CLINICAL RESEARCH ARTICLE

Behavioural and cognitive outcomes following an early stress-reduction intervention for very preterm and extremely preterm infants

Jeannette Milgrom[1,2], Paul R. Martin[3], Carol Newnham[1], Christopher J. Holt[1,4], Peter J. Anderson[5,6], Rod W. Hunt[6,7,8], John Reece[4], Carmel Ferretti[1], Thomas Achenbach[9] and Alan W. Gemmill[1]

BACKGROUND: The landmark findings of the Mother–Infant Transaction Program (MITP) showing improved neurodevelopment of preterm infants following parent-sensitivity training delivered in the neonatal intensive care unit have not been consistently replicated. This study evaluated an MITP-type intervention in terms of neurobehavioural development to preschool age.
METHODS: A randomised controlled trial involved 123 very preterm and extremely preterm infants allocated to either a parent-sensitivity intervention (PremieStart, $n = 60$) or to standard care ($n = 63$). When children were 2 and 4.5 years corrected age, parents completed the Child Behavior Checklist (CBCL). General development was assessed at 2 years with the Bayley Scales of Infant Development (Bayley-III). At 4.5 years, cognitive functioning was assessed with the Wechsler Preschool and Primary Scale of Intelligence (WPPSI-III) and executive functioning with the NEPSY-II.
RESULTS: There were no significant between-group differences in behaviour problems at 2 or 4.5 years, general development at 2 years, or cognitive and executive functioning at 4.5 years.
CONCLUSION: Advances in the quality of neonatal intensive care may mean that MITP-type interventions now have limited additional impact on preterm infants' long-term neurobehavioural outcomes. The gestational age of infants and the exact timing of intervention may also affect its efficacy.

Pediatric Research (2019) 86:92–99; https://doi.org/10.1038/s41390-019-0385-9

INTRODUCTION

Preterm birth (11% of live births worldwide) is a leading cause of neurodevelopmental disability in middle and high-income countries.[1] Preterm infants (born at <37 weeks gestational age: GA) are at elevated risk of long-term developmental difficulties across various functional domains.[2] Whilst individual outcomes range across a spectrum from normal, healthy development to profound disability, meta-analyses confirm moderate-to-large deficits in several neurodevelopmental domains including cognitive and executive functioning, attentional and behavioural problems and academic achievement.[3]

Early stressful experiences may play a role in the long-term outcomes of children born preterm.[4–6] Stressors may include, but are not limited to, necessary medical interventions in early life, as well as over-stimulation.[4–6] For example, preterm infants are exposed to the noisy, often light-filled and busy environment of the Neonatal Intensive Care Unit (NICU) and may experience acute and chronic stress from sensory over-stimulation, as well as from painful but necessary medical and nursing procedures (e.g. various skin punctures including heel pricks or intravenous medication, intubation, extubation, eye examination, changes to airway management).[4,7] Such stressors may be further compounded by a lack of opportunity for sustained nurturance through physical closeness and touch due to separation of preterm infants from their mothers, as well as caregiver interactions that are inappropriate for a preterm infant's developmental stage.[8]

This situation has stimulated research investigating the benefits of actively intervening to reduce preterm infants' stress exposure while still in the NICU as an approach distinct from post-discharge early intervention programs that extend into early childhood.[9] Previous studies of early stress-reduction interventions commencing in the NICU have combined various elements including: control of visual, auditory and tactile stimuli; 'clustering' of NICU care practices; providing infants with appropriate swaddling, nesting or positioning; ensuring regular periods of uninterrupted rest, and reducing sources of stress in caregiver-infant interactions.[10] Broadly, stress-reduction interventions that commence while preterm infants are in the NICU focus either on individualised 'developmental care' plans implemented by clinical staff[11] or on training parents (usually mothers) to recognise and minimise stress in care-giving interactions with infants. Various promising

[1]Parent-Infant Research Institute, Department of Clinical & Health Psychology, Austin Health, Melbourne, VIC 3081, Australia; [2]Melbourne School of Psychological Sciences, University of Melbourne, Melbourne, VIC 3010, Australia; [3]Research School of Psychology, The Australian National University, Canberra, ACT 2000, Australia; [4]Discipline of Psychological Sciences, Australian College of Applied Psychology, Melbourne, VIC 3000, Australia; [5]Monash Institute of Cognitive and Clinical Neurosciences, Melbourne, VIC 3800, Australia; [6]Clinical Sciences, Murdoch Children's Research Institute, Melbourne, VIC 3052, Australia; [7]Department of Neonatal Medicine, The Royal Children's Hospital, Melbourne, VIC 3052, Australia; [8]Departments of Obstetrics and Gynaecology, and Paediatrics, University of Melbourne, Melbourne, VIC 3052, Australia and [9]Department of Psychiatry, University of Vermont, Burlington, VT 05405, USA
Correspondence: Alan W. Gemmill (alan.gemmill@austin.org.au)

Behavioural and cognitive outcomes following an early stress-reduction…
J Milgrom et al.

93

results have been reported from individual studies of this kind aimed at reducing the sources of environmental stress outlined above.[11–14] However, systematic reviews suggest that evidence for improved long-term developmental outcomes is inconclusive.[9,10,14]

Several parent-focused NICU-based interventions in this field have been variations on the Mother–Infant Transaction Program (MITP) based on the work of Rauh and colleagues[15] in Vermont, USA in the 1980s (the Vermont Intervention Program for Low Birth Weight Infants). The MITP builds parents' skills in adapting to the bidirectional dynamics of their interactions with infants in a manner that is sensitive and responsive to the developmental stage, physiological state, and regulatory capacities of the immature preterm infant.[15–18] In the first MITP study,[15] the intervention began during hospitalisation in NICU and was extended by four home visits spaced out until 90 days after discharge. Children's development was followed for several years but intervention effects appeared to emerge only gradually, increasing through childhood, culminating in a difference of 10.6 IQ points favouring intervention children by 9 years of age, a substantial and clinically important difference.[18] The apparent longer-term efficacy of the original MITP study stimulated subsequent research including three further MITP-based randomised controlled trials (RCTs)[19–24] which, to date, have produced mixed results.

