European Food Safety Authority

The EFSA Journal (2008) 660, 1-54

Assessment of the results of the study by McCann *et al.* (2007) on the effect of some colours and sodium benzoate on children's behaviour [1]

Scientific Opinion of the Panel on Food Additives, Flavourings, Processing Aids and Food Contact Materials (AFC)

(Question No EFSA-Q-2007-171)

Adopted on 7 March 2008

PANEL MEMBERS[*]

Fernando Aguilar, Herman Autrup, Sue Barlow[*], Laurence Castle, Riccardo Crebelli, Wolfgang Dekant, Karl-Heinz Engel, Natalie Gontard, David Gott[*], Sandro Grilli, Rainer Gürtler, John Chr. Larsen, Catherine Leclercq, Jean-Charles Leblanc, F. Xavier Malcata, Wim Mennes, Maria Rosaria Milana, Iona Pratt, Ivonne Rietjens, Paul Tobback, Fidel Toldrá.

SUMMARY

Following a request from the European Commission, the Panel on Food Additives, Flavourings, Processing Aids and Food Contact Materials (AFC) was asked to assess the results of a recent study on the effect of mixtures of additives on children's behaviour and provide an opinion on the findings, taking into account, if possible, other available scientific literature in the related area.

A recent study by McCann *et al*. (2007) has concluded that exposure to two mixtures of 4 synthetic colours plus a sodium benzoate preservative in the diet result in increased hyperactivity in 3-year old and 8- to 9-year old children in the general population. In an earlier study by the same research team there was some evidence for adverse behavioural effects of a

[1] For citation purposes: Scientific Opinion of the Panel on Food Additives, Flavourings, Processing Aids and Food Contact Materials (AFC) on a request from the Commission on the results of the study by McCann *et al.* (2007) on the effect of some colours and sodium benzoate on children's behaviour. *The EFSA Journal* (2008) 660, 1-54.

[*] Two members of the Panel did not participate in the discussion on the subject referred to above because of possible conflict with declared interests.

mixture of 4 synthetic colours and sodium benzoate in 3-year old children on the Isle of Wight (Bateman *et al*,. 2004).

In this recent study the effects of two combinations of Tartrazine (E102), Quinoline Yellow (E104), Sunset Yellow FCF (E110), Ponceau 4R (E124), Allura Red AC (E129), Carmoisine (E122) and sodium benzoate (E211) on children's behaviour were studied. Five of the six food colours belong to the class of synthetic azo dyes and one, Quinoline Yellow (E104), is a quinophthalone. Sodium benzoate is used as a preservative.

The study involved one hundred and fifty three 3-year old and one hundred and forty four 8- to 9-year old children, selected to represent a broad range of behaviour in the general population including children with normal to high level behavioural activity. Children who were medicated for ADHD were not included. A global hyperactivity aggregate (GHA) score was the main outcome of the study, and this parameter was based on aggregated z-scores of observed behaviours and ratings by teachers, class room observers and parents, plus, for 8- to 9- year old children, a computerised test of attention.

Mix A containing Tartrazine (E102), Ponceau 4R (E124), Sunset Yellow FCF (E110), Carmoisine (E122) and sodium benzoate significantly increased GHA scores for all 3-year old children compared to the placebo control GHA scores (effect size 0.20 [CI 0.01 to 0.39], $p<0.05$).

Mix B containing Sunset Yellow FCF (E110), Carmoisine (E122), Quinoline Yellow (E104), Allura Red AC (E129) and sodium benzoate had no effect on GHA scores in 3-year old children as compared to the placebo control GHA scores (effect size 0.17 [CI -0.03 to 0.36]).

This result persisted when analysis was restricted to 3-year old children who consumed more than 85% of juice and had no missing data (complete case group); in this analysis the effect of Mix A in the 3-year old children was still significantly increased compared to placebo control (effect size 0.32 [CI 0.05 to 0.60, $p<0.05$) but for Mix B no significant effect on GHA scores was observed (effect size 0.21 [CI -0.06 to 0.48]).

For the 8- to 9- year old children a significant effect of Mix A (effect size 0.12 [CI 0.02 to 0.23], $p<0.05$) or Mix B (effect size 0.17 [CI 0.07 to 0.28], $p<0.01$) was seen when analysis was restricted to those children consuming at least 85% of drinks with no missing data (complete case group). When all 8- to 9- year old children that completed the study were taken into account, Mix A had no effect on the GHA scores compared to the placebo control (effect size 0.08 [CI -0.02 to 0.17]) and Mix B had a significant effect on GHA scores (effect size 0.12 [CI 0.03 to 0.22] $p<0.05$).

The authors concluded that exposure to synthetic colours or a sodium benzoate preservative (or both) in the diet result in increased hyperactivity in 3-year old and 8- to 9-year old children in the general population.

Based on surveys conducted from 2002 to 2005, the target colours are more frequently used in sweets but also occur commonly in soft drinks and benzoate is frequently present in soft drinks. Children consuming brightly coloured sweets may be exposed to levels comparable to those considered in the protocol of the McCann *et al*. study for one or more of the food colours studied. Comparable levels may also be reached in those children who consume brightly coloured soft drinks. The level of exposure to sodium benzoate is also likely to occur.

The Panel considers that the steps taken for score normalisation and aggregation are mathematical transformations that might affect the assumptions of normality and independence of the data which are essential for the whole statistical analysis. Therefore, the authors' primary analysis was repeated using a more justifiable and conventional statistical model, and this was

supplemented by a set of additional analyses with the aim of aiding the interpretation of the results.

The Panel considers the re-analysis undertaken by EFSA, in which all single variables (minus the individual baseline value for that variable) were considered without normalisation, so that each subject served as its own reference, as the most adequate. This re-analysis was undertaken both at the level of the individual parameters as well as on the aggregated scores.

Based on the results obtained it was concluded that the analysis with the recalculated GHA score led to broadly similar conclusions to that in the original paper by McCann *et al*, except for the following:

(1) The Mix A versus placebo comparison was not statistically significant for the 3-year olds when all subjects were included (entire sample), while the significance for the $\geq$ 85% consumption and complete case groups was increased slightly;

(2) For the 8- to 9- year age group, the Mix A versus placebo comparison was no longer statistically significant in any of the three consumption groups.

In addition the data were analysed on the basis of a modified GHA score in which the parental scores were not included. The results from this analysis no longer revealed any statistically significant effects of Mix A or Mix B versus placebo, except for Mix B versus placebo in 8- to 9-year old completers.

A further analysis was carried out on the whole data set, comprising analysis of the single variables of parental scores, teacher scores and observer scores, and, in the case of 8- to 9-year old children, computer-based scores. There is a suggestion from these analyses that the statistically significant effects seen in the 3-year olds (Mix A versus placebo) and in the 8-to 9-year olds (Mix B versus placebo) are largely driven in the data by the parental scores and, in the older males in both comparisons, by the computer score.

The Panel notes that some, but not all, earlier studies have also reported effects of food colours on child behaviour, the majority of these studies being conducted on children described as hyperactive or with a clinical diagnosis of ADHD.

The Panel concludes that the McCann *et al.* study provides limited evidence that the two different mixtures of synthetic colours and sodium benzoate tested had a small and statistically significant effect on activity and attention in some children selected from the general population, although the effects were not observed for all children in all age groups and were not consistent for the two mixtures. The findings may thus be relevant for specific individuals within the population, showing sensitivity to food additives in general or to food colours in particular.

However, it is not possible to assess the overall prevalence of such sensitivity in the general population and reliable data on sensitivity to individual additives are not available.

The clinical significance of the observed effects also remains unclear, since it is not known whether these small alterations in attention and activity would interfere with schoolwork and other intellectual functioning. The clinical significance could possibly be clarified by assessments that used scales for functional impairment and diagnostic interviews, especially if a high proportion of children with high symptom scores were to be included in such a study.

There are thus a number of uncertainties that are apparent from this new research, some of which are echoed in earlier research. These include:

- the limited consistency of the results with respect to age and gender of the children, the effects of the two mixtures of additives tested and the type of observer (parent, teacher or independent observer);
- the unknown clinical relevance of the novel metric, i.e. the GHA score;
- the unknown relevance of the small effect size (as was also seen in the meta analysis of earlier studies by Schab and Trinh, (2004));
- the fact that the study has not been designed to identify the effects of individual additives;
- a lack of information on dose-response;
- the lack of a biologically plausible mechanism for induction of behavioural effects from consumption of food additives.

The Panel concludes that the McCann *et al.* study provides limited evidence that the two different mixtures of synthetic colours and sodium benzoate tested had a small and statistically significant effect on activity and attention in children selected from the general population excluding children medicated for ADHD, although the effects were not statistically significant for the two mixtures in both age groups.

Since mixtures and not individual additives were tested in the study by McCann *et al.*, it is not possible to ascribe the observed effects to any of the individual compounds.

The clinical significance of the observed effects also remains unclear.

In the context of the overall weight of evidence and in view of the considerable uncertainties, such as the lack of consistency and relative weakness of the effect and the absence of information on the clinical significance of the behavioural changes observed, the Panel concludes that the findings of the study cannot be used as a basis for altering the ADI of the respective food colours or sodium benzoate.

Key words:

Hyperactivity, ADHD, children's behaviour, Southampton study, McCann, food additives, food colours.

Tartrazine, FD&C Yellow No. 5, E102, CAS 1934-21-0, Trisodium-5-hydroxy-1-(sulfonatophenyl)-4-(4-sulphonatophenylazo)-H-pyrazole-3-carboxylate, food colouring substance, EINECS number 217-699-5.

Ponceau 4R, New Coccine, E124, CAS Registry Number 2611-82-7, Trisodium 2-hydroxy-1-(4-Sulphonato-1-naphthylazo)-naphtalene-6,8-disulphonate, food colouring substance, EINECS number: 220-036-2.

