Food additives and hyperactive behaviour in 3-year-old and 8/9-year-old children in the community: a randomised, double-blinded, placebo-controlled trial

Donna McCann, Angelina Barrett, Alison Cooper, Debbie Crumpler, Lindy Dalen, Kate Grimshaw, Elizabeth Kitchin, Kris Lok, Lucy Porteous, Emily Prince, Edmund Sonuga-Barke, John O Warner, Jim Stevenson

Lancet 2007; 370: 1560–67

Published Online September 6, 2007
DOI:10.1016/S0140-6736(07)61306-3

See Comment page 1524

See Department of Error page 1542

School of Psychology (D McCann PhD, A Barrett BSc, A Cooper MSc, D Crumpler BSc, L Dalen PhD, E Kitchin BSc, K Lok MSc, L Porteous BSc, E Prince MSc, Prof E Sonuga-Barke PhD, Prof J Stevenson PhD) **and School of Medicine** (K Grimshaw MSc), **Department of Child Health, University of Southampton, Southampton, UK; and Department of Paediatrics, Imperial College, London, UK** (Prof J O Warner MD)

Correspondence to: Prof Jim Stevenson, School of Psychology, Faculty of Medicine, Health and Life Sciences, University of Southampton, Southampton SO17 1BJ, UK
jsteven@soton.ac.uk

Summary

Background We undertook a randomised, double-blinded, placebo-controlled, crossover trial to test whether intake of artificial food colour and additives (AFCA) affected childhood behaviour.

Methods 153 3-year-old and 144 8/9-year-old children were included in the study. The challenge drink contained sodium benzoate and one of two AFCA mixes (A or B) or a placebo mix. The main outcome measure was a global hyperactivity aggregate (GHA), based on aggregated z-scores of observed behaviours and ratings by teachers and parents, plus, for 8/9-year-old children, a computerised test of attention. This clinical trial is registered with Current Controlled Trials (registration number ISRCTN74481308). Analysis was per protocol.

Findings 16 3-year-old children and 14 8/9-year-old children did not complete the study, for reasons unrelated to childhood behaviour. Mix A had a significantly adverse effect compared with placebo in GHA for all 3-year-old children (effect size 0·20 [95% CI 0·01–0·39], p=0·044) but not mix B versus placebo. This result persisted when analysis was restricted to 3-year-old children who consumed more than 85% of juice and had no missing data (0·32 [0·05–0·60], p=0·02). 8/9-year-old children showed a significantly adverse effect when given mix A (0·12 [0·02–0·23], p=0·023) or mix B (0·17 [0·07–0·28], p=0·001) when analysis was restricted to those children consuming at least 85% of drinks with no missing data.

Interpretation Artificial colours or a sodium benzoate preservative (or both) in the diet result in increased hyperactivity in 3-year-old and 8/9-year-old children in the general population.

Introduction

Artificial food colours and other food additives (AFCA) have long been suggested to affect behaviour in children.[1] Ben Feingold made his initial claims of the detrimental effect of AFCA on childhood behaviour more than 30 years ago.[2] The main putative effect of AFCA is to produce overactive, impulsive, and inattentive behaviour—ie, hyperactivity—which is a pattern of behaviour that shows substantial individual differences in the general population. Children who show this behaviour pattern to a large degree are probably diagnosed with attention-deficit hyperactivity disorder (ADHD). Despite the failure of early studies[3] to identify the range of proposed adverse affects, a recent meta-analysis[4] of double-blinded, placebo-controlled trials has shown a significant effect of AFCA on the behaviour of children with ADHD. The possible benefit in a reduction in the level of hyperactivity of the general population by the removal of AFCA from the diet is less well established. Evidence from our previous study on the Isle of Wight has suggested adverse effects on hyperactivity, measured by parental ratings for 3-year-old children on a specific mix of additives.[5] These findings needed replication on 3-year-old children, and to establish whether the effects could be seen with a wider range of measures of hyperactivity. The present community-based, double-blinded, placebo-controlled food challenge was designed to extend the age range studied to include 8/9-year-old children to determine whether the effects could also be detected in middle childhood.

Methods

Participants

Figures 1 and 2 present details of recruitment and participation in the study, for 3-year-old and 8/9-year-old children, respectively. The study sample was drawn from a population of children aged between 3 years and 4 years, 2 months, registered in early-years settings (nurseries, day nurseries, preschool groups, playgroups) and from children aged between 8 and 9 years attending schools in Southampton, UK. To ensure that the study sample included children from the full range of socioeconomic backgrounds, schools were recruited based on the number of children receiving free school meals (an index of social disadvantage). The distribution of the percentage of children receiving free meals in the schools taking part indicated the proportions for the city as a whole. To further check on how representative the sample was, teachers completed a hyperactivity questionnaire[6] for all 3-year-old and 8/9-year-old children.