One of these, the PremieStart trial reported here, was a randomised controlled trial involving 123 very preterm (29–32 weeks GA), and extremely preterm (<28 weeks GA) infants allocated to either an MITP-type intervention or to standard care in two NICUs in Melbourne, Australia.[19] Our primary hypothesis was that at 2 and 4.5 years corrected age, on measures of cognitive, executive and behavioural functioning, the performance of the preterm children who received the early stress-reduction intervention would be superior to the children in the control condition receiving standard clinical care. We have previously published some positive early findings at 6 months corrected age in this cohort showing that the intervention enhanced maternal sensitivity in early mother–infant interactions as intended and was associated with improved early communication abilities in infants.[19]

METHODS
Design and sample
The study design, sample characteristics and early results are described more fully elsewhere.[19] Briefly, the study was a parallel, two-group randomised controlled trial (RCT) comparing the PremieStart parent-sensitivity training program to standard clinical care (Trial Registration Number: ACTRN12606000412538), designed to be conducted and reported in compliance with the CONSORT statement.[25] The trial was approved by the human research ethics committees of the Royal Women's Hospital and the Mercy Hospital for Women, both located in Melbourne, Australia. The study was powered to detect a minimum clinically important difference (MCID) of half a standard deviation in the main outcomes with 80% power at $\alpha = .05$.[19] Women who delivered at <30 weeks gestation at the NICUs of both participating hospitals were approached when infants were 30–32 weeks postmenstrual age. Exclusion criteria were as follows: (i) insufficient spoken and written English, (ii) triplets or higher multiples, (iii) infants with congenital abnormalities, (iv) infants or mothers judged to be too severely medically ill to participate by their attending physicians, (v) maternal drug and alcohol abuse or dependence, or (vi) residing >100 km from Melbourne. Of 732 women assessed for eligibility, 378 met exclusion criteria (>100 km from Melbourne, $n = 185$; infant/mother too severely ill/medically unstable, $n = 76$; non-English speaking, $n = 32$; drug and alcohol issues, $n = 27$; other, $n = 43$; multiple criteria, $n = 15$). A further

162 declined participation, and 83 did not respond to contact from the researchers, giving a final sample of $n = 109$ women who provided informed consent. Once informed consent was obtained, women were allocated randomly to treatment conditions in a 1-to-1 ratio using a pre-generated, permuted blocks schedule stratified by study site, operated by an independent administrator.[19] Fifty-four women were allocated to the intervention and 55 women to the control group. Together, these 109 women had a total of 123 infants (including 6 pairs of twins in the intervention group and 8 pairs of twins in the control group).

Intervention group. The content of the PremieStart parent-sensitivity training program is described in more detail elsewhere[12,19] and was commenced while preterm infants were in NICU. Adhering to a structured, manualised protocol, nine sessions were delivered in NICU (one per week), followed by a home-booster session 1 month after discharge. Standard clinical procedures for the care of preterm infants at both hospitals also occurred for infants in this group. Components of PremieStart include: intensive training of parents to recognise signs of infant stress such as 'shut-down' mechanisms, alert-available behaviour, facial expressions, quality of motor behaviours, posture and muscle tone, how to provide graded stimulation, how to avoid overwhelming infants, touch, movement, massage, 'kangaroo care' (skin-to-skin nesting of infants), and multi-sensory stimulation. Also, the intervention includes elements aimed at normalising parental feelings, challenging dysfunctional thoughts, and encouragement of parental diary keeping (as an opportunity to diarise the unique experiences of holding, feeding, reading, bathing, singing etc as mother of a preterm baby). The program was specifically designed to support mothers by enhancing the mother's knowledge, skills and ability in understanding her preterm infant's behaviour. The aim is to help parents develop skills in understanding and relating to their infant and to promote sensitive interactions by providing modelling, verbal instruction, direct demonstration, practical experience in handling infants, and offering emotional support to the parents to deal with the challenges of parenting a preterm infant. Two psychologists with extensive experience in preterm populations (CN, a neuropsychologist and CF, a clinical psychologist) provided the training. The psychologists did not discuss the intervention with nursing and medical staff and had limited contact with them. Mothers in the intervention group were explicitly asked not to discuss details of the study with staff or other mothers in the NICU. Before the trial, one psychologist (C.N.), an author of the PremieStart program, trained the other psychologist (C.F.) in delivery of the program. Furthermore, delivery of the program followed a manualised protocol outlining the content to be covered in each of the nine sessions. This enabled the delivery of the intervention in a standardised format by the two psychologists administering the program. Weekly meetings were held between the two psychologists to ensure standardisation and to address any issues. As a check on treatment fidelity and treatment adherence, both psychologists completed session-by-session compliance checklists for each participant in the intervention group to ensure full adherence to the manual content.

Standard care control group. As in the intervention group, standard developmental care procedures for preterm infants were in place at both NICUs. At the time of this work these included individualised care plans for the infants and parent attendance at group educational classes covering topics such as massage, recognising signs of infant distress and kangaroo care. In addition, mothers in the control group received a single psychoeducation session which included an introduction to stress and anxiety management techniques within a cognitive behavioural framework. Cognitive strategies covered cognitive distortions and applying adaptive coping strategies, and behavioural strategies

Behavioural and cognitive outcomes following an early stress-reduction…
J Milgrom et al.