Carmoisine, Azorubine, CI Acid Red 14 and CI food red 3, E122, CAS 3567-69-9, Disodium 4-hydroxy-3-(4-sulfonato-1-naphthylazo)naphthalene-1-sulfonate, food colouring substance EINECS number 222-657-4.

Quinoline Yellow, D&C Yellow No. 10, E104, CAS 8004-92-0, 2-(2-quinolyl)indan-1,3-dione-disulphonate, food colouring substance, EINECS number 305-897-5.

Allura Red AC, E129, CAS 25956-17-6, Food Red No. 40, FD&C Red No. 40, disodium, 2-hydroxy-1-(2-methoxy-5-methyl-4-sulphonatophenylazo)naphthalene-6- sulphonate, food colouring substance, EINECS number 247-368-0.

Sunset Yellow FCF, E110, Food Yellow No. 5, FD&C Yellow No. 6, E 110, CAS 2783-94-0, Disodium 2-hydroxy-1-(4-sulfonatophenylazo)naphthalene-6-sulfonate.

Sodium benzoate, benzoic acid, E 211, E 210, CAS 532-32-1, CAS 65-85-0, food preservative, EINECS number 208-534-8.

TABLE OF CONTENTS

BACKGROUND AS PROVIDED BY THE COMMISSION

The European Commission has been informed of a recent study, funded by the UK Food Standards Agency, examining the effect which the consumption of certain food additives may have on children's behaviour. The studies were undertaken with two age groups (3-year old children and 8- to 9-year old children) and involved the following food additives in 2 different mixtures/formulations: Tartrazine, Ponceau 4R, Carmoisine, Quinoline Yellow, Allura Red AC and sodium benzoate.

Before additives are authorised they must first be evaluated for their safety. Council Directive 89/107/EEC states that all food additives must be kept under continuous observation and must be re-evaluated whenever necessary in the light of changing conditions of use and new scientific information. Therefore when the European Commission is informed about new scientific evidence relating to a permitted food additive it requests the European Food Safety Authority to give an opinion on this new research.

TERMS OF REFERENCE AS PROVIDED BY THE COMMISSION

In accordance with Article 29 (1) (a) and 31 of Regulation (EC) No 178/2002, the European Commission asks the European Food Safety Authority to assess the results of the study and provide an opinion on the findings, taking into account, if possible, other available scientific literature in the related area.

ACKNOWLEDGEMENTS

The European Food Safety Authority wishes to thank the members of the ad hoc Working Group on the study from the Southampton University on the effect of some colours and benzoic acid on hyperactivity for the preparation of this opinion:

Jan Buitelaar, John Chr. Larsen, Manfred Doepfner, Manfred Gerlach, Michael Kenward, Alicja Mortensen, Iona Pratt (WG Chair), Diane Purper-Ouakil, Ivonne Rietjens, Terje Sagvolden, Stephen Senn, Stephan Strobel.

The Panel wishes to thank Prof. J. Stevenson, one of the authors of the McCann *et al*. study, for the fact that he has met with the ad hoc Working group to discuss methodological and other issues. In addition, the Panel wishes to thank the Food Standards Agency and Prof. J. Stevenson for providing EFSA with the study details.

The Panel also wishes to thank the Assessment Methodology Unit of EFSA for their assistance with the statistical re-analysis.

INTRODUCTION

It has been suggested for a number of years that exposure to synthetic food colours and other food additives may have behavioural effects, especially in young children, resulting in overactive, impulsive and inattentive behaviour (Feingold, 1975; Overmeyer and Taylor, 1999; Schab and Trinh, 2004). If severe, children who show this behaviour are likely to be diagnosed with Attention Deficit Hyperactivity Disorder (ADHD).

Attention Deficit Hyperactivity Disorder (ADHD) or Hyperkinetic Disorder (HKD) is a behavioural disorder, characterised by problems with sustained attention, impulsivity and hyperactivity, which adversely affects these children's behaviour. ADHD typically has onset in early childhood (WHO, 2007; American Psychiatric Association, 2000). Hypotheses about the cause of ADHD have evolved from simple one-cause theories to the view that it is a complex, multi-factorial disorder caused by the confluence of many different types of risk factors (i.e., genetic, biological, environmental, psychosocial), with each factor contributing to the vulnerability to the disorder (Biederman and Faraone, 2005; Sagvolden *et al.*, 2005). This multi-factorial view of ADHD is consistent with the observed heterogeneity in the genetics, pathophysiology and clinical manifestation of the disorder.

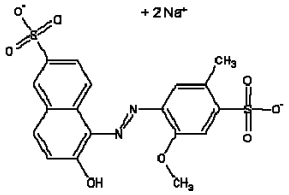
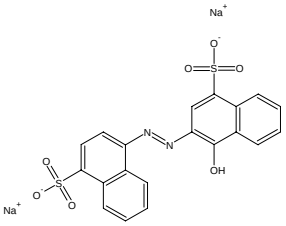
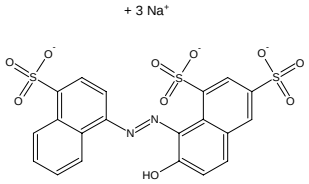
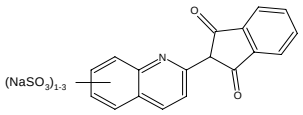
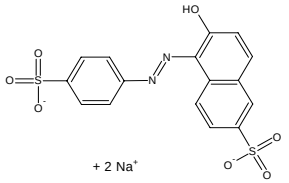
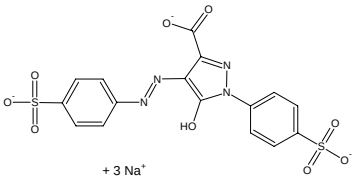
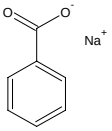
Most recently, a study by McCann *et al*. (2007) has concluded that synthetic colours plus a sodium benzoate preservative in the diet result in increased hyperactivity in 3-year old and 8- to 9- year old children in the general population. An earlier study by the same research team (the so-called Isle of Wight study) reported some evidence for adverse behavioural effects of a mixture of 4 synthetic colours and sodium benzoate, as measured by parental ratings for 3-year old children on the Isle of Wight (Bateman *et al.* 2004). In addition a meta analysis of double-blind Placebo-controlled trials has shown a small but statistically significant effect of synthetic food colours on the behaviour of children with hyperactivity (Schab and Trinh 2004). Other older studies however failed to identify similar behavioural effects (NIH, 1982; Harley *et al.,* 1978a; Harley *et al*., 1978b; Mattes and Gittelman, 1981; Karvale and Forness, 1983).

The earlier study (Bateman *et al.,* 2004) did not allow firm conclusions about the clinical significance of the observed effects of a series of food colours and the preservative sodium benzoate on children's behaviour, mainly because of limitations in the study design (COT, 2007). These limitations included among others the presence of a large Placebo effect, and the fact that statistically significant effects on behaviour were only observed with parental observations and not with assessments made by independent researchers.

Subsequently the UK Food Standards Agency (FSA) set up an ad-hoc working group to consider these limitations in study design and to make recommendations on a new study design. Based on the findings of this working group, the FSA commissioned a new study via open competition in 2004, incorporating the study design changes that had been recommended by the working group. The results of this new study were recently published (McCann *et al.* 2007).

In this study, the effects of two combinations of Tartrazine (E102), Quinoline Yellow (E104), Sunset Yellow FCF (E110), Ponceau 4R (E124), Allura Red AC (E129), Carmoisine (E122) and sodium benzoate (E211) on children's activity levels and attention were evaluated. Table 1 presents the chemical structures of the six synthetic colours included in the study. Five of the six food colours belong to the class of synthetic azodyes and one, Quinoline Yellow (E104), is a quinophthalone. Sodium benzoate is used as a preservative.

Table 1. Chemical structure of the additives included in the study.

Substance	Chemical formula	Chemical structure
Allura Red AC (E129)	$C_{18}H_{14}N_2Na_2O_8S_2$	
Azorubine = Carmoisine (E122)	$C_{20}H_{12}N_2Na_2O_7S_2$	
Ponceau 4R (E124)	$C_{20}H_{11}N_2Na_3O_{10}S_3$	
Quinoline Yellow (E104)	$C_{18}H_9NNaO_5S$ $C_{18}H_9NNa_2O_8S_2$ $C_{18}H_9NNa_3O_{11}S_3$	
Sunset Yellow FCF (E110)	$C_{16}H_{10}N_2Na_2O_7S_2$	
Tartrazine (E102)	$C_{16}H_9N_4Na_3O_9S_2$	
Sodium benzoate (E211)	$NaC_7H_5O_2$	

ASSESSMENT

EFSA's AFC Panel, in addressing the Terms of Reference provided by the European Commission, has assessed this new study in the light of previous opinions on the compounds, and has also considered more recent studies which have become available since the publication of the available opinions on these colours and benzoate, in order to evaluate the relevance of these findings for human health.

To assist the Panel in this task a number of experts in behaviour, child psychiatry, allergy, and statistics were invited to join the ad hoc Working Group (see Acknowledgements). This ad hoc Working Group met on four occasions and prepared a detailed analysis of the McCann *et al.* study.

1. Study design and conduct

The study design and conduct are described in detail in McCann *et al.* (2007). The study consisted of a community-based double-blind, Placebo-controlled randomised cross-over food challenge in 3-year old children and in 8- to 9-year old children with two mixtures (Mix A and Mix B), each consisting of 4 different colours and sodium benzoate. A mixed fruit juice drink was used as vehicle for the food colour / sodium benzoate mixtures, and a Placebo drink was used in the wash-out periods. The Placebo and the two additive mixes were identical except for the additives, and there were no nutritional differences in the composition of the drinks. Table 2 presents an overview of the additive composition of the mixtures A and B and also presents data reflecting the actual intake levels of the additives achieved.

Table 2. Overview of the composition of the mixtures A and B and of the actual intake levels achieved.

