline/ich-e-6-r2-guideline-good-clinical-practice-step-5_en.pdf (accessed April 2021).

62. Great Britain. *Data Protection Act 2018*. London: The Stationery Office; 2018. URL: www.legislation.gov.uk/ukpga/2018/12/contents/enacted (accessed 10 February 2021).

63. National Institute for Health Research. *Research Governance Guidelines*. 2019. URL: www.nihr.ac.uk/documents/research-governance-guidelines/12154 (accessed February 2021).

64. DAMOCLES Study Group. A proposed charter for clinical trial data monitoring committees: helping them to do their job well. *Lancet* 2005;**365**:711–22. https://doi.org/10.1016/S0140-6736(05)17965-3

65. National Institute for Health and Care Excellence (NICE). *Guide to the Methods of Technology Appraisal*. London: NICE; 2013.

66. Department of Health and Social Care (DHSC). *NHS Reference Costs 2017–2018*. London: DHSC; 2018.

67. Curtis L. *Unit Costs of Health and Social Care 2017*. Canterbury: PSSRU, University of Kent; 2017.

68. Government Digital Service. *Foster Carers: Help with the Cost of Fostering*. 2018. URL: www.gov.uk/foster-carers/help-with-the-cost-of-fostering (accessed 1 January 2020).

69. Valuation Office Agency. *Local Housing Allowance: LHA Bedroom Calculator 2018*. 2018. URL: https://lha-direct.voa.gov.uk/BedroomCalculator.aspx (accessed 1 January 2020).

70. Joint Formulary Committee. *British National Formulary*. London: BMJ Group and Pharmaceutical Press. URL: www.medicinescomplete.com (accessed 1 January 2020).

71. Netten A, Knight J, Dennett J, Cooley R, Slight A. *A 'Ready Reckoner' for Staff Costs in the NHS. Volume II. Methodology*. Canterbury: PSSRU, University of Kent; 1998.

72. Thompson SG, Barber JA. How should cost data in pragmatic randomised trials be analysed? *BMJ* 2000;**320**:1197–200. https://doi.org/10.1136/bmj.320.7243.1197

73. van Hout BA, Al MJ, Gordon GS, Rutten FF. Costs, effects and C/E-ratios alongside a clinical trial. *Health Econ* 1994;**3**:309–19. https://doi.org/10.1002/hec.4730030505

74. Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Med Decis Making* 1998;**18**(Suppl. 2):68–80. https://doi.org/10.1177/0272989X98018002S09

75. Barber JA, Thompson SG. Analysis of cost data in randomized trials: an application of the non-parametric bootstrap. *Stat Med* 2000;**19**:3219–36. https://doi.org/10.1002/1097-0258(20001215)19:23<3219::AID-SIM623>3.0.CO;2-P

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

76. Fenwick E, Byford S. A guide to cost-effectiveness acceptability curves. *Br J Psychiatry* 2005;**187**:106–8. https://doi.org/10.1192/bjp.187.2.106

77. Weichle T, Hynes DM, Durazo-Arvizu R, Tarlov E, Zhang Q. Impact of alternative approaches to assess outlying and influential observations on health care costs. *Springerplus* 2013;**2**:614. https://doi.org/10.1186/2193-1801-2-614

78. Sainsbury Centre for Mental Health. *The Chance of a Lifetime: Preventing Early Conduct Problems and Reducing Crime*. London: Sainsbury Centre for Mental Health; 2009.

79. Scott S, Knapp M, Henderson J, Maughan B. Financial cost of social exclusion: follow up study of antisocial children into adulthood. *BMJ* 2001;**323**:191. https://doi.org/10.1136/bmj.323.7306.191

80. Stevens KJ. Working with children to develop dimensions for a preference-based, generic, pediatric, health-related quality-of-life measure. *Qual Health Res* 2010;**20**:340–51. https://doi.org/10.1177/1049732309358328

81. Furber G, Segal L, Leach M, Cocks J. Mapping scores from the Strengths and Difficulties Questionnaire (SDQ) to preference-based utility values. *Qual Life Res* 2014;**23**:403–11. https://doi.org/10.1007/s11136-013-0494-6

82. Manca A, Hawkins N, Sculpher MJ. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. *Health Econ* 2005;**14**:487–96. https://doi.org/10.1002/hec.944

83. Ahern S, Burke LA, McElroy B, Corcoran P, McMahon EM, Keeley H, *et al.* A cost-effectiveness analysis of school-based suicide prevention programmes. *Eur Child Adolesc Psychiatry* 2018;**27**:1295–304. https://doi.org/10.1007/s00787-018-1120-5

84. Deidda M, Boyd KA, Minnis H, Donaldson J, Brown K, Boyer NR, *et al.* Protocol for the economic evaluation of a complex intervention to improve the mental health of maltreated infants and children in foster care in the UK (The BeST? services trial). *BMJ Open* 2018;**8**:e020066. https://doi.org/10.1136/bmjopen-2017-020066

85. Rocks S, Stepney M, Glogowska M, Fazel M, Tsiachristas A. Understanding and evaluating new models of child and adolescent mental health services in south-east England: a study protocol for an observational mixed-methods study. *BMJ Open* 2018;**8**:e024230. https://doi.org/10.1136/bmjopen-2018-024230

86. Wailoo AJ, Hernandez-Alava M, Manca A, Mejia A, Ray J, Crawford B, *et al.* Mapping to estimate health-state utility from non-preference-based outcome measures: an ISPOR Good Practices for Outcomes Research Task Force report. *Value Health* 2017;**20**:18–27. https://doi.org/10.1016/j.jval.2016.11.006

87. Ara R, Rowen D, Mukuria C. The use of mapping to estimate health state utility values. *PharmacoEconomics* 2017;**35**:57–66. https://doi.org/10.1007/s40273-017-0548-7