94

included relaxation and breathing techniques. This was followed by a non-therapeutic, 10-min contact with one of the study psychologists every week for the remainder of the period equivalent to the scheduling of parent-sensitivity training sessions in the intervention group.

Measures

Behavioural outcomes at 2 and 4.5 years. The Total Problems score as well as the Internalising and Externalising syndrome subscales of the Child Behavior Checklist (CBCL) were the primary measures of child behaviour problems.[26] The CBCL is one of the most widely used standardised measures for evaluating maladaptive behavioural and emotional problems in children 18 months and older. Many studies have demonstrated high rates of concordance between the CBCL and actual psychiatric diagnoses.[27] Completed by a parent or other caregiver, the CBCL contains 99 items, scored 0 = not true, 1 = sometimes true, and 2 = very true or often true, based on the preceding 2 months, to yield empirically based syndrome scores. The CBCL assesses internalising (emotional reactivity, anxious/depressive, somatic complaints, withdrawal) and externalising (attention problems, aggression) behaviours. Last, Deficient Emotional Self-Regulation (DESR) is a well validated and widely used measure, calculable from aggression/anxiety-depression/attention scores on the CBCL, for assessing emotion regulation problems.[28]

General development at 2 years. At 2 years corrected age, the Bayley Scales of Infant & Toddler Development 3rd Edition (Bayley-III)[29] was administered by psychologists blinded to treatment allocation to assess cognitive, language, and motor development. Higher scores on the Bayley-III indicate better performance on cognitive, language and motor tasks.

Cognitive outcomes at 4.5 years. At 4.5 years corrected age, the Wechsler Preschool and Primary Scale of Intelligence 3rd Edition (WPPSI-III)[30] was administered to assess general cognitive functioning. The WPPSI-III is an individually administered clinical instrument for assessing the intelligence of young children 2 years + 6 months to 7 years + 3 months of age. The tests provide composite scores in Verbal IQ (acquired knowledge, verbal reasoning and comprehension, and attention to verbal stimuli), Performance IQ (fluid reasoning, spatial processing, attention to detail, and visual-motor integration), and a composite measure that represents general intellectual ability – Full Scale IQ. Processing Speed (visual-motor processing speed and accuracy) is also reported.

At the 4.5-year follow-up, children also completed subtests from the NEPSY-II.[31] The NEPSY-II is a neuropsychological assessment that assesses executive functioning, attention, language, memory and learning, sensorimotor functioning, social perception and visuospatial processing in children between 3 and 16 years. The NEPSY-II has good reliability for subtests. Full testing time is ~1 ½ −2 h. The relevant subtests used in the present study were those that could be reliably administered to 4-year olds, drawn from three of the NEPSY-II's six domains in which developmental problems are commonly reported in preterm children: Attention and Executive Functioning domain – Statue; Language domain - Speeded Naming; Memory and Learning domain - Narrative Memory.

Both the WPPSI-III and NEPSY-II were administered by psychologists blinded to treatment allocation.

Statistical analysis

All analyses followed intention-to-treat (ITT) principles with subjects analysed in the group to which they were originally randomised. The ITT approach was implemented through the use of linear-mixed modelling (LMM) and, as a further sensitivity analysis, we investigated whether there had been any differential attrition between the two groups in respect of key baseline characteristics. LMM accommodates repeated measurements, and takes into account all available data so that cases with missing data are not fully excluded. Multiple imputation (MI) of missing data was considered; however, the pattern of missing data meant that assumptions for the use of MI were not adequately met. Further, some provisional analyses using MI resulted in near identical outcomes to what is reported here. In all cases, results remained qualitatively unchanged. Mean group scores are reported as cross-sectional descriptive statistics. Effect sizes are given as Cohen's d or η_p^2 with associated 95% confidence intervals (CIs).

All comparisons were tests of the null hypothesis of no difference between the groups and were evaluated against a (critical p-value) of .01 in order to account for multiple testing. All assumptions were adequately met for the reported analyses.

The primary behavioural outcome measure was appropriate for use at both the 2-year and 4.5-year time points and this was analysed using LMM around a mixed factorial design with intervention vs. control as the between-subjects factor and time point (2 vs. 4.5 years) as the within-subjects factor. Compound symmetry provided the covariance structure with the best model fit for this analysis. All of the other outcomes were analysed at a single time point via a direct comparison of the covariate adjusted means between the two groups. The following baseline variables, collected when infants were still hospitalised, were entered as covariates: maternal age; gestational age (GA); birth weight; sex; length of stay in hospital (as a proxy for severity of medical illness); twin status (twin vs. singleton), and a 5-point scale measuring the presence and severity of intraventricular haemorrhage.[19] Choice of covariates was guided by established principles taking into account a number of factors, including: the theoretical importance of the covariates, relationships among covariates, psychometric properties of the covariates, and the impact of the number of covariates on the statistical power of the main analyses. Computations were executed in IBM SPSS Statistics, version 25.

RESULTS

Table 1 shows the baseline characteristics of the 123 preterm infants and their mothers. Participant retention at both follow-ups was relatively high with data returned for 107 children (87%) at 2 years and 96 children (78%) at 4.5 years. Figure 1 shows the flow of participants through the trial from baseline to 4.5 years. To check that attrition of the sample over time had not introduced systematic bias in the baseline characteristics of the two groups by follow-up, we conducted comparisons of responders vs. non-responders via t-tests and χ^2 tests. These confirmed that the two groups remained comparable on key baseline variables at the 2-year and 4.5-year time points. The mean age (corrected for prematurity) at 2-year follow-up was 2.05 years in the intervention group (SD = 0.08) and was also 2.05 years in the control group (SD = 0.08). At 4.5-year follow-up, the mean age was 4.64 years in the intervention group (SD = 0.18) and 4.67 years in the control group (SD = 0.19). For one child in the control group, impairment due to cerebral palsy and an intellectual disability meant that Bayley-III, WPPSI-III and NEPSY-II testing were neither possible nor valid.