Additive (E number)	ADI mg/ kg bw	Mix A Daily dose in mg for 3- / 8- to 9-year old	Mix B Daily dose in mg for 3- / 8- to 9-year old	Mix A Daily dose in mg/kg bw* for 3-year olds (% ADI)	Mix B Daily dose in mg/kg bw* for 3-year olds (% ADI)	Mix A Daily dose in mg/kg bw* for 8- to 9-year olds (% ADI)	Mix B Daily dose in mg/kg bw* for 8- to 9-year olds (% ADI)
Tartrazine (E102)	7.5	7.5 / 9.4	-	0.5 (6.7%)	-	0.3 (4%)	-
Ponceau 4R (E124)	4.0	5.0 / 6.3	-	0.33 (8.25%)	-	0.2 (5%)	-
Sunset Yellow FCF (E110)	2.5	5.0 / 6.3	7.5 / 15.6	0.33 (13.2%)	0.5 (20%)	0.2 (8%)	0.5 (20%)
Carmoisine (E122)	4.0	2.5 / 3.1	7.5 / 15.6	0.17 (4.25%)	0.5 (12.5%)	0.1 (2.5%)	0.5 (12.5%)
Quinoline Yellow (E104)	10.0	-	7.5 / 15.6	-	0.5 (5%)	-	0.5 (5%)
Allura Red AC (E129)	7.0	-	7.5 / 15.6	-	0.5 (7.1%)	-	0.5 (7.1%)
Sodium benzoate (E211)	5.0	45 / 45	45 / 45	3.0 (60%)	3.0 (60%)	1.45 (36.3%)	1.45 (36.3%)

* The doses per kg bw were calculated by COT (2007) using average body weights for the two age groups obtained from UK National Diet and Survey data. For comparison the ADI values for these different additives are also included.

The children were selected from families volunteering from nurseries, preschool groups and playgroups for the 3-year old children and from schools in the Southampton area for 8- to 9-year old children. The children who were included in the study were selected to represent a

broad range of behaviour in the general population including children with normal to high level activity. Children who were medicated for ADHD were not included. In total one hundred and fifty three 3-year old and one hundred and forty four 8- to 9-year old children were included in the study.

The families were instructed to maintain the children during the course of the study on diets that were free of the food colours used in the study and also free of sodium benzoate. Compliance was monitored by means of a diary in which parents reported consumption levels of the test mixtures as well as compliance with the dietary requirements.

In order to investigate the hypothesis that the children's behaviour in response to the challenge with the food colours in question could be influenced by allelic variation in a number of genes that have previously been implicated in ADHD (Thapar *et al*., 1999; Swanson *et al*., 2000; Kuntsi and Stevenson, 2001), buccal swabs were collected from the children for genotypic analyses of cellular DNA (COT, 2007). The genes studied included genes from the dopamine, adrenergic and histamine neurotransmitter systems (COT, 2007).

The amounts of the different colours in Mix A given to the 3-year old children were identical to those used in the Isle of Wight study (Bateman *et al*,. 2004). For the 8- to 9-year old children the intakes of the different colours in Mix A were lower on a kg bw basis than for the 3-year old children whereas for Mix B the intakes on a mg per kg bw basis were the same for both age groups and higher than for Mix A. For sodium benzoate the intake in mg/kg bw was about 2 times higher for the 3-year old children as compared to the 8- to 9-year old children, but similar for Mix A and Mix B for each of age groups. The researchers indicated that the intakes of the different colours in both Mix A and Mix B for the 3-year olds and for Mix A in the 8- to 9-year olds were approximately equivalent to the amount of food colouring in two 56 gram bags of sweets. The intakes for Mix B for the 8- to 9-year old equated to about four bags of sweets a day.

During the 6 weeks of the study children received batches of the drinks on a weekly basis. Wash-out weeks (week 1, 3 and 5) in which the children received a Placebo drink, were alternated with challenge weeks (week 2, 4 and 6) during which the children received either Placebo drink, Mix A or Mix B in randomised order. The ingredients of the Placebo drink were free of the colours and preservative being tested in the challenge, and for 8- to 9-year olds the volumes of the different juices that made up the mixture and were consumed on a daily basis were as follows: 150 ml tropical juice, 80 ml red grape juice, 10 ml prune juice, 140 ml blackcurrant juice, 10 ml beetroot juice, 20 ml pear juice, 160 ml orange juice and 55 ml water, together making up a final volume of 625 ml per day. For the 3-year olds the volumes were reduced proportionately to provide 300ml a day. The Placebo drink was developed so that when each of the additive mixes in turn was introduced there were no detectable differences in taste, colour or smell. The Placebo and the two additive mixes were therefore identical except for the additives. It was thought essential to ensure that any behavioural effect attributable to ingredients of the Placebo mix, including idiosyncratic reactions from individual children were kept constant across the challenge types. In order to make the Mix palatable for children the sweetener aspartame was included, since studies on aspartame and hyperactivity have produced uniformly negative results (e.g. Wolraich *et al.,* 1994). In contrast there is evidence that sugar can affect inattention using one of the same computerised tests adopted in this study (Wender & Solanto, 1991).

Table 3 presents a schematic overview of the crossover trial.

Table 3. Schematic overview of the crossover trial

	Week 0	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
Sequence 1	Normal diet* (27,25***)	WO**	Mix A (27,24)	WO	Mix B (23,24)	WO	Mix C (21,23)
Sequence 2	Normal diet (25,24)	WO	Mix A (24,22)	WO	Mix C (23,21)	WO	Mix B (22,21)
Sequence 3	Normal diet (26,23)	WO	Mix B (25,22)	WO	Mix A (25,23)	WO	Mix C (20,21)
Sequence 4	Normal diet (24,24)	WO	Mix B (24,24)	WO	Mix C (21,21)	WO	Mix A (20,21)
Sequence 5	Normal diet (27,25)	WO	Mix C (24,21)	WO	Mix A (21,22)	WO	Mix B (23,22)
Sequence 6	Normal diet (24,23)	WO	Mix C (23,20)	WO	Mix B (19,20)	WO	Mix A (18,20)

*Normal diet to set baseline levels

** Wash-out period

*** Number of children with a GHA score (Number of children in 3-Year Group, Number of children in 8- to 9-Year group) (derivation of the GHA score is described below)

Behaviour at home was assessed by parents, behaviour in the classroom was assessed by teachers and by independent observers. For the 8- to 9-year old children behaviour was also assessed by a computerised attention test. Behaviour was scored at the end of each treatment week (week 2, 4 and 6) using standardised and validated ADHD-behaviour assessment tools. The following measurement tools were used:

1. ADHD rating scale IV (teacher version). A questionnaire was completed to describe the frequency of the specific behaviours displayed over the past week, for every week of the study.
2. The abbreviated Weiss-Werry-Peters (WWP) hyperactivity scale. The WWP has been used in a number of studies to assess hyperactivity. Parents rated their child's behaviour during the previous week for seven items previously used.
3. Classroom Observation Code (COC): The COC assesses the occurrence of 12 mutually exclusive behaviours during structured didactic teaching and during periods of independent work under teaching supervision.
4. Conners' Continuous performance test II (CPTII). This is a test using visual stimuli of 14 minutes duration and is widely used to evaluate attention and the response inhibition component of executive control. This test was only used for the 8- to 9-year old children.

Ratings of behaviour from each of the individual measurement tools were combined, un-weighted, into an overall Global Hyperactivity Aggregate (GHA) score, representing a novel metric developed by the researchers, combining observational and computerised behavioural scores in one parameter. ADHD rating, WWP and COC were used to calculate the GHA score for 3-year old children, with an additional measure (CPTII, being itself an aggregate of four computer scores) as a fourth instrument for 8- to 9-year old children.

The primary analysis of the data was based on the GHA score. A high GHA score indicates more activity. The authors indicated that although the designs for the two age groups were similar, the difference in composition of the GHA score, and in the dose of the additives used, meant that data from the two age groups could not be analysed jointly. Therefore the age groups were analysed in parallel but independently.

The data were analysed using linear mixed model methods in SPSS (Gueorguiva and Krystal, 2004; Mallinckrodt *et al.*, 2004). Two linear mixed models were fitted for each age group. Although in the original publication (McCann *et al.,* 2007) no details were given on the parameterisation of the model, the raw output of the analysis provided to EFSA gave detailed information. The first model is a basic mixed effect model where a random effect is put on the subject and a fixed effect on the treatment. Model 2 undertaken by the authors included some additional fixed effects (see the separate statistical report for further details).

A compound symmetry covariance matrix provided best fit for the models fitted to the data of the first age group while an unstructured covariance matrix gave the best fit for the second age group. The choice of covariance structure was done based on log-likelihood ratio comparisons for Model 1 only, accounting for the total number of parameters to be estimated.

The analyses were replicated for three sets of data: the full study population sample, a high consumption subset (data included if the child consumed ≥ 85% of drinks in each treatment week) and a complete case subset data (high consumption and no missing GHA scores).

To test whether there was evidence of carry-over effects, the scores of the previous active challenge period and baseline were added as factors in the mixed model. No effect due to the type of challenge in the previous period on the current scores could be demonstrated. From this, it was concluded by the study authors that the wash-out periods were sufficiently long to have prevented carry-over effects.