88. McCabe C, Edlin R, Meads D, Brown C, Kharroubi S. Constructing indirect utility models: some observations on the principles and practice of mapping to obtain health state utilities. *PharmacoEconomics* 2013;**31**:635–41. https://doi.org/10.1007/s40273-013-0071-4

89. Round J, Hawton A. Statistical alchemy: conceptual validity and mapping to generate health state utility values. *Pharmacoecon Open* 2017;**1**:233–9. https://doi.org/10.1007/s41669-017-0027-2

90. Jonsson U, Alaie I, Löfgren Wilteus A, Zander E, Marschik PB, Coghill D, Bölte S. Annual Research Review: quality of life and childhood mental and behavioural disorders – a critical review of the research. *J Child Psychol Psychiatry* 2017;**58**:439–69. https://doi.org/10.1111/jcpp.12645

91. Ravens-Sieberer U, Erhart M, Wille N, Wetzel R, Nickel J, Bullinger M. Generic health-related quality-of-life assessment in children and adolescents: methodological considerations. *PharmacoEconomics* 2006;**24**:1199–220. https://doi.org/10.2165/00019053-200624120-00005

92. Mukuria C, Rowen D, Harnan S, Rawdin A, Wong R, Ara R, Brazier J. An updated systematic review of studies mapping (or cross-walking) measures of health-related quality of life to generic preference-based measures to generate utility values. *Appl Health Econ Health Policy* 2019;**17**:295–313. https://doi.org/10.1007/s40258-019-00467-6

93. Sharma R, Gu Y, Sinha K, Aghdaee M, Parkinson B. Mapping the Strengths and Difficulties Questionnaire onto the Child Health Utility 9D in a large study of children. *Qual Life Res* 2019;**28**:2429–41. https://doi.org/10.1007/s11136-019-02220-x

94. Olazagasti MAR, Klein RG, Mannuzza S, Belsky ER, Hutchison JA, Lashua-Shriftman EC, *et al.* Does childhood attention-deficit/hyperactivity disorder predict risk-taking and medical illnesses in adulthood? *J Am Acad Child Adolesc Psychiatry* 2013;**52**:153–62. https://doi.org/10.1016/j.jaac.2012.11.012

95. Bakermans-Kranenburg MJ, Van IJzendoorn MH, Pijlman FT, Mesman J, Juffer F. Experimental evidence for differential susceptibility: dopamine D4 receptor polymorphism (DRD4 VNTR) moderates intervention effects on toddlers' externalizing behavior in a randomized controlled trial. *Dev Psychol* 2008;**44**:293–300. https://doi.org/10.1037/0012-1649.44.1.293

96. Belsky J, Bakermans-Kranenburg MJ, van IJzendoorn MH. For better and for worse: differential susceptibility to environmental influences. *Curr Dir Psychol Sci* 2007;**16**:300–4. https://doi.org/10.1111/j.1467-8721.2007.00525.x

97. Klein Velderman M, Bakermans-Kranenburg MJ, Juffer F, van IJzendoorn MH. Effects of attachment-based interventions on maternal sensitivity and infant attachment: differential susceptibility of highly reactive infants. *J Fam Psychol* 2006;**20**:266–74. https://doi.org/10.1037/0893-3200.20.2.266

98. Wakschlag LS, Choi SW, Carter AS, Hullsiek H, Burns J, McCarthy K, *et al.* Defining the developmental parameters of temper loss in early childhood: implications for developmental psychopathology. *J Child Psychol Psychiatry* 2012;**53**:1099–108. https://doi.org/10.1111/j.1469-7610.2012.02595.x

99. Goodyer IM, Reynolds S, Barrett B, Byford S, Dubicka B, Hill J, *et al.* Cognitive behavioural therapy and short-term psychoanalytical psychotherapy versus a brief psychosocial intervention in adolescents with unipolar major depressive disorder (IMPACT): a multicentre, pragmatic, observer-blind, randomised controlled superiority trial. *Lancet Psychiatry* 2017;**4**:109–19. https://doi.org/10.1016/S2215-0366(16)30378-9

100. Robling M, Bekkers MJ, Bell K, Butler CC, Cannings-John R, Channon S, *et al.* Effectiveness of a nurse-led intensive home-visitation programme for first-time teenage mothers (Building Blocks): a pragmatic randomised controlled trial. *Lancet* 2016;**387**:146–55. https://doi.org/10.1016/S0140-6736(15)00392-X

101. Cottrell DJ, Wright-Hughes A, Collinson M, Boston P, Eisler I, Fortune S, *et al.* Effectiveness of systemic family therapy versus treatment as usual for young people after self-harm: a pragmatic, phase 3, multicentre, randomised controlled trial. *Lancet Psychiatry* 2018;**5**:203–16. https://doi.org/10.1016/S2215-0366(18)30058-0

102. Doyle O, Delaney L, O'Farrelly C, Fitzpatrick N, Daly M. Can early intervention improve maternal well-being? Evidence from a randomized controlled trial. *PLOS ONE* 2017;**12**:e0169829. https://doi.org/10.1371/journal.pone.0169829

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

103. Asmussen K, Feinstein L, Martin J, Chowdry H. *Foundations for Life: What Works to Support Parent Child Interaction in the Early Years*. London: Early Intervention Foundation; 2016.

104. Edwards RT, Jones C, Berry V, Charles J, Linck P, Bywater T, *et al*. Incredible years parenting programme: cost-effectiveness and implementation. *J Child Serv* 2016;**11**:54–72. https://doi.org/10.1108/JCS-02-2015-0005

105. Office for National Statistics. *Local Area Analysis of Qualifications Across England and Wales*. 2014. URL: https://webarchive.nationalarchives.gov.uk/20160105191238/http://www.ons.gov.uk/ons/rel/census/2011-census-analysis/local-area-analysis-of-qualifications-across-england-and-wales/rpt—local-area-analysis-of-qualifications-across-england-and-wales.html (accessed April 2021).