Treatment adherence

Session checklists showed that, among the 54 women in the intervention group, compliance was 100% with all sessions attended by all participants and all content items delivered by the psychologists.

Behavioural outcomes at 2 and 4.5 years

Results on the CBCL are summarised in Table 2. After controlling for key baseline covariates in the ITT analysis, the factorial analyses

Behavioural and cognitive outcomes following an early stress-reduction…
J Milgrom et al.

95

Table 1. Participant characteristics at baseline

|	Intervention	Control
Mothers	$n = 54$	$n = 55$
Mean age in years (SD)	34.2 (4.7)	33.3 (5.5)
Born in Australia, n (%)	34 (62.9)	38 (69.1)
Completed high school, n (%)	13 (24.1)	14 (24.5)
Higher diploma or degree, n (%)	27 (50)	29 (52.7)
Annual family income in $Australian		
Up to $60,000	9	9
$60,000–80,000	6	6
>$80,0000	27	28
Not disclosed	12	12
Infants	$n = 60$	$n = 63$
Singletons, n (%)	48 (80)	47 (75)
Twins, n (%)	12 (20)	16 (25)
Female, n (%)	30 (50)	33 (52.4)
Mean birth weight in grams (SD)	966.8 (240.5)	1046.1 (332.2)
Mean gestational age in weeks (SD)	27.4 (1.5)	27.8 (1.7)
Antenatal maternal corticosteroids, n (%)	54 (90)	52 (82.5)
Unplanned caesarean extractions, n (%)	9 (15)	11 (17.5)
Intraventricular haemorrhage		
Grades 1–4 (n)	1 (3)	1 (5)
	2 (3)	2 (7)
	3 (2)	3 (1)
	4 (2)	4 (3)
Length of stay in hospital, days (SD)	89.98 (38.12)	85.02 (37.16)
Died before reaching term-age, n	2	0

on each of the CBCL subscales revealed only one significant result; there was a significant main effect of time for Internalising Problems such that scores in both groups rose across time, $F(1, 75.98) = 18.20$, $p < .001$, $\eta^2_p = .19$ (95% CIs, .06 to .34). There was some descriptive evidence that intervention group CBCL Total Problems scores and DESR scores increased more steeply between 2 and 4.5 years than in the control group; however, this was not statistically significant ($p = .08$ in each case). Also, comparisons between the two groups for each subscale at both ages failed to reveal any significant differences. Despite the lack of significance, it is interesting to note that there was a change in the pattern of effect sizes at 2 and 4.5 years. Even though the average magnitude of effect across the subscales did not alter appreciably across the two time points, at 2 years, three of the four measured effects showed a direction in favour of the intervention group, whereas at 4.5 years, all of the effects were in a direction that favoured of the control group.

General development at 2 Years
At 2 years of age (Table 3), after controlling for key baseline covariates in the ITT analysis, no significant between-group differences were detected on the Bayley-III Cognitive, Language, or Motor composites. One of the three observed effects (Cognitive composite) was in the direction of more positive outcomes for the intervention group and two were in the opposite direction (Table 3).

Cognitive functioning at 4.5 years
At 4.5 years of age (Table 4), analysis revealed no significant between-group differences on the WPPSI-III Full Scale IQ, Verbal IQ, Performance IQ or Processing Speed. The intervention and control groups did not differ significantly on the NEPSY-II Narrative Memory Contrast Score, the Statue Total Scaled Score or the Speeded Naming Scaled Combined Score. The pattern of effect sizes in terms of direction of effect and average magnitude of effect was similar for both outcome measures (Table 4).
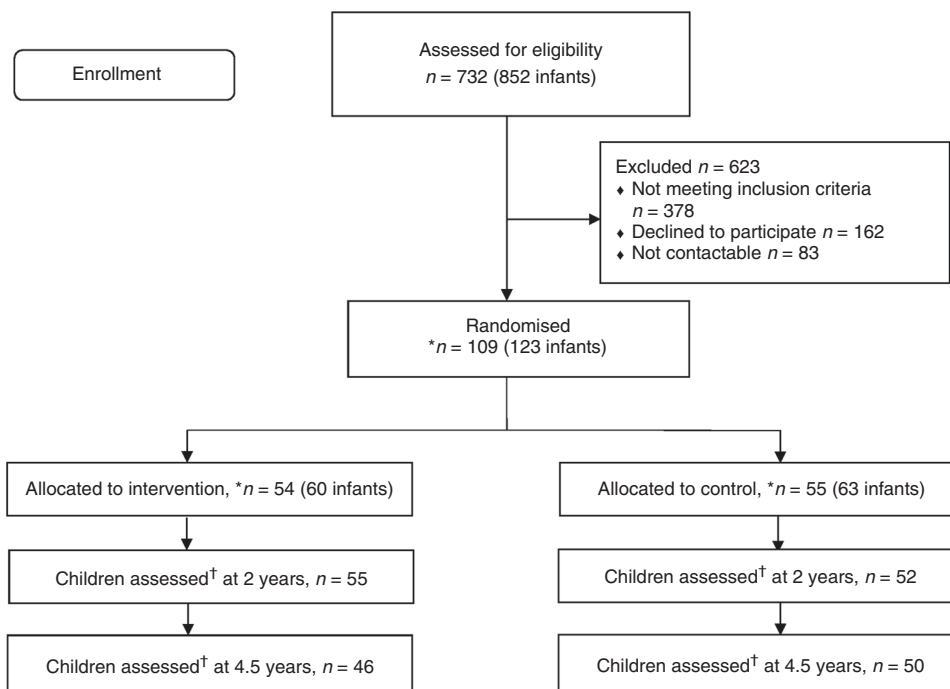
Fig. 1 Participant Flow through the Study. * n = mothers randomised; (n) in parenthesis = number of infants (there were 14 pairs of twins). † Not all assessed infants had data on all subscales of outcome measures, see Tables 2–4