2. AFC Panel comments on study design and conduct

The Panel notes that:

- The study tested two mixtures and no individual compounds. Testing of mixtures cannot identify the hazards of individual compounds. The choice to test mixtures was based on the fact that the Isle of Wight study (Bateman *et al.* 2004) had also tested a mixture, and part of the objective of the new study was to investigate whether the findings of the Isle of Wight study could be replicated with a better study design. Mix A reflects the mixture tested in this earlier study. Mix B reflects a mixture representative for sweets as they are consumed in the UK;
- Parents, teachers and independent observers scored the behaviour of the children. The outcome of their scores was combined, and for 8- to 9-year old children the results of the CPTII computer tests were also included, to give the GHA scores. In this way observational scores were combined with computer scores;
- Including parental scores into the GHA score the study design does not completely overcome the criticisms of the earlier Isle of Wight study that effects on behaviour were only observed via parental observations and not via assessments made by independent researchers;
- The GHA score combines three measures of behaviour for the 3-year old children, adding the computer-based measure (CPTII) for the 8- to 9-year old children only. The Panel noted that using this aggregated score is adequate from a statistical point of view,

because it is considered not to affect the integrity of the statistical approach. Combining the measures does not increase the chance of introducing statistical differences which do not actually exist;

- The aggregated score is adequate to score an overall change in behaviour, but that it is not a clinically accepted and validated outcome in that it has not been assessed whether it shows meaningful relations with external variables such as prognosis and impairment of functioning in other areas of behaviour;
- The combined GHA scores reflect a global score that may be difficult to interpret behaviourally and statistically. Therefore the Panel notes that also subsequent analyses on each of the individual behaviour variables would be helpful in order to assess the relative contributions of each behavioural pattern;
- The drinks were given at home to guarantee better compliance. However, the time of day at which the children consumed the drink was not regulated. In addition, the durations of the possible behavioural effects are not known. These factors combined may have influenced the results since transient effects may not have been observed in some of the tests;
- The study was not designed to explore possible dose: response relationships or possible subgroups of "responders";
- Self-volunteering of the subjects included in the study might have introduced a selection bias;
- Although the study designs for the two age groups were similar, the difference in composition of the GHA scores and the dose of additives used meant that data from the two age groups cannot be analysed jointly;
- A one week wash-out period was chosen which was also the period used in the Isle of Wight study (Bateman *et al.* 2004). There was no evidence of carry-over effects from a challenge week into the next challenge week based on statistical analyses. Analysis of the assessment of behaviour during the wash-out period would have provided a clue on the efficiency of the wash-out period and/or possible effects of the Placebo;
- Based on data on toxicokinetics in experimental animals, it may be concluded that the test substances, with the possible exception of Quinoline Yellow, will be eliminated from the body during the wash-out periods. Quinoline Yellow has a longer half-life and some parent compound or breakdown products thereof may still be present in the body after the one week wash-out period (see Appendix A).

2.1. Comments on statistical analysis

General remarks and discussion on the design

Basic principles of cross over trials are that every subject (child) receives each of the treatments being evaluated over a standard period of time while each subject serves as its own control and the outcome variables are assessed in the same way in each period of the treatment. The major advantage of such a design is the reduction of the sample size (number of subjects) needed to achieve a certain statistical power. However, several disadvantages are inherent to a cross over trial design: the treatment must have a reversible effect; the statistical analysis is rather complicated and if not carried out correctly, may lead to erroneous conclusions. Also carry-over effects are possible, *i.e.* the residual effect of one treatment on the outcome of a

subsequent treatment. This is particularly a potential concern if no prior pharmacokinetic (PK) and pharmacodynamic (PD) assessment of the tested food additives is available to support the chosen length of the wash-out period. In the McCann *et al*. (2007) study the claim of absence of carry-over effects was not supported by preliminary investigation on the pharmacokinetics of the compounds, although a formal statistical test for carry-over was performed and no statistically significant carry-over effect was found.

Discussion on the statistical methodology used in the Mc Cann et al. paper

Analyses were based on two linear mixed models. The first and simplest model contained treatment effects only, in addition to those for random variability. With a cross-over design it is recommended that adjustment be made for possible period effects; indeed, in this study, there is clear evidence of differences between the different treatment weeks.

The second model was an extension of the first in which, in addition, period effects and a number of *between-subject* terms, including baseline GHA score, were introduced. The incorporation of between-subject effects is pointless in a cross-over design like this because such effects are anyway effectively eliminated from the analysis by the within-subject nature of the treatment comparisons. This is illustrated clearly in Tables 3 and 4 in the McCann *et al.* study (McCann *et al*., 2007), where the differences in treatment effects estimates from the two models are of little or no importance. The choice of covariance structure is very unlikely to be crucial in the analysis of data from a design like this, provided within-subject dependence is accommodated.

The steps taken for score normalisation and aggregation are mathematical transformations that might affect the assumptions of normality and independence of the data which are essential for the whole statistical analysis. Moreover, in such cross-over trials, since each subject serves as its own control, individual scores should be compared to individual baseline scores, not to the group mean baseline score.

2.2. Comments on dietary exposure levels used in the study

The researchers indicated that the daily dietary exposure to the different colours for both Mix A and Mix B for the 3-year olds and for Mix A in the 8-to 9-year olds were approximately equivalent to the amount of food colouring in two 56 g bags of sweets and that the daily dietary exposures for Mix B for the 8- to 9-year olds equated about four bags of sweets a day (McCann *et al*., 2007).

The Panel was provided with information allowing it to assess if the level of exposure considered in the study was likely to occur, based on current legislation, current actual uses and use levels in foods consumed by children and current levels of consumption of these foods by children.

The amounts in Mix A given to 3-year olds were identical to those used in the previous (Isle of Wight) study, while for 8- to 9-year olds the amounts of the colours in Mix A were increased by 25% to reflect the greater food intake by the older children.

According to COT (COT, 2007, based on information from the Food Standards Agency, UK), for 8- to 9-year olds, the amounts of the colours in Mix B reflected what a child could reasonably consume in one day, based on average consumption of foods containing colours at their maximum permitted levels (MPL).

The two main sources of added colours in children are soft drinks and confectionery. The Panel noted that according to current legislation, all six target colours may be used singly or in

combination to a cumulative MPL of 300 mg/kg in confectionery and to a cumulative MPL of 100 mg/l in soft drinks. Mix A provided overall 20 mg of artificial colours for 3-year old children and 25 mg for children aged 8 to 9 years. Mix B provided overall 30 mg of colours for 3-year old children and 62.4 mg for children aged 8 to 9 years. These levels of dietary exposure could be reached through consumption of 67 g to 208 g of confectionery containing artificial colours at their cumulative MPL. They could also be reached through consumption of 200 ml to 624 ml of soft drinks containing artificial colours at their cumulative MPL.

In addition, three of the target colours (Sunset Yellow FCF, Carmoisine and Ponceau 4R) have a lower individual MPL in the legislation: 50 mg/l in soft drinks and 50 mg/kg in confectionery. The individual dietary exposure to these colours in the study protocol was highest in Mix B: up to 7.5 mg in 3-years old children and up to 15.6 mg in 8- to 9-years old children (for Sunset Yellow FCF and Carmoisine). These levels of dietary exposure to either Sunset Yellow FCF or Carmoisine could be reached through consumption of respectively 150 g and 312 g of sweets or through the consumption of respectively 150 ml and 312 ml of soft drinks containing one of these colours at its individual MPL.

The main source of benzoates in children are soft drinks and the MPL for benzoates is 150 mg/l. Both Mix A and Mix B provided 45 mg of sodium benzoate in the two age groups. This level of exposure could be reached through consumption of 300 ml of soft drink containing benzoate at its MPL.

The Panel noted that the dietary exposures to the colours used in the study, were well below the ADIs of the individual substances. As shown in Table 2, the dietary exposures in the 3-year old children ranged from 4.3% to 13.2% of the ADIs for the individual colours in Mix A. For the 8- to 9-year old children these values for Mix A ranged from 2.5% to 8% of the ADIs for the individual colours. For the colours in Mix B the dietary exposure of the 3-year old children and the 8- to 9-year old children were similar, ranging from 5 to 20% of the ADIs of the individual colours. The dietary exposures for the two age groups were different for sodium benzoate, amounting to respectively 60% and 36.3% of the ADI for sodium benzoate for the 3-year and 8- to 9-year old children respectively.

Connolly and co-workers have investigated the frequency of occurrence of the food colours and sodium benzoate used in the McCann *et al.* study in a recent 7-day dietary survey of 594 Irish children aged 5-12 years (Unpublished data by Connolly *et al*., 2008). The food consumption data, coded at brand level, were combined with the Irish National Food Ingredient Database in which all ingredients listed on the label of food items, including additives, are recorded, (Gilsenan *et al*., 2002). In the case of sodium benzoate, the presence as a natural ingredient in the food was not considered. Among the 5,551 individual food items coded at brand level that were consumed during the survey, 279 (5%) contained at least one of the target additives. The percentage of child food-eating occasion containing the target artificial colour ranged from 138/72,024 (0.2%) for Tartrazine to 555/772,024 (0.8%) for Sunset Yellow FCF. Tartrazine, which is authorised for use in "processed mushy and garden peas (canned)" was found to occur most frequently in the food group "Peas, Beans and Lentils"; The other five colours occurred most frequently in "Chocolate and non chocolate confectionary". Other food groups containing the target colours were "Cakes, Pastries & Buns" (in particular for Carmoisine) and carbonated beverages (in particular for Sunset Yellow FCF). The frequency of occurrence for sodium benzoate was 2183/72,024 (3%); its most frequent source was beverages. The total number of observed child-days was 4158. At least one target additive occurred in 30.5% of child-days, two additives occurred in 7.7% of child-days, three additives in 5.1%, four additives in 2.8%, five additives in 2.2% and six or seven additives in 0.7% of child days. The Panel noted that the data provided do not allow the estimation of the percentage of children who were exposed to the colours or their combination in at least one of the survey day. This percentage would have been useful but was not provided in the Connolly Report.

A usage survey conducted by the Union of European Beverage Associations (UNESDA) in 2005 was made available to the Panel (Tennant, 2006). The survey report indicates that all but one of the six colours considered in the study by *McCann et al*. are commonly used in soft drinks, with Carmoisine being an uncommon artificial colour in these products. However, other surveys described underneath suggest that Carmoisine is also commonly used.