106. Tappenden P, Chilcott J, Brennan A, Squires H, Stevenson M. Whole disease modeling to inform resource allocation decisions in cancer: a methodological framework. *Value Health* 2012;**15**:1127–36. https://doi.org/10.1016/j.jval.2012.07.008

107. Chen W, Taylor E. Parental Account of Children's Symptoms (PACS), ADHD Phenotypes and its Application to Molecular Genetic Studies. In Oades RD, editor. *Attention-Deficit/Hyperactivity Disorder and the Hyperkinetic Syndrome: Current Ideas and Ways Forward*. Hauppauge, NY: Nova Science Publishers; 2006. pp. 3–20.

Appendix 1 Primary outcome: Preschool Parental Account of Children's Symptoms

More information about the measure is available on request from the corresponding author. Additional detail is also available in Chen and Taylor.[107]

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

Appendix 2 Additional detail regarding multiple imputation analysis

There is no information on treatment of missing values in the manuals for PPACS, our primary outcome measure. A multiple imputation was used on the individual questions in each questionnaire at baseline and at follow-up to optimise our treatment of missing data. Any child who has a PPACS at follow-up with at least 50% of the questions completed was included in this multiple imputation PPACS outcome analysis. We based our imputation on completed questions and subscale scores from the PPACS interview at outcome and total scores at earlier time points. For multiple imputation at baseline, we based our imputation on completed questions and subscale scores from the PPACS interview at baseline only. The primary analysis also incorporated multiple imputation for anyone whose PPACS at outcome was completely missing (or had fewer items completed than the threshold). At the 5-month follow-up, these were imputed based on randomised group, sex of child, age of child at the 5-month assessment, baseline PPACS, CBCL and SDQ scores. At the 2-year follow-up, these were imputed based on randomised group, sex of child, age of child at the 2-year assessment, and PPACS, CBCL and SDQ scores (at the 5-month follow-up, if available, or else using these PPACS, CBCL and SDQ scores at baseline).

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

Appendix 3 Participant recruitment by month from trial commencement to recruitment target (n = 300)

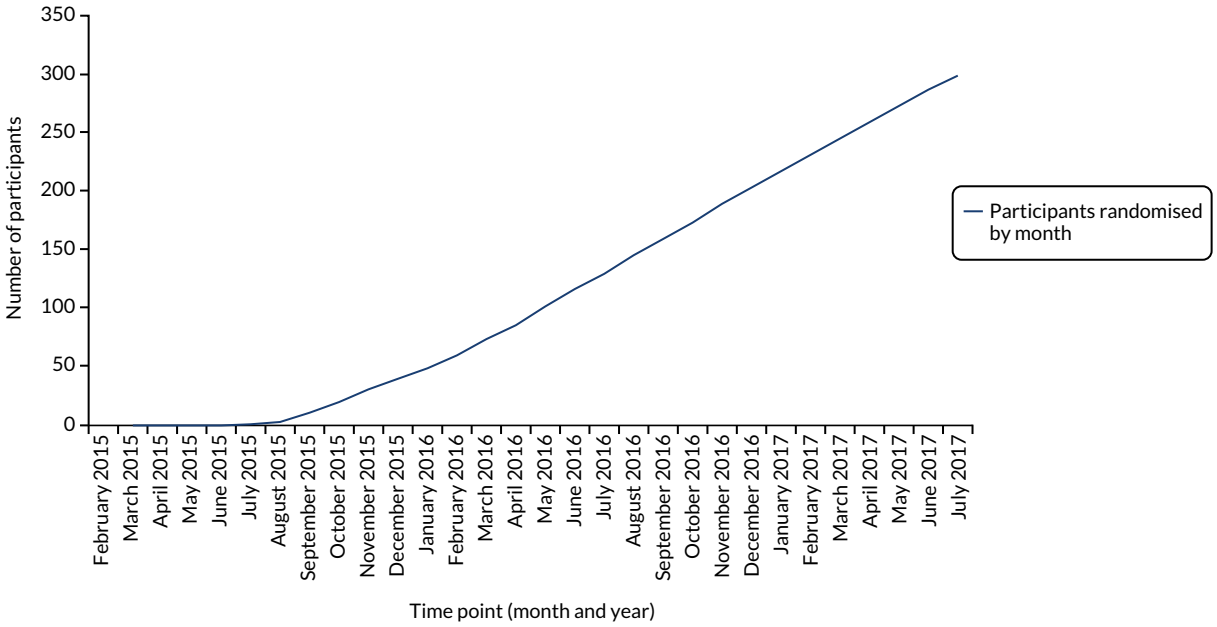


FIGURE 12 Participant recruitment by month from trial commencement to recruitment target.

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

Appendix 4 Secondary caregiver analysis

TABLE 24 Baseline characteristics of secondary caregivers by treatment allocation

Characteristic	Participating		Not participating	
	VIPP-SD (*N* = 26)	Usual care (*N* = 25)	VIPP-SD (*N* = 117)	Usual care (*N* = 115)
Sex (male), *n* (%)	23 (88)	20 (80)	111 (95)	110 (96)
Age (years), mean (SD)	35.4 (6.4)	38.3 (8.9)	37.1 (7.2)	37.6 (6.9)
Parental status*, n *(%)				
Parent (including step or adoptive)	26 (100)	23 (92)	116 (99.1)	113 (98.3)
Other	0 (0)	2 (8)	1 (0.9)	2 (1.7)
Ethnicity*, n *(%)				
Asian	3 (12)	2 (8)	12 (10)	10 (9)
Black	1 (4)	2 (8)	9 (8)	13 (11)
Mixed ethnicity	2 (8)	1 (4)	6 (5)	3 (3)
White	19 (73)	17 (68)	85 (73)	84 (73)
Other	1 (4)	3 (12)	5 (4)	5 (4)
Relationship status*, n *(%)				
Married/civil partnership/cohabiting	25 (96)	24 (96)	103 (88)	104 (90)
Divorced/widowed/legally separated	0 (0)	1 (4)	1 (1)	2 (2)
Single and none of the above	0 (0)	0 (0)	5 (4)	8 (7)
In a relationship but not cohabiting	1 (4)	0 (0)	8 (7)	1 (1)
Employment status*, n *(%)				
Employed	26 (100)	16 (64)	92 (79)	88 (77)
Paid parental leave	0 (0)	0 (0)	0 (0)	0 (0)
Self-employed	0 (0)	5 (20)	20 (17)	23 (20)
Full-time student	0 (0)	0 (0)	0 (0)	0 (0)
Looking after home and children	0 (0)	4 (16)	5 (4)	4 (3)
Highest qualification*,[a] n *(%)				
GCSE or lower	2 (8)	3 (12)	22 (19)	19 (17)
A level/NVQ/BTEC	4 (15)	5 (20)	26 (22)	23 (20)
Graduate	20 (77)	17 (68)	68 (59)	72 (63)