Parents who returned an expression of interest form were contacted by phone and a home visit arranged. On this visit, a research assistant and the study dietitian,

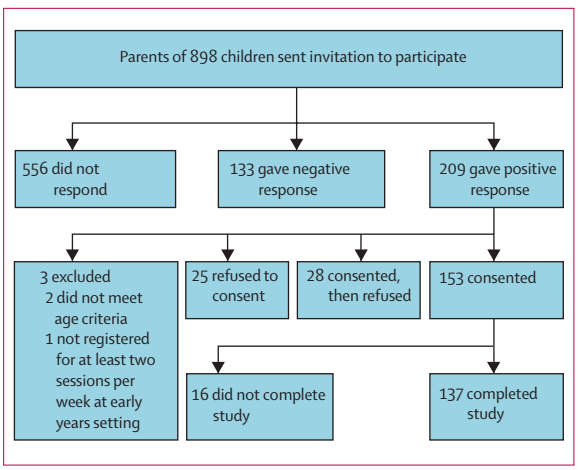

Figure 1: **Enlistment of 3-year-old participants**

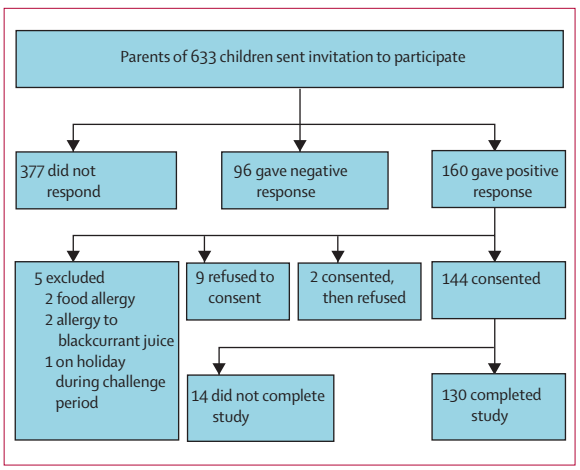

Figure 2: **Enlistment of 8/9-year-old participants**

provided full information about the study and its dietary implications, and written informed consent was obtained. The study dietitian also obtained a report based on 24-h recall by the parent of the child's pretrial diet, which allowed an assessment of baseline levels of the number of foods containing additives consumed by the child in the previous 24 h. The study was approved by the local research ethics committee (reference no 04/Q1702/61) and written informed consent was obtained from parents. Participating early-years settings received £250 and each school £500 as a contribution towards school funds for the benefit of all children.

Study design and challenge protocols

The study design and challenge protocols for both ages were similar. Children were entered into this study with a within-subject crossover between two active mixes (A and B) and a placebo drink.

The two active mixes differed both in the quantities of additives and the specific additives included. Mix A was similar to the active challenge used in the Isle of Wight study,[5] and mix B was selected to indicate the current average daily consumption of food additives by 3-year-old and 8/9-year-old children in the UK.[7] Both mixes included sodium benzoate, which had been included in the challenge on the Isle of Wight study and in previous studies.[8,9]

Mix A for 3-year-old children included 20 mg of artificial food colourings (5 mg sunset yellow [E110], 2·5 mg carmoisine [E122], 7·5 mg tartrazine [E102], and 5 mg ponceau 4R [E124, Forrester Wood, Oldham, UK]) and 45 mg of sodium benzoate [E211, Sigma Aldridge, Gillingham, UK]). Active mix B included 30 mg of artificial food colourings (7·5 mg sunset yellow, 7·5 mg carmoisine, 7·5 mg quinoline yellow [E110], and 7·5 mg allura red AC [E129]) and 45 mg of sodium benzoate.

Mix A amounts for 8/9-year-old children were multiplied by 1·25 to account for the increased amount of food consumed by children at this age. Therefore, mix A included 24·98 mg of artificial food colourings (6·25 mg sunset yellow, 3·12 mg carmoisine, 9·36 mg tartrazine, and 6·25 mg ponceau 4R) and 45 mg of sodium benzoate. Active mix B included 62·4 mg of artificial food colourings (15·6 mg sunset yellow, 15·6 mg

	3-year-old children in total sample analysis (n=153)	8/9-year-old children in total sample analysis (n=144)
Racial background		
White	126 (82%)	130 (90%)
Other	15 (10%)	14 (10%)
Missing data	12 (8%)	..
Marital status		
Married/partner	127 (83%)	115 (80%)
Single/separated/divorced/widowed	26 (17%)	29 (20%)
NSSC (father)		
Higher occupations	34 (22%)	37 (26%)
Intermediate occupations	26 (17%)	18 (13%)
Lower occupations	51 (33%)	44 (31%)
Never worked/long-term unemployed	4 (3%)	7 (5%)
No father present	26 (17%)	29 (20%)
Missing data	12 (8%)	9 (6%)
NSSC (mother)		
Higher occupations	31 (20%)	38 (26%)
Intermediate occupations	18 (12%)	26 (18%)
Lower occupations	66 (43%)	32 (22%)
Never worked/long-term unemployed	26 (17%)	32 (22%)
Missing data	12 (8%)	16 (11%)
Mother's education		
School attendance up to age 16 years (no qualifications or certificates below "A" level)	53 (35%)	60 (42%)
"A" levels	61 (40%)	42 (29%)
University degree/postgraduate qualification	27 (18%)	27 (19%)
Missing data	12 (8%)	15 (10%)

Data are n (%). NSSC=national statistics social class.[20] Higher=managerial and professional. Intermediate=self-employed. Lower=routine work. "A" levels=pre-university, school examinations in the UK.