Data from three *ad hoc* surveys in which analytical determinations of artificial colours were performed in retail products were also provided to the Panel: an unpublished survey conducted in 2005 by the Food Safety Authority of Ireland (FSAI) in 34 retail ready to drink soft drinks, a survey by the UK Food Standards Agency in 201 retail ready to drink soft drinks selected for being distinctly coloured (FSA, 2003) and a survey by the UK Food Standards Agency in 196 retail samples of brightly coloured packaged sweets (FSA, 2002).

The frequency of occurrence of each of the artificial colours under study and the range of analytical values observed in soft drinks and sweets (when the colour was present) are reported in Table 4.

Among colours, Tartrazine was present with the lowest frequency: in respectively 6%, 1.5% and 3% of Irish soft drinks, UK soft drinks, and UK sweets. The colour most frequently present was Quinoline Yellow: in respectively 21%, 37% and 56% of Irish soft drinks, UK soft drinks, and UK sweets. In sweets, the overall concentration of the target colours was up to 208 mg/kg in sweets, lower than the cumulative MPL of 300 mg /kg in sweets. In the same table, the quantity of either sweets or beverages that needs to be consumed to lead to the level of exposure of the experimental protocol was calculated for each group of children in order to verify if the levels of exposure considered are in line with potential level of exposure in children consuming products present on the market. Calculations were based on the upper concentration values observed in the surveys on retail products. For sweets the quantity varied from a minimum of 42 g needed to reach the dietary exposure to Allura Red AC in Mix B for 3-year old children to a maximum of 363 g needed to reach the dietary exposure to Carmoisine in Mix B for 8- to 9-year old children. In the case of beverages, the quantity varied from a minimum of 42 ml needed to reach the dietary exposure to Carmoisine in Mix A for 3-year old children to a maximum of 371 ml needed to reach the dietary exposure to Allura Red AC in Mix B for 8- to 9-year old children.

According to UK FSA diary survey on the consumption of soft drinks by young children (FSA 2003 b), high level consumers of 1.5 to 4.5 years drank around 500 ml (just over one and a half 330 ml cans) a day.

In the UK survey, co-occurrence of two of the colours under study was observed in 41 soft drinks, co-occurrence of three of the colours under study was observed only in two soft drinks, co-occurrence of three or four of the colours under study was not observed. In the Irish survey, co-occurrence of two of the colours under study was observed in 4 soft drinks, co-occurrence of three of the colours under study was observed in one soft drink, co-occurrence of three or four of the colours under study was not observed. The co-occurrence of two, three or four of the target colours was more frequent in sweets, as observed in the UK survey. When two or more colours under study occurred in the same product, the overall concentration reached the cumulative MPL of 100 mg /l in a number of soft drinks (up to 106 mg/l in the UK survey).

In conclusion, the target artificial colours and benzoate were found to be used in foods consumed by children in surveys conducted in years 2002 to 2005. The target colours were more frequently used in sweets but also occurred commonly in soft drinks, benzoate was very frequently present in beverages. Children consuming brightly coloured sweets may be exposed to levels comparable to those considered in the protocol of the McCann *et al*. study for one or more of the food colours studied. Comparable levels may also be reached in those children who

consume brightly coloured soft drinks. The level of exposure to sodium benzoate is also likely to occur.

Table 4. Actual use levels of colours used in the Southampton study and quantities of sweets and beverages corresponding to the different Mixes

	Analytical survey of the UK Food Standards Agency conducted in 2002 in England in brightly coloured retail ready to drink soft drinks (UK FSA, 2003)		**Survey of the Food Safety Authority of Ireland conducted in 2005 (unpublished)**		**Quantity of soft drink (ml) at the highest observed concentration(*) corresponding to the exposure to individual colours in Mix A and Mix B**		**Analytical survey of the UK Food Standards Agency conducted in England in 2000/2001 in brightly coloured sweets (UK FSA, 2002)**		**Quantity of sweets (g) at the highest observed concentration(*) corresponding to the exposure to individual colours in Mix A and Mix B**	
	Occurrence in excess of LOD	Range of analytical data above LOD (mg /l)	Occurrence in excess of LOQ	Range of analytical data above LOQ (mg /l)	3-year olds	8- to 9- year olds	Occurrence	Range	3–year olds	8- to 9-year olds
Allura Red AC	6/201	9-42	2 / 34	20–32	Mix B (7.5 mg): 179 ml	Mix B (15.6 mg): 371 ml	40/196	3-177	Mix B (7.5 mg): 42 g	Mix B (15.6 mg): 88 g
Ponceau 4R	32/201	1-47	6/34	3-22	Mix A (5 mg): 106 ml	Mix A (6.25 mg): 133 ml	38/196	2-39	Mix A (5 mg): 128 g	Mix A (6.25 mg): 160 g
Tartrazine	3/201	3-28	2/34	5-25	Mix A (7.5 mg): 268 ml	Mix A (9.36 mg): 334 ml	6/196	3-63	Mix A (7.5 mg): 119 g	Mix A (9.36 mg): 149 g
Quinoline Yellow	75/201	1-92	7/34	1-34	Mix B (7.5 mg): 81 ml	Mix B (15.6 mg): 170 ml	110/196	2-200	Mix B (7.5 mg): 37.5 g	Mix B (15.6 mg): 78 g
Sunset Yellow FCF	61/201	1-61	7/34	11-49	Mix A (5 mg): 82 ml Mix B (7.5 mg): 123 ml	Mix A (6.25 mg): 102 ml Mix B (15.6 mg): 256 ml	61/196	1-106	Mix A (5 mg): 47 g Mix B (7.5 mg): 71 g	Mix A (6.25 mg): 59 g Mix B (15.6 mg): 147 g
Carmoisine	64/201	1-45	13/34	1-59	Mix A (2.5 mg): 42 ml Mix B (7.5 mg): 127 ml	Mix A (3.12 mg): 53 ml Mix B (15.6 mg): 264 ml	53/196	2-43	Mix A (2.5 mg): 58 g Mix B (7.5 mg): 174 g	Mix A (3.12 mg): 72 g Mix B (15.6 mg): 363 g

*The concentration value used is the highest range of analytical data observed in the UK and Ireland survey (evidenced in bold character in the columns reporting the ranges of analytical data)

3. Results of the study

3.1. Overview of the findings of the study (as reported by the authors)

For reasons unrelated to effects, 16 of the one hundred and fifty three 3-year old children and 14 of the one hundred and forty nine 8- to 9-year old children did not complete the study.

Tables 5 and 6 provide an overview of the results from Model 1 of the two linear mixed models used by the authors, obtained for three groups of participants from each age group: (1) the full study population sample (2) those with at least 85% consumption of the trial drinks (≥ 85% consumption), and (3) those with at least 85% consumption of the trial drinks and observations from all three periods (complete case). The latter, (2) and (3), represent subgroups of the original trial samples.

Table 5. GHA score estimates during challenge period for 3-year old children (taken from McCann *et al.* 2007‡)

	Entire sample (n=140)	Group with ≥85% consumption (n=130)	Complete case group, ≥85% consumption and no missing data (n=73)
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)

Values given in the Table represent estimates (95% CI) of the differences in GHA mean scores between the Placebo and the challenge for each population group, based on the mean baseline scores at week 0 and the mean scores following treatment.

*p<0·05.

‡ The figures quoted above are those used by the authors in their discussions of the new findings and represent the outcomes from application of statistical model 1 or model 2.

Table 6. GHA score estimates during challenge period for 8- to 9-year old children (taken from McCann *et al.* 2007‡).

	Entire sample (n=136)	Group with ≥85% consumption (n=119)	Complete case group, ≥85% consumption and no missing data (n=91)
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 (0·06 to 0·28)†

Values given in the Table represent estimates (95% CI) of the differences in GHA mean scores between the Placebo and the challenge for each population group, based on the mean baseline scores at week 0 and the mean scores following treatment.

*p<0·05. †p<0·01

‡ The figures quoted above are those used by the authors in their discussions of the new findings and represent the outcomes from application of statistical model 1 or model 2.

Mix A significantly increased GHA scores for all 3-year old children compared to the Placebo control GHA scores (effect size 0.20 [CI 0.01 to 0.39], p<0.05).

Mix B had no effect on GHA scores in 3-year old children as compared to the Placebo control GHA scores (effect size 0.17 [CI -0.03 to 0.36]).

This result persisted when analysis was restricted to 3-year old children who consumed more than 85% of juice and had no missing data; in these analysis the effect of Mix A in the 3-year old children was still significantly increased compared to Placebo control (effect size 0.32 [CI 0.05 to 0.60, $p<0.05$) but for Mix B no significant effect on GHA scores was observed (effect size 0.21 [GI -0.06 to 0.48]).

For the 8- to 9-year old children a significant effect of Mix A (effect size 0.12 [CI 0.02 to 0.23], $p<0.05$) and of Mix B (effect size 0.17 [CI 0.07 to 0.28], $p<0.01$) was observed when analysis was restricted to those children consuming at least 85% of drinks with no missing data. When all children that completed the study were taken into account in 8- to 9-year old children Mix A had no effect on the GHA scores compared to the Placebo control (effect size 0.08 [CI -0.02 to 0.17]), and Mix B had a significant effect on GHA scores (effect size 0.12 [CI 0.03 to 0.22] $p<0.05$).

Post-hoc analysis reported in the statement of the Committee on Toxicity (COT, 2007) revealed that the parental reports were the main contributors to the changes in GHA scores for the 3-year old children, whereas in the 8- to 9-year old children the largest contribution to the GHA score was reported to come from the computer-based task (COT, 2007). The researchers suggested that parents may have been more sensitive to or more exposed to behavioural changes in their children than the independent observers or teachers.

As also reported by the UK COT, the research team found that the observed increases in the GHA scores with Mix A in 3-year olds and 8- to 9-year olds and with Mix B in 8- to 9-year olds were statistically significantly associated with differences in genotype, specifically with two genetic polymorphisms thought to impair histamine clearance (COT, 2007). This analysis was carried out in the subgroup of children with at least 85% consumption of the trial drinks.