A level, Advanced level; BTEC, Business and Technology Education Council; GCSE, General Certificate of Secondary Education; NVQ, National Vocational Qualification.
a Data were missing for one participant in each group for non-participating caregivers for this variable.

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

TABLE 25 Primary ITT analysis of adjusted mean difference on secondary outcomes of children's behaviour problems (total scores) at the 5- and 24-month follow-ups reported by participating secondary caregivers

	Trial group						
	VIPP-SD		Usual care				
Outcome	*n*	Mean (SD)	*n*	Mean (SD)	Mean difference[a] (95% CI)	Effect size[b] (95% CI)	*p*-value[a]
CBCL[c] total score – secondary caregiver report							
Baseline	26	43.8 (23.6)	25	37.1 (15.2)			
5 months	24	35.3 (19.7)	24	31.4 (14.2)	−5.2 (−13.3 to 3.0)	−0.36 (−0.94 to 0.21)	0.21
24 months	23	33.0 (23.1)	24	31.5 (16.4)	1.61 (−7.78 to 10.99)	0.10 (−0.48 to 0.67)	0.73
SDQ[c] total score – secondary caregiver report							
Baseline	26	13.9 (5.0)	25	11.3 (5.7)			
5 months	24	11.2 (4.7)	24	10.5 (3.7)	−0.78 (−2.89 to 1.33)	−0.21 (−0.78 to 0.36)	0.46
24 months	23	10.3 (6.0)	24	10.6 (5.6)	2.32 (−0.78 to 5.41)	0.41 (−0.14 to 0.96)	0.14

a Difference in mean is difference between treatment groups from linear regression analysis of the outcome measure on the baseline score of that same measurement, on treatment centre, on randomised group, on length of follow-up, on age of child and on number of parents/caregivers participating (all treated as fixed effects). Positive differences represent greater adjusted decreases in symptoms in the VIPP than in the usual-care group.
b Effect size is standardised difference in mean (adjusted as above). This is the difference in mean divided by the SD of controls at follow-up.
c Higher scores indicate higher problems.

TABLE 26 Complier-average causal effects[a] analysis on secondary outcomes of child behaviour reported by participating secondary caregivers

	Trial group						
	VIPP-SD		Usual care				
Outcome	*n*	Mean (SD)	*n*	Mean (SD)	Mean difference[b] (95% CI)	Effect size[c] (95% CI)	*p*-value
CBCL[d] – secondary caregiver reported							
Baseline	26	43.8 (23.6)	25	37.1 (15.2)			
5 months	24	35.3 (19.7)	24	31.4 (14.2)	−8.16 (−17.2 to 0.90)	−0.58 (−1.21 to 0.06)	0.08
24 months[e]	23	33.0 (23.1)	24	31.5 (16.4)	1.61 (−7.78 to 10.99)	0.10 (−0.48 to 0.67)	0.73
SDQ[d] – secondary caregiver reported							
Baseline	26	13.9 (5.0)	25	11.3 (5.7)			
5 months	24	11.2 (4.7)	24	10.5 (3.7)	−0.87 (−2.89 to 1.15)	−0.24 (−0.78 to 0.31)	0.40
24 months[e]	23	10.3 (6.0)	24	10.6 (5.6)	2.32 (−0.78 to 5.41)	0.41 (−0.14 to 0.96)	0.14

a CACE results are based on the assumption that there is no effect of being randomised to VIPP-SD, or of receiving one, two or three VIPP-SD visits. Receiving four, five or six VIPP-SD visits counts as receiving the intervention.
b Difference in mean is difference between treatment groups from linear regression analysis of the outcome measure on the baseline score of that same measurement, on treatment centre, on randomised group, on length of follow-up, on age of child and on number of parents/caregivers participating (all treated as fixed effects). Positive differences represent greater adjusted decreases in symptoms in the VIPP-SD than in the usual-care group.
c Effect size is standardised difference in mean (adjusted as above). This is the difference in mean divided by the SD of controls at follow-up.
d Higher scores represent higher levels of behaviour problems.
e At the 24-month follow-up, all families with two participating caregivers who were allocated to the VIPP-SD group received the intervention in practice. Hence, these CACE results are identical to the ITT results.