Table 2. Behavioural problems (CBCL) at 2 and 4.5 years

Score[a]	Intervention		Control		Adjusted mean difference[b] (95% CIs)	Effect size[c] (95% CIs)	p value[d]
	n	Mean (SD)	n	Mean (SD)			
2 Years							
Internalising problems	50	5.46 (4.93)	51	6.16 (6.13)	0.01 (−2.45 to 2.48)	0.00 (−0.09 to 0.09)	.99
Externalising problems	50	9.14 (5.76)	51	10.94 (8.32)	−2.01 (−4.95 to 0.93)	0.27 (−0.12 to 0.66)	.18
Total problems	50	23.00 (13.41)	51	26.57 (18.27)	−1.94 (−4.54 to 0.65)	0.18 (−0.21 to 0.57)	.14
DESR	50	10.48 (6.31)	51	12.53 (9.50)	−2.09 (−5.41 to 1.23)	0.25 (−0.14 to 0.64)	.22
4.5 Years							
Internalising problems	45	8.76 (7.17)	39	7.13 (6.14)	1.89 (−0.76 to 4.53)	−0.31 (−0.11 to 0.74)	.16
Externalising problems	45	10.31 (7.25)	40	10.60 (6.63)	0.84 (−2.35 to 4.03)	−0.11 (−0.30 to 0.55)	.60
Total problems	42	27.68 (17.22)	40	26.13 (15.22)	3.80 (−3.39 to 10.98)	−0.23 (−0.18 to 0.68)	.30
DESR	45	12.62 (8.34)	40	12.20 (7.68)	1.60 (−1.99 to 5.20)	−0.12 (−0.29 to 0.56)	.38

DESR deficient emotional self-regulation
[a]Observed values before adjustment for key covariates described in Methods
[b]Mean between-group difference from LMM after controlling for GA, birth weight, length of hospital stay, presence/grade of intraventricular haemorrhage, sex and twin pairs
[c]Reported effect sizes are Cohen's *d* based on the adjusted mean difference. To aid interpretation, the value of *d* has been signed such that a positive value favours the intervention group
[d]Tabled *p*-values are uncorrected for multiple testing

Table 3. Bayley scales at 2 years

Score[a]	Intervention		Control		Adjusted mean difference[b] (95% CIs)	Effect size[c] (95% CIs)	p value[d]
	n	Mean (SD)	n	Mean (SD)			
Cognitive composite	55	107.00 (16.52)	52	105.87 (17.98)	1.46 (−5.56 to 8.48)	0.08 (−0.31 to 0.45)	.68
Language composite	52	96.90 (18.74)	50	101.26 (17.85)	−4.00 (−11.70 to 3.69)	−0.20 (−0.19 to 0.57)	.30
Motor composite	52	103.25 (17.16)	52	106.08 (20.13)	−4.03 (−11.89 to 3.82)	−0.20 (−0.19 to 0.57)	.31

[a]Observed values before adjustment for key covariates described in Methods.
[b]Mean between-group difference from LMM after controlling for GA, birth weight, length of hospital stay, presence/grade of intraventricular haemorrhage, sex and twin pairs
[c]Reported effect sizes are Cohen's *d* based on the adjusted mean difference. To aid interpretation, the value of *d* has been signed such that a positive value favours the intervention group
[d]Tabled *p*-values are uncorrected for multiple testing

DISCUSSION

Despite promising findings in infancy in this same RCT cohort,[19] the results of the present study yielded no evidence that an MITP-type intervention led to sustained behavioural or cognitive benefits for preterm children at later ages. At 2 years there were no significant group differences in child behaviour (as measured by the major domains of the CBCL). In both groups, the average behavioural scores reported here were not within a range that would be considered of clinical concern[26] Likewise, there were no differences at 2 years in language, cognitive and motor development (as measured by the Bayley-III). Similarly, at 4.5 years there were no significant group differences in either behavioural problems or in cognitive performance (WPPSI-III) and executive function (NEPSY-II). Furthermore, the observed effect sizes in these domains, even if they had been detectable as statistically significant in a larger sample, would have failed to meet the threshold of the pre-specified minimum clinically important difference (0.5 SD difference in primary outcomes) on which this RCT was originally powered.[19]

These findings are in agreement with some results from other MITP-based studies but in apparent conflict with others. The PremieStart trial is one of three RCTs that have tried to replicate the positive results of the original MITP program (one in Australia, one in Norway, and one in Sweden). All three have used similar methodologies, similar sample sizes and directly comparable primary outcome measures. In one of these cohorts (the Tromsø Intervention Study on Preterms conducted in Sweden), treatment effects in the primary behavioural and cognitive domains were absent up to 2 years of age, some effects in cognitive performance had apparently emerged by 5 years,[32] but none were sustained at 7 and 9 years.[24] In the second of these RCTs (based at Oslo University Hospital in Norway), no significant treatment effects on primary outcomes were reported by the time of the last published follow-up when children were 3 years of age.[20]

In the present study cohort, we previously reported improved early weight gain at term-equivalent age and superior communication abilities at 6 months corrected age in infants in the intervention group.[19] This seemed compatible with brain imaging results from our previous pilot RCT, with this same intervention program, which found improved frontal white matter microstructure and connectivity at term-equivalent age.[12] These very early effects on brain maturation appeared to be potentially important biomarkers of future trajectory since some deficits in the later neurodevelopmental outcomes of preterm infants are highly

Behavioural and cognitive outcomes following an early stress-reduction...
J Milgrom et al.