Table 1: **Characteristics of parents of children enlisted in study**

	Mix A		Mix B		Placebo	
	n	mean (SD)	n	mean (SD)	n	mean (SD)
3-year-old children						
Entire sample (n=140)	131	−0·11 (1·03)	134	−0·14 (1·03)	129	−0·32 (1·11)
≥85% consumption (n=130)	104	−0·11 (1·03)	108	−0·15 (1·07)	99	−0·39 (1·07)
Complete case (n=73)	73	−0·14 (1·04)	73	−0·26 (1·05)	73	−0·44 (0·98)
8/9-year-old children						
Whole sample (n=136)	132	0·25 (0·97)	133	0·33 (1·10)	127	0·19 (1·03)
≥85% consumption (n=119)	104	0·26 (0·93)	112	0·32 (1·09)	103	0·19 (1·04)
Complete case (n=91)	91	0·27 (0·92)	91	0·35 (1·08)	91	0·19 (1·06)

Table 2: **Mean GHA scores for 3-year-old and 8/9-year-old children by challenge type**

	Entire sample (n=140)	Group with ≥85% consumption (n=130)	Complete case group (n=73)
Model 1			
Intercept	−0·31 (−0·49 to −0·13)*	−0·33 (−0·53 to −0·13)†	−0·44 (−0·68 to −0·21)†
Challenge type			
Mix A *vs* placebo	0·20 (0·01 to 0·40)‡	0·24 (0·02 to 0·47)‡	0·31 (0·04 to 0·58)‡
Mix B *vs* placebo	0·16 (−0·04 to 0·35)	0·16 (−0·07 to 0·38)	0·19 (−0·08 to 0·46)
Model 2			
Intercept	−0·54 (−0·89 to −0·18)*	−0·51 (−0·92 to −0·11)	−0·58 (−1·08 to −0·09)‡
Challenge type			
Mix A *vs* placebo	0·20 (0·01 to 0·39)‡	0·28 (0·05 to 0·51)‡	0·32 (0·05 to 0·60)‡
Mix B *vs* placebo	0·17 (−0·03 to 0·36)	0·19 (−0·04 to 0·41)	0·21 (−0·06 to 0·48)
Week of study			
Week 2 *vs* week 6	0·15 (−0·05 to 0·34)	0·15 (−0·08 to 0·38)	0·19 (−0·08 to 0·46)
Week 4 *vs* week 6	0·17 (−0·03 to 0·36)	0·23 (0·00 to 0·46)‡	0·19 (−0·09 to 0·46)
Sex	0·18 (−0·10 to 0·45)	0·22 (−0·07 to 0·51)	0·05 (−0·31 to 0·40)
Baseline GHA score	0·46 (0·26 to 0·66)†	0·54 (0·31 to 0·76)†	0·36 (0·06 to 0·66)‡
Pretrial diet	0·08 (−0·02 to 0·19)	0·07 (−0·04 to 0·18)	0·09 (−0·04 to 0·23)
Maternal education level	−0·01 (−0·29 to 0·28)	−0·04 (−0·34 to 0·26)	−0·03 (−0·41 to 0·35)
Maternal social class	0·15 (−0·44 to 0·13)	−0·23 (−0·53 to 0·08)	−0·21 (−0·58 to 0·16)

Data are estimate (95% CI). *p<0·01. †p<0·001. ‡p<0·05. Complete case=≥85% consumption and no missing data. Model 1=challenge type alone. Model 2=challenge type with additional factors controlled.

Table 3: **General GHA estimates in linear mixed models during challenge period for 3-year-old children**

carmoisine, 15·6 mg quinoline yellow, and 15·6 mg allura red AC) and 45 mg of sodium benzoate.

Doses for mixes A and B for 3-year-old children were roughly the same as the amount of food colouring in two 56-g bags of sweets. For 8/9-year-old children, the dose for mix A was equal to about two bags of sweets a day and for mix B about four bags of sweets a day.

After a week on their typical diet (week 0: baseline diet), the artificial colours to be used in the challenges and sodium benzoate were withdrawn from their diet for 6 weeks. Over this period when challenge with active or placebo drinks were given, additive withdrawal continued (week 1: withdrawal period but receiving placebo; weeks 2, 4, and 6: challenge with randomisation to two active periods and one placebo period; weeks 3 and 5: washout continuing on placebo). During this period, 3-year-old children received the challenge and washout-placebo drinks on a weekly basis and consumed mixed fruit juices (placebo or active) at home (300 mL/day for 3-year-old children, 625 mL/day for 8/9-year-old children), provided in identical sealed bottles. At the beginning of the study, children were assigned by the study administrator by a random-number generator to receive one of six possible sequences of placebo, active mix A, or active mix B challenges across weeks 2, 4, and 6.

A masked testing by two independent panels of 20 young adults showed that the active and placebo juice drinks could not be differentiated. When asked if the mix contained additive, 16 (40%), 21 (52%), and 26 (65%) adults responded positively for mix A, mix B, and placebo, respectively. We recorded no significant differences between these proportions (Friedman test, χ^2=4·412, df=2). Therefore, no reliable differences were seen between the look and taste of the drinks. The only difference in the composition of the placebo and active mixes was the presence of the AFCA in the active mix with some variation in the proportions of the fruit juices to ensure matching colour and taste for the placebo and active drinks. The child's family and the research team were masked to the challenge allocation. The study administrator assigned the challenge sequence and assisted in the preparation and packaging of juice drinks that were then delivered by the masked research team to homes every week, when questionnaires and other forms were obtained and dispensed. Parents completed a daily diary of juice consumption and compliance with the diet over the study period. Parents also recorded a mistake event when a child consumed a portion of food containing the artificial colours or sodium benzoate. Any bottles containing juice not consumed in the previous week were obtained, returned to the study office, and measured to help validate, if possible, parental reports of juice consumption by children.