The authors concluded on the basis of these results that synthetic colours and/or a sodium benzoate preservative in the diet may exacerbate hyperactive behaviours (inattention, impulsivity, and overactivity) in 3-year old and 8- to 9- year old children in the general population.

3.2. AFC Panel assessment of observed effects in the study

The Panel noted that:

- Small but statistically significant effects of Mix A but not of Mix B on GHA scores in 3-year old children were described. In 8- to 9-year old children, when taking all children that completed the study into account, a small but significant effect of Mix B but not of Mix A on GHA scores was observed in 8- to 9-year old children. Thus, the statistically significant effects were not found for the same mixture in the two age groups;
- For Mix A the doses on a mg/kg bw/day basis were higher in the 3-year old children and this may have contributed to the difference in the magnitude of the effect of Mix A in the two age groups. For Mix B the doses in both age groups were similar but the effects were significant only in the 8- to 9-year old age group;
- The effects of Mix A on behavioural parameters in the 3-year old children were consistent with those of the Isle of Wight study, showing an increase in GHA score in the 3-year old children in the present McCann *et al.* study and an effect on the basis of only parental observations in the Isle of Wight study;

- The main contributors to the statistically significant effect on the GHA scores in the 3-year old children were the parental scores, as described in more detail in Section 3.3 below. The scores from teachers and independent observers were not a major component in the overall GHA scores. The use of the GHA scores does not therefore completely overcome the criticisms on the earlier Isle of Wight study;
- Since each subject serves as its own control, no further explanation of variation in the trial can be achieved by fitting subject level covariates. Only covariates that changed over time for a given subject could have any further explanatory power. Therefore the extra information obtained from model 2 used by the authors of the study is minimal, as shown in their publication in Tables 3 and 4. Apart from the fact that a period effect was fitted the additional factors will not explain the variation of the GHA scores;
- The clinical significance of the observed effects (a) for the individual children in the study and (b) for the population as a whole remains unclear, since the effects were small in magnitude and these small alterations in attention and activity may not interfere with schoolwork and other intellectual functioning.

3.3. Statistical re-analysis and AFC Panel assessment

The Panel considered that the steps taken for score normalisation and aggregation are mathematical transformations that might affect the assumptions of normality and independence of the data which are essential for the whole statistical analysis. Therefore, the authors' primary analysis was repeated using a more justifiable and conventional statistical model, and this was supplemented by a set of additional analyses with the aim of aiding the interpretation of the results.

Details of the statistical re-analysis can be found in the separate statistical report.

The re-analysis consisted of two parts. First, the authors' primary analysis was repeated, with minor changes to reflect a more appropriate statistical treatment and, second, a set of supplementary analyses were carried out.

For the primary analysis the Global Hyperactivity Aggregate (GHA) score was recalculated following the same steps as in the original analysis, except for the omission of the final re-normalisation step.

The remainder of the supplementary analyses consisted of the calculation of various descriptive statistics and formal analysis of each of the individual component measures.

For all formal analyses, both primary and supplemental, a linear mixed model was used that was similar to that of the first analysis reported in the Lancet paper (McCann *et al.* 2007). The model included only within-subject effects, namely those associated with the experimental intervention and with periods. Random subject effects were also included, and in this setting imply an analysis identical to that with a compound symmetry covariance structure. The "week" variable was also included as a fixed effect in the model. Consumption subgroup analyses matched those of the original paper.

The p-values were calculated for the contrast 'Mix A vs. Placebo' and 'Mix B vs. Placebo' for the whole dataset, the $\geq$ 85% consumers and the children consuming at least 85% of drinks with no missing data (complete case group). Models were run for all combinations of sexes and age groups.

Table 7 presents a summary of all statistically significant cases found in the re-analysis using the Global Hyperactivity Aggregate (GHA) score similar to that carried out in the McCann *et al. s*tudy.

Table 7. Summary of all significant cases found in the statistical reanalysis in the treatment group comparisons for the GHA score, using a similar approach to that used in the McCann *et al.* study

Test	Year Group	Sex	Estimate	Std Err	P-Value
entire sample					
Mix B vs. Placebo	8/9Y	M	0.1115	0.04394	0.0124
Mix B vs. Placebo	8/9Y	both	0.05963	0.02981	0.0466
>85% consumption					
Mix B vs. Placebo	8/9Y	both	0.08348	0.03276	0.0116
Mix A vs. Placebo	3Y	both	0.1962	0.08074	0.0161
Mix B vs. Placebo	8/9Y	M	0.1116	0.04644	0.0179
complete case					
Mix B vs. Placebo	8/9Y	both	0.08546	0.03536	0.0167
Mix A vs. Placebo	3Y	both	0.2359	0.09764	0.0169
Mix B vs. Placebo	8/9Y	M	0.1118	0.04993	0.0273

Based on these results it is concluded that the primary analysis with the recalculated GHA score led to broadly similar conclusions to that in the original paper by McCann *et al*., except for the following:

(1) The Mix A versus Placebo comparison was not statistically significant for the 3-year olds when all subjects were included (entire sample), while the significance for the ≥ 85% consumption and complete case groups was increased slightly;

(2) for the 8- to 9-years age group, the Mix A versus Placebo comparison was no longer statistically significant in any of the three consumption groups.

The Panel considers the re-analysis presented here in which all single variables (minus the individual baseline value for that variable) were reanalysed without normalisation, so that each subject served as its own references baseline, as the most adequate.

In addition the data were analysed on the basis of a modified GHA score in which the parental scores were not included. The results from this analysis no longer revealed any statistically significant effects of Mix A or Mix B versus Placebo, except for Mix B in the 8- to 9-year old completers when both sexes are pooled (p=0.042).

A further analysis was carried out on the whole data set, comprising analysis of the single variables of parental scores, teacher scores and observer scores, and, in the case of 8- to 9-year old children, computer-based scores. Table 8 presents a summary of all statistically significant cases found in the re-analysis of the single variables for all three in treatment groups.

Table 8. Summary of all statistically significant cases found in the re-analysis of single variables for all three treatment groups.

Test	Score	Year Group	Sex	Estimate	Std Err	P-Value
Entire sample						
Mix A vs. Placebo	CPTCom	8/9Y	M	5.0607	2.1497	0.0206
Mix B vs. Placebo	Parent	8/9Y	both	0.9017	0.4186	0.0322
>85% consumption						
Mix A vs. Placebo	CPTCom	8/9Y	M	5.0867	2.0922	0.0172
Mix B vs. Placebo	CPTCom	8/9Y	M	6.0375	2.0077	0.0035
Mix B vs. Placebo	Parent	8/9Y	both	0.9687	0.4584	0.0359
Complete case						
Mix A vs. Placebo	CPTCom	8/9Y	M	5.5414	2.2611	0.0166
Mix B vs. Placebo	CPTCom	8/9Y	M	6.3442	2.2427	0.006
Mix A vs. Placebo	Parent	3Y	both	1.2448	0.5585	0.0274

The main results are as follows.

- No statistically significant component effects were observed that did not coincide with effects seen already in the authors' overall GHA analysis;
- For the 3-year olds, only the Mix A versus Placebo effect with the parental score was statistically significant in the complete case group. The teacher or observer scores showed no evidence of an effect in the 3-year olds in any consumption group;
- For the 8-to 9- year olds, statistically significant Mix B versus Placebo effects were seen for the parental scores in the entire sample group and the $\geq$85% consumption group, and for the computer scores of the males in the $\geq$85% consumption group and the complete case group;
- For the computer scores there were statistically significant Mix A versus Placebo effects in the 8- to 9- year old males in all three consumption groups;
- Consistency across consumption groups could not be observed except for the computer score for males.

In conclusion, there is a suggestion from these analyses that the statistically significant effects seen in the 3-year olds (Mix A versus Placebo) and in the 8- to 9-year olds (Mix B versus Placebo) are largely driven in the data by the parental scores and, in the older males in both comparisons, by the computer score.

From the data presented in the separate statistical report, it can be derived that a 'Week' effect was shown on both single and aggregated scores but only for the "Week 4 vs. Week 6" comparison in the latter. The general trend was that hyperactivity generally went up from Week 2 to Week 4 and then significantly decreased from Week 4 to Week 6. The size of this period effect was globally of the same order of magnitude as those observed for the treatment effects, in both single and aggregated scores. It illustrates a large intra-individual variability over time and interpretation of all statistical results should be done in the light of this result.

4. Overview of previous studies of the effect of food colours and sodium benzoate on behaviour.

Since Feingold's (1975) initial report that a diet free of synthetic food colours and flavours and naturally occurring salicylates resulted in an improvement in behaviour of hyperactive children, several studies have investigated the possible relationship between exposure of young children to synthetic food colours and other food additives and behavioural effects. The AFC Panel has (briefly) reviewed these previous studies, in particular reports involving the food additives used in the McCann *et al.* study. These studies are summarised in Table 9.