97

Table 4. Cognitive development and executive functioning at 4.5 years

Score[a]	Intervention		Control		Adjusted mean difference[b] (95% CIs)	Effect size[c] (95% CIs)	p value[d]
	n	Mean (SD)	n	Mean (SD)			
WPPSI-III							
Verbal IQ	42	103.64 (17.15)	48	110.46 (11.99)	−5.09 (−11.64 to 1.47)	−0.32 (−0.09 to 0.72)	.13
Performance IQ	46	107.43 (15.66)	50	106.52 (15.72)	2.46 (−3.88 to 8.78)	0.16 (−0.25 to 0.55)	.44
Processing speed	40	97.75 (14.65)	47	101.89 (12.34)	−3.94 (−10.25 to 2.37)	−0.25 (−0.15 to 0.66)	.22
Full Scale IQ	41	106.51 (15.37)	47	108.57 (12.99)	−0.34 (−6.40 to 5.72)	−0.02 (−0.38 to 0.42)	.91
NEPSI-II							
Narrative memory[e]	43	8.86 (3.17)	46	8.54 (93.76)	0.22 (−1.46 to 1.90)	0.05 (−0.36 to 0.47)	.80
Speeded naming[f]	45	9.07 (2.75)	43	10.37 (3.34)	−1.25 (−2.65 to 0.13)	−0.38 (−0.80 to 0.03)	.08
Statue[g]	41	9.88 (3.07)	42	10.60 (2.76)	−0.39 (−1.92 to 1.14)	−0.11 (−0.32 to 0.54)	.62

[a]Observed values before adjustment for key covariates described in Methods
[b]Mean between-group difference from LMM after controlling for GA, birth weight, length of hospital stay, presence/grade of intraventricular haemorrhage, sex and twin pairs
[c]Reported effect sizes are Cohen's d based on the adjusted mean difference. To aid interpretation, the value of d has been signed such that a positive value favours the intervention group
[d]Tabled p-values are uncorrected for multiple testing
[e]Scaled contrast score
[f]Scaled combined score
[g]Total scaled score

correlated with abnormalities in early brain morphology and microstructure.[33,34] However, given the results of the current study and the two other MITP-based trials discussed above, evidence for a positive effect of MITP-type early interventions on subsequent behavioural and cognitive development remains mixed. Notably, since the 9-year outcomes of the original MITP study were published,[18] no subsequent controlled trial based on the MITP has found the same enduring effect on children's later cognitive development.

A number of factors may have led to these variable findings across time and between studies. First, it is possible that the efficacy of MITP-type interventions may be conditional on GA at birth. Degree of prematurity is a key indicator of child outcomes, with a steep increase in the prevalence and severity of developmental problems as GA decreases. Even late preterm infants (those born between 34 and 37 weeks gestation) have an elevated risk of problems compared with children born after a full-term pregnancy.[14] The GA inclusion range of our sample was <30 weeks with no lower limit, resulting in an average of around 27 weeks GA and including infants born as early as 23 weeks GA. This compares with an average GA of 32 weeks in the original MITP trial in Vermont[18] and ranges of ≥30 to <36 weeks in the Oslo sample[21] and ≤36 weeks in the Tromsø study.[35] Consequently, the average birth weight of our sample (around 1000 g) was 200–300 g lower than in these other cohorts. Whilst the current study was not powered for a secondary analysis of a possible GA by treatment interaction, it seems biologically plausible that any modest benefit of parent-sensitivity training may ultimately have a negligible impact in relation to the more severe developmental vulnerabilities faced by very preterm and extremely preterm infants.

Second, the format of the original MITP was designed to deliver the bulk of intervention sessions on consecutive days, in the week immediately prior to infants' discharge from the NICU. This format was broadly retained in the two MITP-type studies conducted in Scandinavia. However, our MITP-type intervention, PremieStart, was deliberately modified to be delivered once weekly rather than daily, allowing the training to take place over a more extended timeframe throughout NICU hospitalisation, beginning around 32 weeks postmenstrual age. These variations in timing may have had an impact on the variability of findings between studies.

Third, the standard medical and nursing care available to preterm infants has advanced considerably in recent decades and continues to do so, as evidenced by ongoing improvements in survival rates.[36] Standard care practice in NICUs now routinely includes various elements shared in common with an MITP-type approach. For example, the clinical care delivered in both hospitals in the present study included not only individualised care plans for each of the infants, but also parental invitation to hospital-arranged educational classes covering topics such as infant massage, recognising signs of infant distress and kangaroo care. Control group mothers also received a 10-min weekly contact with the study psychologists. Thus the background 'standard' care received by children in later MITP-type RCTs, including the one reported here, is likely to have been of relatively high quality across both intervention and control groups compared to that available in Vermont in the 1980s.

Fourth, failure over time to replicate positive findings from an initially ground-breaking research study, the so-called 'decline effect', is a recognised phenomenon across biomedical and other scientific research.[37,38] Many factors can potentially contribute to such effects[38] including publication bias, outcome reporting bias, regression to the mean, broader advances in the quality of medical care and population health, and even the effects of early enthusiasm for, and meticulous adherence to, a novel medical innovation among pioneering researchers, study clinicians and study participants (analogous to some elements of the Hawthorne effect).[39] For example, it remains possible that the originators of the MITP delivered their parent-sensitivity training intervention in a more effective manner than subsequent trials have been able to replicate.