Global hyperactivity aggregate (GHA)

Three measures of behaviour were used to calculate GHA for 3-year-old children, with an additional measure for 8/9-year-old children. First, the abbreviated ADHD rating scale IV (teacher version)[6] was used. A total score was obtained for ten of the 18 items (inattentive=5, hyperactive=5) in this questionnaire, which was completed to describe the frequency of the specific behaviours displayed over the past week, for every week of the study. Parent behaviour was the second measure, by use of the abbreviated Weiss-Werry-Peters (WWP) hyperactivity scale,[10] which has been used in several studies to assess hyperactivity.[11,12] Interparent agreement is good for ratings of childhood behaviour (r=0·82).[13] Parents rated their child's behaviour during the previous week for seven items previously used (switching activities; interrupting or talking too much; wriggling; fiddling with objects or own body; restless; always on the go; concentration),[4] from which we obtained a total score. For 8/9-year-old children, we used an abbreviated ADHD rating scale IV (parent

version)[14] to measure parent behaviour, whereby a ten-item questionnaire was completed by parents every week.

A third measure was the classroom observation code,[15] which assesses the occurrence of 12 mutually exclusive behaviours during structured didactic teaching and during periods of independent work under teacher supervision. To develop this measure, the behaviours had been selected to indicate components of ADHD that are shown in the classroom. After observers (psychology graduates) were given extensive training, the inter-rater reliability of the classroom observation code, tested before and during the study, exceeded 0·87. Children were observed for 24 min every week (three observation sessions of 8 min each) and a total weekly mean score was derived from the total score over every session. The code was slightly modified for 3-year-old children, since preschool children in the UK are not usually given structured or didactic teaching sessions and tend to engage in activities rather than in tasks. Observation took place over a range of activities and the off-task category in the code was scored when the child switched activities.

A fourth measure for 8/9-year-old children was the Conners continuous performance test II (CPTII),[16] a test using visual stimuli of 14-min duration and is widely used to assess attention and the response inhibition component of executive control. We used four scores (SE of reaction time, % of commission errors, d´ [discriminability index], and β) to derive a weekly aggregate score. This subset of indicators from the CPTII has been shown to be highly correlated with the ADHD rating scale.[17]

The GHA was developed to measure individual differences in hyperactivity using different sources (teacher, parent ratings, direct observation, and a computerised test) and covering the components of hyperactivity (overactivity, impulsivity, and inattention). Weekly scores for every child were standardised to time 0 at baseline (T0). Weekly standardised (z) aggregate scores were calculated as: (score minus mean score at T0) divided by SD at T0. The GHA was an equally weighted aggregate of the weekly z-scores, and calculated only when at least two (or three for 8/9-year-old children) of these behaviour scores were present for any week (one of which being for the classroom observation code) and averaged across the number of available scores. A high GHA indicates more hyperactivity.

Statistical analysis

Although the study designs for the two age groups were similar, the difference in composition of the GHA, and in the dose of AFCA used, meant that data from the two studies could not be analysed jointly. Therefore, we treated the studies as parallel but independent.

Linear mixed-model methods[18,19] in SPSS (version 14.0) were used to analyse data. Several possible covariates were thought to be significantly related to GHA (eg, sex). Two models were tested separately for each age for the effects on GHA in challenge weeks. Model 1 used the challenge type alone as a fixed effect testing for mix A against placebo and mix B against placebo. In model 2, in addition to challenge type, the effects of the following factors were adjusted for: week during study, sex, GHA in baseline week, number of additives in pretrial diet, maternal educational level, and social class. A compound symmetry covariance matrix provided the best-fit model for 3-year-old children and an unstructured covariance matrix for 8/9-year-old children. The study was powered to detect differences between the active and placebo periods and, accordingly in each case, the effects of mix A and mix B were compared with that of placebo. We anticipated that the additional controls on placebo effects would result in an effect size smaller than that achieved in the Isle of Wight study.[5] A sample of 80 children had