TABLE 27 Secondary outcomes on child behaviour according to the perception of male and female caregivers: ITT analysis

Outcome	Trial group: VIPP-SD *n*	VIPP-SD Mean (SD)	Usual care *n*	Usual care Mean (SD)	Mean difference[a] (95% CI)	Effect size[b] (95% CI)	*p*-value
CBCL[c] total score: female caregiver reported							
Baseline	146	41.2 (21.6)	147	42.6 (21.4)			
5 months	135	32.7 (20.7)	143	37.8 (21.1)	3.97 (0.57 to 7.36)	0.19 (0.03 to 0.35)	0.02
24 months	136	30.9 (23.8)	142	36.1 (23.8)	3.62 (−1.11 to 8.35)	0.15 (−0.05 to 0.35)	0.13
CBCL[c] total score: male caregiver reported							
Baseline	31	40.8 (23.6)	25	37.8 (13.8)			
5 months	29	33.6 (19.2)	24	27.5 (11.9)	−7.71 (−14.83 to -0.59)	−0.65 (−1.25 to -0.05)	0.03
24 months	28	31.1 (21.1)	24	25.6 (11.5)	−4.75 (−12.88 to 3.37)	−0.41 (−1.12 to 0.29)	0.24
SDQ[d] total score: female caregiver reported							
Baseline	145	14 (4.7)	147	14.1 (4.7)			
5 months	135	11.4 (5.1)	143	12.3 (5.2)	1.05 (0.06 to 2.05)	0.20 (0.01 to 0.39)	0.04
24 months	136	10.4 (5.5)	142	11.1 (5.9)	0.59 (−0.56 to 1.75)	0.10 (−0.09 to 0.3)	0.31
SDQ[d] total score: male caregiver reported							
Baseline	31	13.3 (5.2)	25	10.9 (5.4)			
5 months	29	11.0 (4.7)	24	9.8 (3.4)	−1.37 (−3.41 to 0.66)	−0.41 (−1.01 to 0.2)	0.18
24 months	28	10.4 (5.4)	24	8.8 (3.6)	−0.01 (−2.53 to 2.52)	0.00 (−0.71 to 0.71)	1.00
CBCL[c] externalising subscale: female caregiver reported							
Baseline	146	16.8 (8)	147	17.5 (8.5)			
5 months	135	13.4 (8.2)	143	15.4 (8.6)	1.48 (−0.03 to 2.98)	0.17 (0.00 to 0.35)	0.05
24 months	136	11.7 (8.9)	142	13.6 (9.7)	1.16 (−0.71 to 3.02)	0.12 (−0.07 to 0.31)	0.22
CBCL[c] externalising subscale: male caregiver reported							
Baseline	31	16.1 (8.8)	25	14.9 (6.3)			
5 months	29	13.6 (8.9)	24	11.5 (4.6)	−2.90 (−6.55 to 0.75)	−0.64 (−1.44 to 0.16)	0.12
24 months	28	11.8 (8.8)	24	10.5 (6.0)	−0.96 (−5.09 to 3.17)	−0.16 (−0.85 to 0.53)	0.64
CBCL[c] internalising subscale: female caregiver reported							
Baseline	146	9.1 (7.2)	147	9.6 (7.5)			
5 months	135	7.3 (6.5)	143	8.6 (6.8)	1.14 (−0.02 to 2.30)	0.17 (0.00 to 0.34)	0.05
24 months	136	8.3 (8.4)	142	9.2 (7.6)	0.49 (−1.15 to 2.13)	0.06 (−0.15 to 0.28)	0.55
CBCL[c] internalising subscale: male caregiver reported							
Baseline	31	8.4 (7.4)	25	8.1 (5.4)			
5 months	29	7.0 (5.4)	24	5.7 (5.2)	−1.23 (−3.98 to 1.51)	−0.24 (−0.77 to 0.29)	0.37
24 months	28	7.9 (7.6)	24	5.6 (3.5)	−2.23 (−4.90 to 0.44)	−0.63 (−1.38 to 0.12)	0.10

continued

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

TABLE 27 Secondary outcomes on child behaviour according to the perception of male and female caregivers: ITT analysis (*continued*)

Outcome	Trial group				Mean difference[a] (95% CI)	Effect size[b] (95% CI)	*p*-value
	VIPP-SD		Usual care				
	n	Mean (SD)	*n*	Mean (SD)			
SDQ[d] externalising subscale: female caregiver reported							
Baseline	145	9.5 (2.9)	147	9.5 (3.5)			
5 months	135	7.6 (3.4)	143	8.3 (3.8)	0.76 (0.06 to 1.46)	0.20 (0.02 to 0.39)	0.03
24 months	136	7.0 (3.6)	142	7.6 (4.2)	0.56 (−0.23 to 1.35)	0.13 (−0.05 to 0.32)	0.16
SDQ[d] externalising subscale: male caregiver reported							
Baseline	31	9.0 (3.5)	25	7.6 (3.1)			
5 months	29	7.4 (3.6)	24	6.8 (2.8)	−0.89 (−2.59 to 0.81)	−0.32 (−0.92 to 0.29)	0.30
24 months	28	6.6 (3.6)	24	6.1 (3.4)	0.22 (−1.93 to 2.37)	0.07 (−0.58 to 0.71)	0.84

a Difference in mean is difference between treatment groups from linear regression analysis of the outcome measure on the baseline score of that same measurement, on treatment centre, on randomised group, on length of follow-up, on age of child and on number of parents/caregivers participating (all treated as fixed effects). Positive differences represent greater adjusted decreases in symptoms in the VIPP-SD than in the usual-care group, with the exception of the CBCL prosocial scale (see footnote c).
b Effect size is standardised difference in mean (adjusted as above). This is the difference in mean divided by the SD of controls at follow-up.
c Higher scores on all scales represent higher levels of behaviour problems, with the exception of the prosocial scale, where higher scores indicate lower levels of problems.
d Higher scores represent higher levels of behaviour problems.