Table 9. **Summary of behavioural toxicity studies carried out on food colours and other additives**

Reference	Numbers of subjects, additive tested and dose levels	Outcome
Feingold 1975	194 HDL children from five separate dietary programs following a diet eliminating synthetic food colours (including Tartrazine = Yellow dye 5), flavours and naturally occurring salicylates.	Improvement in behaviour and learning abilities.
Conners *et al.*, 1976	15 hyperkinetic children (6-13 yr) in a double-blind crossover trial involving a control diet and a diet eliminating synthetic flavours, colours and natural silicates as recommended by Feingold.	Teachers and parents observed the children for one month prior to treatment, using standardised rating scales. Both teachers and parents reported fewer hyperkinetic symptoms on the experiment test diet as compared to the pre-treatment baseline. The control diet ratings did not differ from the baseline period ratings. The teachers noted a significant reduction in symptoms on the experimental diet compared to the control diet but parents did not.
Williams *et al.*, 1978	26 Hyperactive children (5-12 yr) given active or Placebo medications in combination with challenge cookies with synthetic food colours (red dyes 2, 3 and 4; blue dyes 1 and 2; Yellow dyes 5 and 6; green dye 3; orange dye B at levels estimated to be equal to one-half the daily dietary intake of children in the USA) or control cookies without the additives. The children were crossed over into each of 4 treatments and assessment was double blind by teachers and parents.	Stimulant medications were more effective than diet in reducing hyperactive behaviour. The behaviour of 3 to 8 children was diet responsive. It was concluded that especially the teachers (but not the parents) ratings provided support of Feingold's hypothesis that food additives trigger the hyperactive response.
Harley *et al.*, 1978a	9 Hyperactive male subjects, selected on the basis of showing a favourable response to the Feingold diet in an earlier study were maintained on a strict elimination (Feingold) diet for 11 weeks, and given multiple trials of Placebo and challenge food materials in a double-blind challenge experiment. Challenge food items (candy bars and cookies) contained a blend of half the average daily intake of 27 mg of certified food colours.	Parental and teacher ratings, classroom behaviour observations and neuropsychological test scores obtained during baseline, Placebo and challenge conditions in general were not found to be adversely affected by the synthetic colour challenge materials.

Reference	Numbers of subjects, additive tested and dose levels	Outcome
Harley *et al.*, 1978b	36 school-age hyperactive boys under experimental (restricted Feingold K-P) and control diet.	Parent's behavioural ratings on 10 hyperactive children indicated a positive response to the experimental diet but laboratory observations showed no diet effect. It was concluded that teacher ratings, objective classroom and laboratory observational data, attention-concentration and other psychological measured yielded no support for the Feingold hypothesis.
Levy *et al.* 1978	22 children (4-8 years) selected as hyperactive tested before and after 4 weeks on the elimination diet, after a Tartrazine (Yellow dye no 5) and Placebo challenge, and after a 4 week wash-out period on the diet, by Conners parent-teacher ratings, objective tests of attention, standard perceptual motor tests and subtests from the Wechsler Intelligence Scale for Children (WISC).	Statistically significant improvement in the mother's ratings of the children's behaviour after the first 4 weeks of the diet, but the effects was not substantiated by the objective tests (teacher, clinician ratings). Tartrazine did not result in a statistically significant deterioration in the children's behaviour when they were challenged under double blind testing conditions with Tartrazine in any of the ratings (mothers, teachers, clinicians or as measured by objective tests).
Goyette *et al.* 1978	Two double blind challenge trials: 1) Challenge with Placebo or synthetic colours at one-half the average daily intake in sixteen diet responsive hyperkinetic children (4.7-11.8 years) all kept on an elimination diet. 2) Eight subjects (3.4-8.4 years) which met the full criteria for 25 percent reduction in symptoms on the diet and a behavioural and clinical diagnosis of hyperkinesis plus five other children who were either borderline responders or had fewer than the required symptoms placed on an elimination diet and challenged with synthetic food dyes. Parent ratings were limited to a 3-hour period immediately following the challenge.	1) Subjects demonstrated a 57 percent mean reduction in behaviour problems as rated by parents and a 34 percent reduction as rated by teachers when placed on the elimination diet. Double blind testing with challenge and Placebo materials revealed no significant challenge differences. The authors indicated that the ratings may have been insensitive due to the long time span of the observation period as compared to the short duration of the effect 2) subjects demonstrated a 45 percent mean reduction in behaviour problems when placed on the elimination diet. A significant challenge effect ($p<0.025$) was observed with more behavioural problems reported during the active challenge period as compared to the Placebo period.
Rose 1978	Two 8-year-old females who had been on a Feingold diet for a minimum of 11 months were studied in a double-blind Placebo controlled study. Data were obtained by trained observers in the subjects regular class settings.	It was concluded that there was a functional relationship between the ingestion of synthetic food colours and an increase in both the duration and frequency of hyperactive behaviours and such effects were absent upon Placebo exposure.
Swanson and Kinsbourne 1980	40 children (20 hyperactive responsive to stimulant mediation and 20 control average 10 years of age) given a diet free of synthetic food dyes and other additives for 5 days and on day 4 and 5 oral challenge at 10.00 am with 100 or 150 mg of a blend of nine food dyes or Placebo	Performance of the hyperactive children on paired-associate learning tests on the day they received the dye blend was impaired relative to their performance after they received the Placebo. The performance in the nonhyperactive group was not affected by the challenge with the food dye blend.

Reference	Numbers of subjects, additive tested and dose levels	Outcome
Weiss *et al.* 1980	22 Young children, maintained on a diet that excluded certain foods, were challenged intermittently with a blend of seven synthetic colours in a double-blind trial. Parents' observations provided the criteria of response.	One child responded mildly to the challenge and one, a 34-moth old female, responded dramatically.
Mattes JM, Gittelman R. 1981	11 Children on Feingold diet and responsive to colours in double blind cross over with order randomized challenged with food colouring at 13 mg mixture of all FDA approved synthetic food colours in proportions to reflect normal consumption patterns) and Placebo. Evaluations by parents, teachers and psychiatrists and psychological testing	No evidence of a food colouring effect
Kavale and Forness 1983	Meta analysis integrating findings from 23 studies testing the Feingold hypothesis.	The primary findings indicate that diet modification is not an effective intervention for hyperactivity.
Egger *et al.*, 1985	76 Selected overactive children were treated with an oligoantigenic diet. 28 Children who improved their behaviour on the diet completed a double-blind crossover Placebo-controlled trial in which foods thought to provoke symptoms were reintroduced.	62 Children improved on the oligoantigenic diet and a normal behaviour was achieved in 21 of these children. In the cross over study, symptoms returned or were exacerbated much more often when children were on active material than on Placebo based diet. Synthetic colours and preservatives were the most common provoking substances but no child was sensitive to these alone.
David, 1987	Double blind challenges with 50 mg or 250 mg Tartrazine or benzoic acid in 24 children (1.6 to 12.4 years) with history of behavioural adverse effects to these colours in clinical settings	In no child was any change in behaviour noted by the parents or the nursing staff after administration of Placebo or test substance.
Gross *et al.*, 1987	39 Children were given the Feingold diet for 1 week followed by administration for 1 week of food containing the synthetic additives and salicylates. All children were classified by public school psychologists as having moderate to severe learning disorders; 18 were also hyperkinetic and 17 were taking medication for motor restlessness. Three raters blind to the respective diets rated the children's behaviour (monitored by video taping for 4 minute intervals at mealtime) for motor restlessness, disorganized behaviour and misbehaviour.	No significant differences were found in behaviours during week 1 and 2. The authors conclude that the Feingold diet has no beneficial effect.
Rowe K.S., 1988	From 55 children who participated in a 6-week open trial of the Feingold diet, 8 of 14 suspected reactors were involved in a double blind Placebo controlled repeated measures study in which 50 mg doses of the azo dyes Tartrazine and Carmoisine were used.	For 2 children there was a clear association between the ingestion of both dyes and behavioural symptoms of irritation, restlessness and sleep disturbance.

Reference	Numbers of subjects, additive tested and dose levels	Outcome
Rowe K.S., 1988	Of 222 children referred for suspected hyperactivity, 55 were subjected to a 6 week trial of the Feingold diet. A double-blind crossover study, employing a single-subject repeated measures design was conducted using 8 of 14 children for which parents claimed that that a particular cluster of behaviours was associated with ingestion of foods containing synthetic colours. Subjects were maintained on a diet free from synthetic additives and were challenged daily for 18 weeks with Placebo or 50 mg of either tartrazine or Carmoisine each for 2 separate weeks.	Forty (72.7%) demonstrated improved behaviour and 26 (47.3%) remained improved following liberalization of the diet over a period of 3-6 months. In the double-blind crossover study two significant reactors were identified whose behavioural pattern featured extreme irritability, restlessness and sleep disturbance.
Kaplan *et al.*, 1989	24 Hyperactive preschool-aged boys Within subject cross-over design (3 weeks baseline, 3 weeks Placebo, 4 weeks experimental diet, low in simple sugars, and eliminating synthetic colours and flavours, chocolate, monosodium glutamate, preservatives, caffeine and any substances the family reported might affect the specific child (i.e. dairy).	According to parental reports (ten-item version of the Conners rating Scale, known as the Abbreviated Symptom Questionnaire (ASQ) asking about restlessness, impulsivity, disturbing other children, short attention span, fidgeting, distractibility, frustration, crying, mood changes and temper outbursts) more than half of the subjects exhibited a reliable improvement in behaviour and negligible Placebo effects. Several non-behavioural variables also tended to improve (halitosis, night awakenings, and latency to sleep onset).
Pollock and Warner, 1990	39 Children whose behaviour was observed by their parents to improve on a synthetic food additive free diet were included in a double-blind Placebo-controlled challenge. 19 Children completed the study. Synthetic food colours included in the study were 50 mg tartrazine, 25 mg sunset yellow, 25 mg Carmoisine, and 15 mg amaranth, all given in one capsule at breakfast.	In these 19 children who completed the study food colours were shown to have an adverse effect on a daily Conners' rating of behaviour, but most parents could not detect these changes.
Carter *et al.*, 1993	78 Children in clinical trial because of hyperactive behaviour placed on a few food items elimination diet. 19 of them (the ones for which foods or additives were disguised that reliably provoked behaviour) in a subsequent Placebo controlled double blind challenge protocol.	59 Children improved in behaviour. Crossover trial on the 19 children showed a significant effect for the provoking food to worsen ratings (by parents and other people with a role in child's care) of behaviour and to impair psychological test performance
Rowe and Rowe, 1994	200 Children assessed for suspected hyperactivity. For the main study 50 reactors plus 34 other children (23 suspected reactors, 11 uncertain reactors) and 20 control subjects aged 2 to 14 years were studied in a 21 day double blind Placebo-controlled repeated measures study in which each child was used as its own control. Placebo or one of six dose levels of Tartrazine (1,2,5,10,20,50 mg) was administered randomly each morning and behavioural ratings were recorded by parents at the end of each 24 h.	Parents of 150 children reported behavioural improvement with the diet and deterioration on the introduction of foods containing synthetic colouring. 24 children were clear reactors. Significant reactions at all dose levels and a dose response was obtained. It was concluded that behavioural changes in irritability, restlessness and sleep disturbance are associated with the ingestion of Tartrazine in some children.