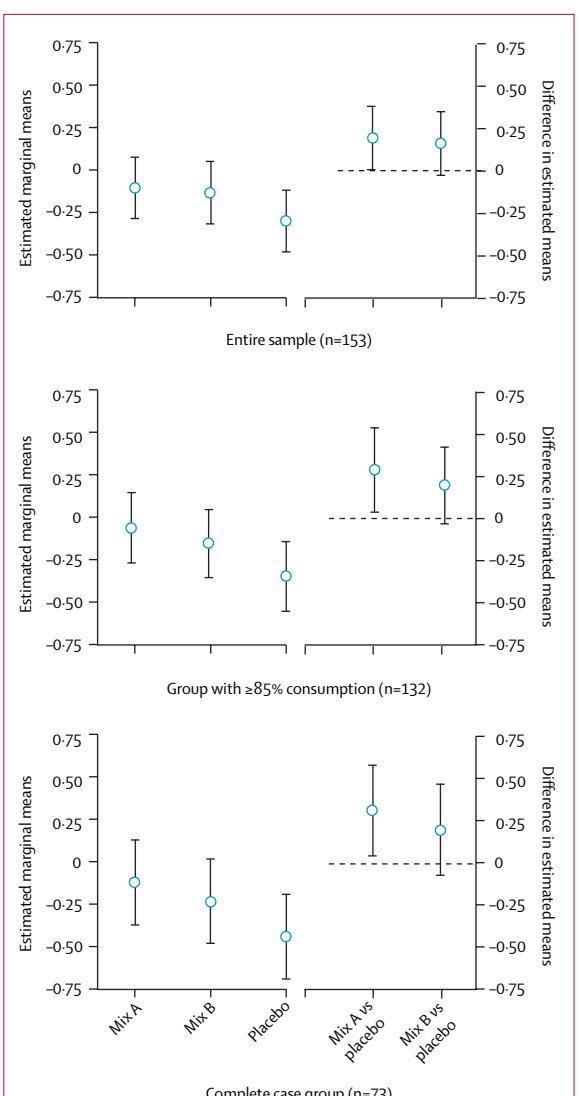

Figure 3: **Estimated marginal means by challenge type and difference in estimated means in GHA under model 2 for 3-year-old children**
Bars=95% CI. Dashed line=zero difference between mean GHA under active mix and mean GHA under placebo.

	Enitre sample (n=136)	Group with ≥85% consumption (n=119)	Complete case group (n=91)
Model 1			
Intercept	0·16 (–0·01 to 0·34)	0·09 (–0·09 to 0·27)	0·11 (–0·10 to 0·32)
Challenge type			
Mix A *vs* placebo	0·08 (–0·02 to 0·18)	0·12 (0·02 to 0·23)*	0·14 (0·03 to 0·24)*
Mix B *vs* placebo	0·12 (0·03 to 0·22)*	0·15 (0·05 to 0·25)†	0·17 (0·06 to 0·28)†
Model 2			
Intercept	0·02 (–0·22 to 0·26)	0·14 (–0·08 to 0·37)	0·14 (–0·12 to 0·39)
Challenge type			
Mix A *vs* placebo	0·08 (–0·02 to 0·17)	0·09 (–0·01 to 0·19)	0·12 (0·02 to 0·23)*
Mix B *vs* placebo	0·12 (0·03 to 0·22)*	0·15 (0·05 to 0·25)†	0·17 (0·07 to 0·28)†
Week of study			
Week 2 *vs* week 6	–0·11 (–0·21 to 0·00)*	–0·19 (–0·29 to –0·08)†	–0·20 (–0·32 to –0·09) †
Week 4 *vs* week 6	0·06 (–0·03 to 0·14)	0·04 (–0·06 to 0·13)	0·03 (–0·07 to 0·13)
Sex	0·16 (–0·03 to 0·35)	0·08 (–0·10 to 0·26)	0·11 (–0·09 to 0·31)
Baseline GHA score	0·78 (0·69 to 0·88)‡	0·79 (0·71 to 0·88)‡	0·79 (0·70 to 0·89)‡
Pretrial diet	0·04 (–0·02 to 0·10)	0·03 (–0·03 to 0·09)	0·02 (–0·05 to 0·09)
Maternal education level	–0·02 (0·20 to 0·16)	–0·02 (–0·19 to 0·15)	0·01 (–0·18 to 0·21)
Maternal social class	0·04 (–0·14 to 0·22)	–0·03 (–0·20 to 0·14)	–0·06 (–0·25 to 0·13)

Data are estimate (95% CI). *p<0·05. †p<0·01. ‡p<0·001. Complete case=≥85% consumption and no missing data. Model 1=challenge type alone. Model 2=challenge type with additional factors controlled.

Table 4: **General GHA estimates in linear mixed models during challenge period for 8/9-year-old children**

80% power at α=0·05 to identify an effect size of 0·32—ie, the magnitude of the difference in GHA mean score changes (SD). This value was lower than that achieved in the previous study (0·51). We were uncertain about the number of children and families who would comply with the demands of a 7-week study, so we set a target of 120 children to reduce the effect of attrition on power, which was eventually exceeded in both age groups.

The analyses were replicated for the full sample, a high consumption group (≥85% consumption of drinks in any challenge week), and a complete case group (≥85% consumption in all challenge weeks and no missing GHA). The high consumption and complete case groups were included to determine whether non-compliance and the method of handling missing data affected the pattern of results. Analysis was per protocol.

This clinical trial is registered with Current Controlled Trials (registration number ISRCTN74481308).

Role of the funding source

The sponsor of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

Of 153 children (mean age 43·5 months [SD 4·5]) enlisted, 79 were boys (43·5 months [4·6]) and 74 were girls (43·4 months [4·3]). Table 1 provides parents' characteristics for the entire sample. We saw no significant differences in these background characteristics between groups assigned to receive the challenge drinks in different orders during each of the six periods. The proportion of children in each of five quintile ranges on the teachers questionnaire[6] was not significantly different for the sample or for the total population (n=898, χ^2 [4]=1·60).

16 (10%) 3-year-old children failed to complete the study. Age, sex, and marital status of the parents had no effect on study completion and children were no more likely to drop out during active challenge weeks than placebo. In only one case was this failure to complete related to problems with the child's behaviour. Of those children lost to the study, 12 had a mean of 41% consumption in the first challenge week and data were missing for four children. 128 (93%) of the 137 children who completed the study consumed more than two-thirds of all drinks, of which 103 (80%) consumed 85% or more (ie, at least six of seven daily drinks per week). Only one of the remaining nine children drank less than 50% of placebo and active drinks during the study period. The occurrence of dietary infractions or mistakes by 3-year-old children was low (0=33% of children, 1–2=31%, 3–4=18·3%, >4=17%). Rate of infractions did not differ during active and placebo weeks.