Appendix 5 Supplementary analysis

TABLE 28 Sensitivity analyses for lack of adjustment for length of follow-up: primary and secondary outcomes on child behaviour – ITT analysis

	Trial group						
	VIPP-SD		Usual care				
Outcome	*n*	Mean (SD)	*n*	Mean (SD)	Mean difference[a] (95% CI)	Effect size[b] (95% CI)	*p*-value
PPACS[c]							
Baseline	151	33.5 (9.0)	149	32.4 (10.6)			
5 months	140	28.8 (9.2)	146	30.3 (9.9)	2.02 (0.24 to 3.80)	0.20 (0.02 to 0.38)	0.03
24 months	140	23.1 (8.2)	142	24.7 (9.9)	1.94 (–0.05 to 3.92)	0.20 (–0.01 to 0.4)	0.06
CBCL[c]							
Baseline	151	40.7 (21.7)	149	42.7 (21.1)			
5 months	140	32.5 (20.6)	145	37.2 (21.0)	2.48 (–0.5 to 5.46)	0.12 (–0.02 to 0.26)	0.10
24 months	141	30.6 (23.4)	144	35.3 (23.7)	3.29 (–1.34 to 7.91)	0.14 (–0.06 to 0.33)	0.16
SDQ[c]							
Baseline	150	13.8 (4.8)	149	14.0 (4.7)			
5 months	140	11.3 (5.1)	145	12.2 (5.2)	0.72 (–0.15 to 1.6)	0.14 (–0.03 to 0.31)	0.10
24 months	141	10.4 (5.4)	144	10.9 (5.8)	0.38 (–0.73 to 1.5)	0.07 (–0.13 to 0.26)	0.50

a Difference in mean is difference between treatment groups from linear regression analysis of the outcome measure on the baseline score of that same measurement, on treatment centre, on randomised group, on age of child and on number of parents/caregivers participating (all treated as fixed effects). Positive differences represent greater decreases in the VIPP-SD than in the usual-care group.
b Effect size is standardised difference in mean (adjusted as above). This is the difference in mean divided by the SD of controls at follow-up.
c Higher scores represent higher behaviour problems.

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

TABLE 29 Baseline characteristics of participating children and primary caregivers with missing PPACS data by treatment allocation

Characteristic	Trial group					
	VIPP-SD			Usual care		
	No missing data (*N* = 133)	Missing at 5 months (*N* = 11)	Missing at 24 months (*N* = 11)	No missing data (*N* = 141)	Missing at 5 months (*N* = 3)	Missing at 24 months (*N* = 7)
Child characteristic						
Sex (male), *n* (%)	67 (50)	4 (36)	7 (64)	60 (43)	1 (33)	2 (29)
Age (months), mean (SD)	22.5 (6.9)	25.1 (5.3)	24.6 (6.0)	23.1 (6.5)	24.0 (4.4)	24.6 (6.1)
Primary caregiver characteristic						
Sex (male), *n* (%)	8 (6)	0 (0)	0 (0)	5 (4)	0 (0)	0 (0)
Age (years), mean (SD)	34.4 (5.3)	30.1 (5.4)	26.1 (4.5)	34.8 (5.9)	30.0 (6.1)	34.4 (5.5)
Ethnicity, n *(%)*						
Asian	14 (11)	1 (9)	0 (0)	14 (10)	1 (33)	1 (14)
Black	3 (2)	0 (0)	0 (0)	15 (11)	0 (0)	0 (0)
Mixed ethnicity	10 (8)	1 (9)	0 (0)	10 (7)	0 (0)	1 (14)
White	98 (74)	9 (82)	11 (100)	99 (70)	1 (33)	4 (57)
Other	8 (6)	0 (0)	0 (0)	3 (2)	1 (33)	1 (14)
Relationship status, n *(%)*						
Married/civil partnership/cohabiting	115 (86)	8 (73)	7 (64)	121 (86)	3 (100)	5 (71)
Divorced/widowed/legally separated	1 (1)	0 (0)	0 (0)	4 (3)	0 (0)	0 (0)
Single and none of the above	10 (8)	2 (18)	2 (18)	15 (11)	0 (0)	2 (29)
In a relationship but not cohabiting	7 (5)	1 (9)	2 (18)	1 (1)	0 (0)	0 (0)
Employment status, n *(%)*						
Employed	60 (45)	4 (36)	4 (36)	61 (43)	2 (67)	3 (43)
Paid parental leave	6 (5)	0 (0)	0 (0)	10 (7)	0 (0)	0 (0)
Self-employed	18 (14)	2 (18)	0 (0)	11 (8)	0 (0)	1 (14)
Student	3 (2)	0 (0)	0 (0)	7 (5)	0 (0)	0 (0)
Looking after home and children	46 (35)	5 (45)	7 (64)	52 (37)	1 (33)	3 (43)
Highest qualification, n *(%)*						
GCSE or lower	12 (9)	3 (27)	3 (27)	13 (9)	0 (0)	1 (14)
A level/NVQ/BTEC	34 (26)	5 (45)	4 (36)	32 (23)	1 (33)	4 (57)
Graduate	87 (65)	3 (27)	4 (36)	96 (68)	2 (67)	2 (29)

A level, Advanced level; BTEC, Business and Technology Education Council; GCSE, General Certificate of Secondary Education; NVQ, National Vocational Qualification.

TABLE 30 Baseline and 5- and 24-month follow-up values of outcome measures in those with missing PPACS data at follow-up