Strengths and limitations
The present study had some limitations including the use of a parental report instrument as the main behavioural outcome measure. Also, parents could not be kept blinded to treatment beyond the point of randomised allocation. This knowledge could have led to changed parental expectations in one or both groups

Behavioural and cognitive outcomes following an early stress-reduction...
J Milgrom et al.

98

which, potentially, could exert an influence on outcomes. These are limitations common to all existing MITP-based RCTs. However, the present study showed good participant retention and the analytical strategy adhered to ITT principles, such that the chance of type-2 error due to systematic attrition bias or inadequate statistical power was minimised. The use of blinded, standardised measures of child cognitive and executive functioning, at both 2 and 4.5 years, further strengthens our confidence in the reliability of these latest null findings. Nonetheless, while we believe that these results are likely to be generalisable to other Australian NICUs, it remains possible that greater treatment effects may have been observed had the study been conducted in another hospital or country, with less well-developed standards of NICU care. It is also possible that children may have received a variety of supports and therapies in the years following the RCT, and we collected no systematic inventory of these.

CONCLUSIONS

In summary, in a sufficiently powered RCT, we found no evidence that an early stress-reduction intervention led to sustained benefits in behavioural or cognitive outcomes for children born very preterm and extremely preterm. Compared to the results of some similar studies, this may suggest that infants born at later gestational ages may be a specific sub-population who could benefit most from continued research into early interventions of this kind. The precise timing of such interventions may also be a worthwhile focus of future research. The results of the present study add to a small, but methodologically good-quality body of research in this field. On balance, this evidence base currently suggests that, despite substantial positive findings in the earliest research, randomised trials of MITP-type interventions most often produce null or mixed results in terms of longer-term child neurobehavioral outcomes. The utility of some specific developmental care components pioneered by MITP-type programs may be growing increasingly redundant as a distinct form of early intervention, as they become progressively absorbed into standard practices for the developmental care of preterm infants. Nonetheless, the foundations of the MITP approach retain considerable theoretical validity[40] as well as empirical support from our broader understanding of the critical factors that help to shape the neurodevelopment of all human infants.[41] A focus on sensitivity and stress-reduction in mother-infant transactions therefore seems likely to remain a significant consideration in research and practice seeking to optimise the wellbeing and developmental prospects of preterm children into the foreseeable future.

Finally, it is important that researchers continue to make available the full, long-term outcomes of MITP-type trials in the published evidence base or in open access repositories, in order that their utility can be collectively assessed without bias. Further follow-up reports of the behavioural, cognitive and academic trajectories of the children involved in the PremieStart trial are planned at 6 years and 9 years of age.

AUTHOR CONTRIBUTIONS
J.M. and C.N. conceived the PremieStart intervention informed by the work of T.A. and colleagues. J.M., P.R.M., P.J.A., C.N., R.W.H., T.A. and A.W.G. contributed to study design and funding applications. J.M., C.N., C.J.H., C.F., R.W.H. and A.W.G. implemented trial procedures and collected data. C.N. and C.F. carried out the parent-sensitivity training. C.J.H. project-managed the trial. J.R. (80%) and A.W.G. (20%) designed and executed the statistical analysis. A.W.G. (55%) and J.M. (45%) drafted the manuscript. All authors contributed to the interpretation of results and approved the final manuscript.

ADDITIONAL INFORMATION
Competing interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES
1. Blencowe, H. et al. National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. *Lancet* **379**, 2162–2172 (2012).
2. Hutchinson, E. A., De Luca, C. R., Doyle, L. W., Roberts, G. & Anderson, P. J. School-age outcomes of extremely preterm or extremely low birth weight children. *Pediatrics* **131**, 2012–2311 (2013).
3. Aarnoudse-Moens, C. S. H., Weisglas-Kuperus, N., van Goudoever, J. B. & Oosterlaan, J. Meta-analysis of neurobehavioral outcomes in very preterm and/or very low birth weight children. *Pediatrics* **124**, 717–728 (2009).
4. Smith, G. C. et al. Neonatal intensive care unit stress is associated with brain development in preterm infants. *Ann. Neurol.* **70**, 541–549 (2011).
5. Doesburg, S. M. et al. Neonatal pain-related stress, functional cortical activity and visual-perceptual abilities in school-age children born at extremely low gestational age. *Pain* **154**, 1946–1952 (2013).
6. Vinall, J. et al. Invasive procedures in preterm children: brain and cognitive development at school age. *Pediatrics* **133**, 412–421 (2014).
7. Newnham, C. A., Inder, T. E. & Milgrom, J. Measuring preterm cumulative stressors within the NICU: the Neonatal Infant Stressor Scale. *Early Hum. Dev.* **85**, 549–555 (2009).
8. Newnham, C. A., Milgrom, J. & Skouteris, H. Effectiveness of a modified mother-infant transaction program on outcomes for preterm infants from 3 to 24 months of age. *Infant Behav. Dev.* **32**, 17–26 (2009).
9. Spittle, A., Orton, J., Anderson, P. J., Boyd, R. & Doyle, L. W. Early developmental intervention programmes provided post hospital discharge to prevent motor and cognitive impairment in preterm infants. *Cochrane Database Syst. Rev.* CD005495 (2015).
10. Symington, A. & Pinelli, J. Developmental care for promoting development and preventing morbidity in preterm infants. *Cochrane Database Syst. Rev.* CD001814 (2006).
11. McAnulty, G. et al. Effects of the Newborn Individualiased Developmnetal Care and Assessment Program (NIDCAP) at age 8 years: preliminary data. *Clin. Pedia.* **49**, 258–270 (2010).
12. Milgrom, J. et al. Early sensitivity training for parents of preterm infants: impact on the developing brain. *Pediatr. Res.* **67**, 330–335 (2010).
13. Als, H. et al. Early experience alters brain function and structure. *Pediatrics* **113**, 846–857 (2004).
14. Guralnick, M. J. Preventive interventions for preterm children: effectiveness and developmental mechanisms. *J. Dev. Behav. Pediatr.* **33**, 352–364 (2012).
15. Rauh, V. A., Achenbach, T. M., Nurcombe, B., Howell, C. T. & Teti, D. M. Minimizing adverse effects of low birthweight: four-year results of an early intervention program. *Child Dev.* **59**, 544–553 (1988).
16. Achenbach, T. M., Phares, V., Howell, C. T., Rauh, V. A. & Nurcombe, B. Seven-year outcome of the Vermont intervention program for low-birthweight infants. *Child Dev.* **61**, 1672–1681 (1990).
17. Rauh, V. A., Nurcombe, B., Achenbach, T. & Howell, C. The Mother-Infant Transaction Program. The content and implications of an intervention for the mothers of low-birthweight infants. *Clin. Perinatol.* **17**, 31–45 (1990).
18. Achenbach, T. M., Howell, C. T., Aoki, M. F. & Rauh, V. A. Nine-year outcome of the Vermont intervention program for low birth weight infants. *Pediatrics* **91**, 45–55 (1993).
19. Milgrom, J. et al. Early communication in preterm infants following intervention in the NICU. *Early Hum. Dev.* **89**, 755–762 (2013).
20. Kyno, N. M. et al. Effect of an early intervention programme on development of moderate and late preterm infants at 36 months: a randomised controlled study. *Infant Behav. Dev.* **35**, 916–926 (2012).
21. Ravn, I. H. et al. Effects of early mother-infant intervention on outcomes in mothers and moderately and late preterm infants at age 1 year: a randomized controlled trial. *Infant Behav. Dev.* **35**, 36–47 (2012).