117 (76%) 3-year-old children had complete GHA data over active and placebo weeks, 19 (12%) had two GHA scores, and one had one score. Of children who left the study, 12 provided one score, and four had missing data.

Table 2 shows the mean GHA scores under each of the three challenge types. For the challenge periods in weeks 2, 4, and 6, preliminary analyses had shown no effect of the type of challenge in the previous challenge period on the GHA; therefore, the washout periods had eradicated carry-over effects. Table 3 shows the results of the linear mixed-model analyses for 3-year-old children. For model 1 (the unadjusted effects of challenge type), all three samples had significant adverse effects of mix A on GHA compared with placebo. The higher GHA scores for mix B were not significantly greater than for placebo. Under model 2, with the effects of other factors controlled, the effect of mix A was significant for the entire sample (table 2, p=0·044), by contrast with that of mix B (p=0·093). When the analyses are restricted to those children with at least 85% juice consumption, the adverse effect of mix A on behaviour was still significant (p=0·016), but non-significant for mix B (p=0·098). The complete case groups showed the same pattern of results (mix A, p=0·020; mix B, p=0·131). Figure 3 shows estimated marginal mean scores after adjustment for factors in model 2.

Of 144 8/9-year-old children (mean age 106·3 months [SD 5·9]) enlisted to the study, 75 were boys (106·4 months [6·1]) and 69 were girls (106·1 months [5·8]). Table 1 provides parents' characteristics for the entire sample. We recorded no significant differences in these background characteristics between groups of children

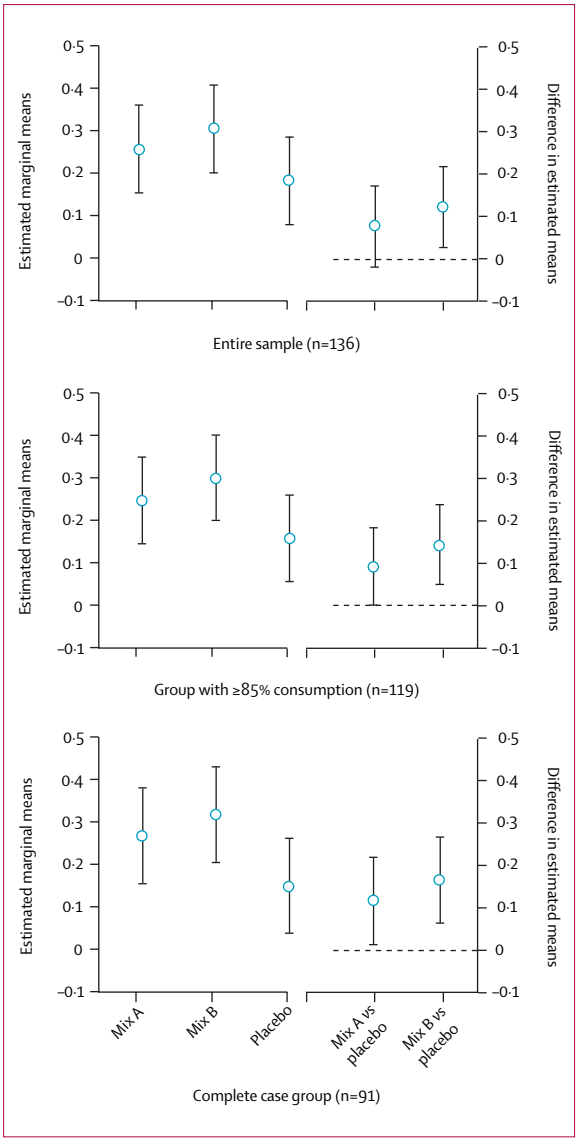

Figure 4: **Estimated marginal means by challenge type and difference in estimated means in GHA under model 2 for 8/9-year-old children**
Bars=95% CI. Dashed line=zero difference between mean GHA under active mix and mean GHA under placebo.

assigned to receive the challenge drinks in different orders over each of the six periods. The proportion of children in each of five quintile ranges on the teachers questionnaire[6] was not significantly different for the sample and for the total population (n=663, χ^2 [4]=5·05).

14 (10%) 8/9-year-old children failed to complete the study; reasons for failure were unrelated to behavioural problems. Age, sex, and marital status of the parents had no effect on study completion and children were no more likely to drop out during active challenge weeks than placebo. Of those children lost to the study, two had a mean of 93% consumption in the first challenge week and data were missing for 12 children. Of the remaining children who completed the study, 98 (75%) consumed 85% or more of the drinks over the challenge weeks (at least six of seven daily drinks per week). Only seven of the remaining 28 children drank less than 50% of placebo and active drinks over the study period. The occurrence of dietary infractions or mistakes by 8/9-year-old children during the study period was low (0=25% of children, 1–2=41%, 3–4=19%, >4=16%). Rate of infractions did not differ during active and placebo weeks.