Outcome	Trial group, mean (SD)					
	VIPP-SD			Usual care		
	No missing data ($n = 133$)	Missing at 5 months ($n = 11$)	Missing at 24 months ($n = 11$)	No missing data ($n = 141$)	Missing at 5 months ($n = 3$)	Missing at 24 months ($n = 7$)
PPACS[a]						
Baseline	33.2 (9)	36.5 (8.7)	37.9 (10.8)	32 (10.5)	37.3 (5.1)	36.7 (15.9)
5-month follow-up	28.5 (9)		34.9 (12.1)	30.4 (9.9)		27.5 (12.8)
24-month follow-up	23.2 (8.3)	21.5 (7.9)		24.6 (9.9)	30 (–)	
CBCL[a]						
Baseline	39.8 (21.2)	52.1 (27.7)	47.1 (27.2)	42.5 (21.1)	40.7 (12.3)	46.0 (23.8)
5-month follow-up	32.4 (20.6)		34.3 (22.5)	37.1 (21)		43.3 (21.7)
24-month follow-up	30.6 (23.7)	30.3 (20.2)	23.2 (–)	35.2 (23.7)	25.6 (13.6)	41.5 (36.1)
SDQ[a]						
Baseline	13.7 (4.8)	14.9 (5.2)	16.3 (5.0)	14.0 (4.7)	12.2 (4.2)	14.4 (4.9)
5-month follow-up	11.2 (5.1)		14 (5.2)	12.2 (5.3)		13.5 (3.0)
24-month follow-up	10.4 (5.5)	10.3 (4.7)	12 (–)	10.9 (5.7)	7.0 (1.4)	16.0 (11.3)
Parenting Scale[b]						
Baseline	3.0 (0.5)	2.8 (0.7)	2.7 (0.4)	3.0 (0.6)	2.7 (0.5)	2.9 (0.5)
5-month follow-up	2.9 (0.5)		2.8 (0.5)	2.9 (0.6)		2.6 (0.2)
24-month follow-up	3.0 (0.5)	3.1 (0.7)	2.8 (–)	3.0 (0.6)	2.8 (0.4)	3.7 (0.8)
PHQ-9[c]						
Baseline	4.1 (3.8)	5.1 (4.8)	6.4 (4.8)	3.9 (3.6)	5.3 (6.8)	10.9 (10.4)
5-month follow-up	3.9 (4.4)		4.3 (4.9)	4.0 (4.4)		11.6 (8.0)
24-month follow-up	4.1 (4.6)	2.3 (3.0)	5.0 (–)	3.9 (4.0)	3.0 (1.4)	11.0 (12.7)
GAD-7[c]						
Baseline	4.9 (4.4)	3.6 (4.2)	4.4 (3.8)	4.7 (4.2)	5.3 (5.0)	5.3 (4.4)
5-month follow-up	4.2 (4.4)		4.1 (4.7)	3.9 (4.0)		6.4 (4.8)
24-month follow-up	4.2 (4.7)	1.9 (2.1)	5.0 (–)	4.1 (4.0)	5.0 (2.8)	11.0 (5.7)
RDAS[d]						
Baseline	49.8 (8.1)	45.8 (6.9)	44.9 (10.1)	50.4 (8.9)	59.4 (2.2)	48.2 (16.1)
5-month follow-up	49.3 (9.4)		51.0 (9.1)	49.9 (9.7)		43.7 (15.5)
24-month follow-up	50 (8.2)	54.2 (6.4)	48.0 (–)	50.3 (7.2)	56.0 (8.5)	50.0 (–)

a Higher scores represent higher behaviour problems.
b Higher scores indicate more ineffective parenting strategies.
c Higher scores represent higher levels of behaviour problems.
d Higher scores indicate greater relationship satisfaction and lower scores indicate greater relationship distress.

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

Appendix 6 Subgroup analysis

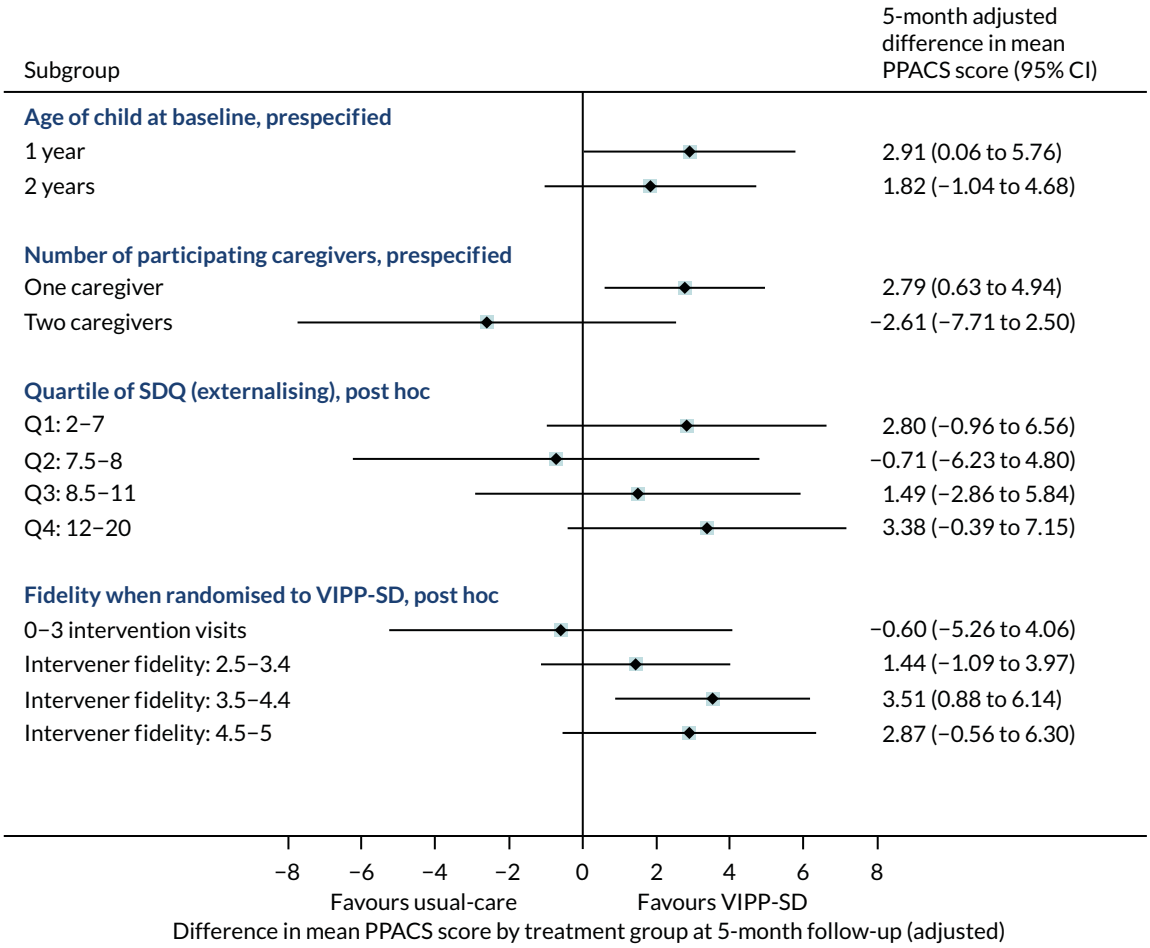


FIGURE 13 Subgroup analysis for baseline behaviour based on quartiles of externalising scores on the SDQ at the 5-month follow-up.