Behavioural and cognitive outcomes following an early stress-reduction…
J Milgrom et al.

99

22. Olafsen, K. S. et al. Joint attention in term and preterm infants at 12 months corrected age: The significance of gender and intervention based on a randomised controlled trial. *Infant Behav. Dev.* **29**, 554–563 (2006).

23. Kaaresen, P. I. et al. A randomized controlled trial of an early intervention program in low birth weight children: outcome at 2 years. *Early Hum. Dev.* **84**, 201–209 (2008).

24. Hauglann, L. et al. Cognitive outcome of early intervention in preterms at 7 and 9 years of age: a randomised controlled trial. *Arch. Dis. Child. Fetal Neonatal Ed.* **100**, F11–F16 (2015).

25. Schulz, K. F., Altman, D. G., Moher, D & CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 340, c332 (2010).

26. Achenbach, T. M. & Rescorla, L. A. *Manual for the ASEBA Preschool Forms & Profiles: An Integrated System of Multi-informant Assessment* (University of Vermont, Research Center for Children, Youth & Families, Burlington, 2000).

27. Warnick, E. M., Bracken, M. B. & Kasl, S. Screening efficiency of the child behavior checklist and strengths and difficulties questionnaire: a systematic review. *Child Adolesc. Ment. Health* **13**, 140–147 (2008).

28. Biederman, J. et al. Severity of the aggression/anxiety-depression/attention child behavior checklist profile discriminates between different levels of deficits in emotional regulation in youth with attention-deficit hyperactivity disorder. *J. Dev. Behav. Pediatr.* **33**, 236–243 (2012).

29. Bayley, N. *Bayley Scales of Infant and Toddler Development: Administration Manual* 3rd edn (Harcourt Assessment, San Antonio, 2006).

30. Wechsler, D. *The Wechsler Preschool and Primary Scale of Intelligence – WPPSI-111 Technical and Intrepretive Manual* 3rd edn (The Psychological Corporation, San Antonio, 2002).

31. Korkman, M., Kirk, U. & Kemp, S. *NEPSY-11 NCS*. (Pearson Inc, Tx, USA, 2007).

32. Landsem, I. P. et al. Does an early intervention influence behavioral development until age 9 in children born prematurely? *Child Dev.* **15**, 12368 (2015).

33. Ullman, H. et al. Neonatal MRI is associated with future cognition and academic achievement in preterm children. *Brain* **138**, 3251–3262 (2015).

34. Thompson, D. K. et al. Regional white matter microstructure in very preterm infants: predictors and 7 year outcomes. *Cortex* **52**, 60–74 (2014).

35. Nordhov, S. M. et al. Early intervention improves cognitive outcomes for preterm infants: randomized controlled trial. *Pediatrics* **126**, E1088–E1094 (2010).

36. Stoll, B. J. et al. Trends in care practices, morbidity, and mortality of extremely preterm neonates, 1993-2012. *JAMA* **314**, 1039–1051 (2015).

37. Schooler, J. Unpublished results hide the decline effect. *Nature* **470**, 437 (2011).

38. Ioannidis, J. P. Contradicted and initially stronger effects in highly cited clinical research. *JAMA* **294**, 218–228 (2005).

39. Sedgwick, P. & Greenwood, N. Understanding the Hawthorne effect. *Br. Med. J.* **351**, h4672 (2015).

40. Sameroff, A. J. in *The Transactional Model of Development: How Children and Contexts Shape Each Other* (ed. Sameroff, A. J.) pp 3–21 (American Psychological Association, Washington, DC, 2009).

41. Feldman, R. & Eidelman, A. Biological and environmental initial conditions shape the trajectories of cognitive and social-emotional development across the first years of life. *Dev. Sci.* **12**, 194–200 (2009).