Of 125 8/9-year-old children, 114 (87%) had full GHA data during active and placebo weeks, six (4%) had two GHA scores, and five (3%) had one score; eight (6%) had no GHA scores. Table 2 also shows mean GHA scores for 8/9-year-old children for the entire sample, the group with at least 85% consumption, and the complete case sample. For the challenge periods in weeks 2, 4, and 6, preliminary analyses showed no effect of the type of challenge in the previous challenge period on the GHA, showing that the washout periods had eradicated carry-over effects. For model 1 (the unadjusted effects of challenge type) the effects of mix A and mix B were significantly greater than that of placebo, with the exception of the entire sample in which the effects of mix A versus placebo fail to reach significance (table 4). Under model 2, in which the effects of other factors were controlled, the effect of mix A for the entire sample was not significant (p=0·123) but mix B did have a significantly adverse effect compared with placebo (p=0·012). When the analyses are restricted to those children who consumed at least 85% juice, the adverse effect of mix A on behaviour remained non-significant (p=0·066) but was significant for mix B (p=0·003). The complete case groups showed significantly higher GHA scores than placebo for mix A (p=0·023) and mix B (p=0·001). Figure 4 shows the estimated marginal means score after adjustment for factors in model 2.

Discussion

In this community-based, double-blinded, placebo-controlled food challenge, we tested the effects of artificial food additives on children's behaviour and have shown that a mix of additives commonly found in children's food increases the mean level of hyperactivity in children aged 3 years and 8/9 years. Our complete case data has indicated that the effect sizes, in terms of the difference between the GHA under active mix and placebo challenges, were very similar for mix B in 3-year-old and 8/9-year-old children. For mix A, the effect for 3-year-old children was greater than for 8/9-year-old children. The effects for mix B were not significant for 3-year-old children because there was greater variability in the response to the active challenges than placebo in this age group. Thus, we recorded substantial individual differences in the response of children to the additives. For both age groups, no significant effect of social and demographic factors was seen on the initial level of GHA or in the moderation of the challenge effects. The moderating

effects of genotype on the child's behaviour response to AFCA are examined in a separate paper (unpublished data).

The effect sizes reported in this study are similar to those calculated in the meta-analysis by Schab and Trinh.[4] They estimated the effects of AFCA on hyperactivity to be 0·283 (95% CI 0·079–0·488), falling to 0·210 (0·007–0·414) when the smallest and lowest quality trials were excluded. It should be noted that this meta-analysis included studies of hyperactivity in clinical samples, whereas the present study was done on children in the general population with the full range of degrees of hyperactivity. These effect sizes recorded by Schab and Trinh are smaller than those reported for stimulant treatment for ADHD in children, for which one meta-analysis[21] reported a range of effect sizes from 0·78 (0·64–0·91) by teacher report to 0·54 (0·40–0·67) by parent report. We report effect sizes that average at about 0·18. Children with ADHD are generally about 2 SD higher on hyperactivity measures than those without the disorder,[22] thus an effect size of 0·2 is about 10% of the behavioural difference between them.

This study provides evidence of deleterious effects of AFCA on children's behaviour with data from a whole population sample, using a combination of robust objective measures with strong ecological validity, based partly on observations in the classroom and ratings of behaviour made independently by teachers and by parents in the different context of the home and applying double-blinded challenges with quantities of additives equal to typical dietary intakes. It also replicates the effects of mix A previously reported on a large sample (n=277) of 3-year-old children,[5] although significant effects were only seen with parental ratings in that study.

The specific deleterious compounds in the mix cannot be determined for the present study and need to be examined in subsequent studies. The effect of artificial colours needs to be differentiated from the effects of preservatives in a 2×2 design. Further investigation would also need to establish whether the age-related difference seen in the present study can be replicated—ie, the effects of mix A being greater for 3-year-old children than for 8/9-year-old children. We examined the effects of additives on changes in behaviour during an extended period in a community-based, double-blinded, placebo-controlled food challenge. A weakness in this approach is the lack of control over when the challenges are ingested in relation to the timing of measures of hyperactivity. This study design also needs extensive resources to obtain multisource and multicontext measures of hyperactivity. We have completed a pilot study showing that changes in hyperactivity in response to food additives can be produced within about 1 h. Therefore, future studies could use more feasible acute double-blinded challenges undertaken in more controlled settings.

The present findings, in combination with the replicated evidence for the AFCA effects on the behaviour of 3-year-old children, lend strong support for the case that food additives exacerbate hyperactive behaviours (inattention, impulsivity, and overactivity) in children at least up to middle childhood. Increased hyperactivity is associated with the development of educational difficulties, especially in relation to reading, and therefore these adverse effects could affect the child's ability to benefit from the experience of schooling.[23] These findings show that adverse effects are not just seen in children with extreme hyperactivity (ie, ADHD),[4] but can also be seen in the general population and across the range of severities of hyperactivity. Our results are consistent with those from previous studies and extend the findings to show significant effects in the general population. The effects are shown after a rigorous control of placebo effects and for children with the full range of levels of hyperactivity.

We have found an adverse effect of food additives on the hyperactive behaviour of 3-year-old and 8/9-year-old children. Although the use of artificial colouring in food manufacture might seem superfluous, the same cannot be said for sodium benzoate, which has an important preservative function. The implications of these results for the regulation of food additive use could be substantial.

Contributors

JS, JOW, and ES-B participated in the conception and design of the study. The Food Standards Agency assisted with the design of the study. DMC directed the execution of the study. AB, AC, DC, LD, EK, LP, and EP undertook assessments of the children and helped to develop the observational methods employed in the study. KG supervised and KL executed the nutritional aspects of the study in relation to the preparation of suitable challenge drinks and advice on diet for parents. DMC and JS analysed the data and wrote the manuscript with input from all the authors.