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

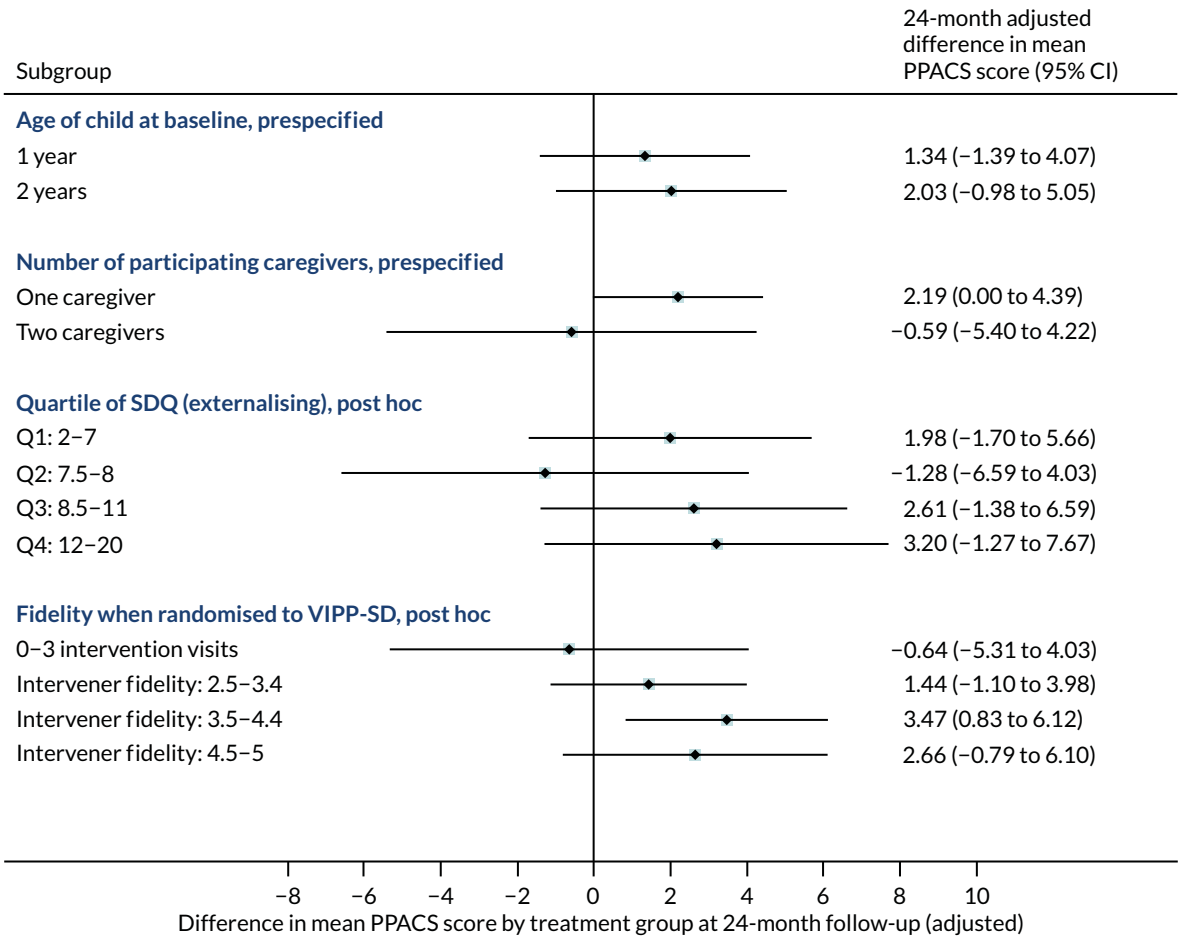


FIGURE 14 Subgroup analysis for baseline behaviour based on quartiles of externalising scores on the SDQ at the 24-month follow-up.

Appendix 7 Breakdown of protocol deviations and violations

The primary reason for protocol deviations during the trial was missing data. This happened when families were able to complete only part of an assessment (e.g. because it was carried out over the telephone), the parent/child did not consent/assent to specific parts of data collection, or, in a small number of cases, faulty equipment meant that the data could not be collected. Further to this, two enrolment errors occurred when participants were erroneously enrolled under the incorrect NHS site. In each case, the case was terminated on the eCRF, the participant was enrolled to the correct site and they were assigned a new participant ID. Another enrolment error related to a family being enrolled as having one participating caregiver when there were, in fact, two participating caregivers in the family.

Protocol violations mostly occurred when participants allocated to the treatment group did not take up any VIPP-SD sessions. Further to this, two protocol violations occurred when participants met the trial's exclusion criteria following enrolment to the trial. One family (allocated to the usual-care group) reported receiving six sessions of an individualised video-feedback intervention during the study. Another family was participating in active family court proceedings at the 24-month time point. *Table 31* shows a breakdown of all protocol deviations and violations.

TABLE 31 Occurrence of protocol deviations and violations

Deviation/violation type	*n*
Protocol deviations	
Participant had missing data	45
Participant did not complete one or more follow-up assessments	28
Enrolment error	3
Protocol violations	
Participant did not take up any VIPP-SD treatment	12
Participant met exclusion criteria following enrolment	2

© 2021 O'Farrelly *et al.* This work was produced by O'Farrelly *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaption in any medium and for any purpose provided that it is properly attributed. See: https://creativecommons.org/licenses/by/4.0/. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

EME
HS&DR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

This report presents independent research funded by the National Institute for Health Research (NIHR). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care

Published by the NIHR Journals Library