Conflict of interest statement

We declare that we have no conflict of interest.

Acknowledgments

We thank the children, families, and teachers in the participating schools and early years settings in the Southampton area for their help and assistance with the study; and Catherine Varcoe-Baylis and Jenny Scoles and the local steering committee for their assistance, especially Ulrike Munford and the representatives from Southampton City Council Children's Services and Learning and the Southampton Early Years Development and Childcare Partnership; and the Food Standards Agency for their significant advice and input into the study. This study received funding from the Food Standards Agency (grant T07040).

References

1 Overmeyer S, Taylor E, Annotation: principles of treatment for hyperkinetic disorder: practice approaches for the UK. *J Child Psychol Psychiatry* 1999; **40:** 1147–57.

2 Feingold BF. Hyperkinesis and learning disabilities linked to artificial food flavors and colours. *Am J Nurs* 1975; **75:** 797–803.

3 Editorial. NIH consensus development conference: defined diets and childhood hyperactivity. *Clin Pediatr* 1982; **21:** 627–30.

4 Schab DW, Trinh NT. Do artificial food colours promote hyperactivity in children with hyperactive syndromes? A meta-analysis of double-blind placebo-controlled trials. *J Dev Behav Pediatr* 2004; **25:** 423–34.

5 Bateman B, Warner JO, Hutchinson E, et al. The effects of a double blind, placebo controlled, artificial food colourings and benzoate preservative challenge on hyperactivity in a general population sample of preschool children. *Arch Dis Child* 2004; **89:** 506–11.

6 DuPaul GJ, Power TJ, Anastopoulos AD, Reid R, McGoey K, Ikeda M. Teacher ratings of ADHD symptoms: Factor structure and normative data. *Psychol Assess* 1997; **9:** 436–44.

7 Gregory JR, Collins EDI Davies PSW, Hughes JM, Clarke PC. National Diet and Nutrition Survey: children aged 1·5 to 4·5 years. Vol 1: Report of the Diet and Nutrition Survey. London: HM Stationery Office, 1995.

8 Egger J, Graham PJ, Carter CM, Gumley D, Soothill JF. Controlled trial of oligoantigenic treatment in the hyperkinetic syndrome. *Lancet* 1985; **325:** 540–45.

9 Carter CM, Urbanowicz M, Hemsley R, et al. Effects of a few food diet in attention-deficit disorder. *Arch Dis Child* 1993; **69:** 564–68.

10 Routh D. Hyperactivity. In: Magrab P, ed. Psychological management of pediatric problems. Baltimore: University Park Press, 1978: 3–8.

11 Thompson MJJ, Stevenson J, Sonuga-Barke E, et al. The mental health of preschool children and their mothers in a mixed urban/rural population. Prevalence and ecological factors. *Br J Psychiatry* 1996; **168:** 16–20.

12 Hayward C, Killen J, Kraemer H, et al. Linking self-reported childhood behavioural inhibition to adolescent social phobia. *J Am Acad Child Adolesc Psychiatry* 1998; **37:** 1308–16.

13 Mash EJ, Johnston C. Parental perceptions of child behaviour problems, parenting self-esteem and mother's reported stress in younger and older hyperactive and normal children. *J Consult Clin Psychol* 1983; **51:** 86–99.

14 DuPaul GJ, Power TJ, Anastopoulos AD, Reid R. AD/HD rating scale IV: checklists, norms and clinical interpretation. New York: Guilford Press, 1998.

15 Abikoff H, Gittleman R. Classroom observation code—a modification of the stony-brook code. *Psychopharmacol Bull* 1985; **21:** 901–09.

16 Conners CK. The Conners continuous performance test. Toronto, ON, Canada: Multi-Health Systems, 1994.

17 Epstein N, Erkanli A, Conners CK, Klaric J, Costello JE, Angold A. Relations between continuous performance test performance measures and ADHD behaviours. *J Abnorm Child Psychol* 2003; **31:** 543–54.

18 Gueorguieva R, Krystal JH. Move over ANOVA: Progress in analysing repeated-measures data and its reflection in papers published in the Archives of General Psychiatry. *Arch Gen Psychiatry* 2004; **61:** 310–17.

19 Mallinckrodt CH, Watkin JG, Molenburghs G, et al. Choice of the primary analysis in longitudinal clinical trials. *Pharm Stat* 2004; **3:** 161–69.

20 Office for National Statistics. Standard occupational classification. London: Stationery Office, 2000.

21 Schachter HA, King J, Langford S, Moher D. How efficacious and safe is short-acting methylphenidate for the treatment of attention-deficit disorder in children and adolescents? A meta-analysis. *Can Med Assoc J* 2001; **165:** 1475–88.

22 Swanson JM, Sergeant J, Taylor E, Sonuga-Barke EJS, Jensen PS, Cantwell DP. Attention-deficit hyperactivity disorder and hyperkinetic disorder. *Lancet* 1998; **351:** 429–33.

23 McGee R, Prior M, Williams S, Smart D, Sanson A. The long-term significance of teacher-rated hyperactivity and reading ability in childhood: findings from two longitudinal studies. *J Child Psychol Psychiatry* 2002; **43:** 1004–17.