Reference	Numbers of subjects, additive tested and dose levels	Outcome
Boris and Mandel, 1994	26 Children who met the criteria for ADHD were treated with a multiple item elimination diet and challenged with several foods, dyes and/or preservatives. A double blind Placebo controlled food challenge was completed in 16 of the 19 children responding favorably.	19 Children responded favorably. On open challenge all 19 children reacted to many foods, dyes and/or preservatives. In the double blind Placebo controlled study there was a significant improvement on Placebo days compared with challenge days. Atopic children with ADHD had a significantly higher response rate than the non-atopic group. It is concluded that the study shows a beneficial effect of eliminating reactive foods and synthetic colours in children with ADHD.
Schab and Trinh, 2004	Meta analysis: ten electronic databases were searched for double-blind Placebo controlled trials evaluating the effect of synthetic food colours. 15 Trials met the primary inclusion criteria.	Meta-analytic modelling determined the overall effect size of synthetic food colours on hyperactivity to be 0.283 (95% CI 0.079 to 0.488), falling to 0.210 (95% CA 0.007 to 0.414) when the smallest and lowest quality trials were excluded. Trials selected for responsiveness before enrolment demonstrated the greatest effects. The authors concluded that despite indications of publication bias and other limitations, this study is consistent with accumulating evidence that neurobehavioural toxicity may characterise a variety of widely distributed chemicals.
Bateman *et al.*, 2004	Children (3-years old) screened for hyperactivity (HA) and atopy (AT). Four groups selected (HA/AT, not HA/AT, HA/not-AT and not-HA/not-AT (n=277) and subjected to a diet eliminating synthetic colours and benzoate preservatives for one week. In subsequent 3 weeks within subject double blind crossover study with, in random order dietary challenge with a drink containing synthetic colourings (20 mg/day)(Sunset Yellow, Tartrazine, Carmoisine, Ponceau 4R; 5 mg of each) and sodium benzoate (45 mg/day)(active period), or a Placebo mixture. Behaviour was assessed by a tester blind to dietary status and by parent's ratings	Significant reductions in hyperactive behaviour during the withdrawing phase and significantly greater increases in hyperactive behaviour during the active than the Placebo period based on parental reports. Effects were not influenced by presence or absence of HA or AT. No significant differences detected based on testing in the clinic.

Initial studies, as summarized in Table 9, investigated the effects on behaviour of Feingold's diet (a diet without synthetic colours and flavours) under double-blind conditions in relatively small groups of hyperactive children (Conners *et al.*, 1976, Williams *et al.*, 1978; Kaplan *et al.* 1989; Harley *et al.*, 1978a). Several studies reported improved behavioural characteristics in part but not all children of their study population (Conners *et al.*, 1976, Williams *et al.*, 1978; Kaplan *et al.* 1989) whereas others did not provide support for the Feingold hypothesis (Harley *et al.*, 1978a). In other studies children identified as possible responders were challenged in double-blind studies (Harley *et al.*, 1978b; Goyette *et al.*, 1978; Mattes *et al.*, 1978; Levy *et al.*, 1978; Weiss *et al.*, 1980).These studies used generally only hyperactive and responsive children, small study groups and parental or teaching ratings, and reported either adverse or no effects. Generally all these studies did not study dose-response dependency and also did not link the adverse behavioural effect to a specific food additive.

The Panel noted that the findings from individual studies have in general not been conclusive. This has been ascribed by others to several logistic and methodological problems (Rowe and Rowe 1994 and references therein), including for example;

- the identification of children from heterogeneous populations;
- problems with dietary compliance;
- Placebo effects;
- the possible lack of inertness of the control substance;
- varying and imprecise diagnostic criteria for hyperactivity;
- doubts about the validity and reliability of behavioural outcome measures particularly those appropriate to the assessment of dye challenge effects;
- uncertainty about the detection of treatment effects when only a small number of children respond;
- confusion about suitable dosage levels of colourings for use in challenge trials, and
- lack of incorporating different dosages into the design.

The Panel also notes that a few controlled studies on the effects of ingested synthetic food colours on behaviour (Rose, 1978; Swanson, 1980; Mattes and Gittelman, 1981; Egger *et al.*, 1985; David, 1987; Rowe, 1988) have reported inconsistent effects on behaviour (changes or no change) after the dye challenge as compared to Placebo.

Only a few studies deal with the possible behavioural effects of specific, individual, food colours. A dose response curve was reported in a study with responsive children exposed to increasing doses of Tartrazine, and this study again revealed that some but not all of the children in a selected responsive study population were responders (Rowe and Rowe, 1994).

A more recent meta-analysis of double blind Placebo-controlled trials combined fifteen individual trials that met the inclusion criteria (Schab and Trinh, 2004). Meta-analytic modelling determined the overall effect size of the artificial food colours on the hyperactivity score to be 0.283 (CI 95%, 0.079 to 0.488), falling to 0.210 (CI 95%, 0.007 to 0.414) when the smallest and lowest quality trials were excluded. Trials that screened for responsiveness before enrolment demonstrated the greatest effects. The authors concluded that despite indications of publication bias and other limitations, this study is consistent with accumulating evidence that neurobehavioural toxicity may characterise a variety of widely distributed chemicals.

On the other hand the older meta-analysis performed by Kavale and Forness (Kavale and Forness 1983) integrating findings from 23 studies testing the Feingold hypothesis concluded that diet modification is not an effective intervention for hyperactivity

In addition to human behavioural studies, animal studies reporting effects of food colours on neurological and behavioural parameters have been reported (Vorhees *et al.*, 1983; Tanaka, 1994; Tanaka, 1996, Tanaka, 2006a; Tanaka, 2006b). However, since the relationship between these experimental parameters in animal studies and ADHD-like symptoms in humans remain unclear the present opinion does not take these animal data into account. Novel behavioural methods have been developed measuring the key ADHD behaviours in children and animal models of ADHD (e.g. Sagvolden, 2000, Sagvolden *et al.*, 2005) and it could be possible to use such methods to evaluate behavioural effects of doses of various synthetic colours and flavours in normal animals as well as animal models of ADHD in order to elucidate possible mechanisms and relations presently lacking in available studies of children.

In summary, the Panel considers that while a number of studies have reported a possible relationship between exposure of young children to synthetic food colours and other food additives and behavioural effects others have not identified an association between exposure to these substances and behavioural effects. The available literature is thus not consistent and does not allow a firm conclusion. The Panel notes that the majority of these studies have been conducted on children described as hyperactive or with a clinical diagnosis of ADHD, thus not being representative of the general population.

5. Possible mechanisms of action

One of the explanations for the suggested induction of neurobehavioural (ADHD-like) responses in children following exposure to synthetic food colours and other food additives, is a CNS-mediated origin. It is assumed that ADHD is linked to altered Central Nervous System (CNS) dopamine function, most likely mediated by faulty dopaminergic modulation of neuronal activity transmitted by the neurotransmitters glutamate and GABA (Sagvolden *et al.*, 2005).

These responses may have a genetic basis. ADHD is not a uniform disorder and it is evident that there are subgroups of affected children responding to different triggers with behaviour classified as ADHD. Genetic factors (Shoukri & Donner, 2007), such as serotonin receptor polymorphisms (Brookes *et al*., 2006), dopamine receptor (Mill *et al*., 2004) and other receptor polymorphisms may play a role (Polanczyk *et al*., 2007). Some of these observations have not been confirmed and may be population-specific findings (Curran *et al*. 2001).

The findings of the present study also suggest that certain genetic polymorphisms, specifically two genetic polymorphisms thought to impair histamine clearance, may result in possible differential sensitivity to the particular additives used in the study, although the increases in GHA scores were not limited to individuals with the specific polymorphisms measured in the study (COT, 2007). There were no associations between behaviour and other genetic polymorphisms investigated in the study, including genetic polymorphisms selected from the dopamine neurotransmitter systems, which have previously been implicated in ADHD (COT, 2007)

Another possible mechanism is a hypersensitivity reaction in a small, sensitive subgroup of the population. ADHD sufferers can react negatively to "allergenic" foods such as milk, egg, wheat etc. (Carter *et al.*, 1993; Boris & Mandel, 1994) and a connection with food allergies has been suggested (Marshall, 1989).

DISCUSSION

The study by McCann *et al*. (2007) reports effects of two combinations of Tartrazine (E102), Quinoline Yellow (E104), Sunset Yellow FCF (E110), Ponceau 4R (E124), Allura Red AC (E129), Carmoisine (Azorubine, E122) and sodium benzoate (E211) on children's behaviour, as measured by the Global Hyperactivity Aggregate (GHA) score, a novel metric developed by the researchers, combining behavioural and computer-based measures in one overall parameter. The Panel notes that the children who were included in the study were selected to represent a broad range of behaviour in the general population including children with normal activity through to those with high activity levels, but that children who were medicated for ADHD were not included.

A small but significant effect of Mix A on GHA scores were observed in 3-year old children (effect size 0.20 [95% CI 0.01 to 0.40], $p<0.05$), while Mix B did not produce a significant

change in the GHA scores for this group. In contrast, in 8- to 9-year old children, a significant effect of Mix A (effect size 0.14 [95% CI 0.03 to 0.24], $p<0.05$) and Mix B (effect size 0.17 [95% CI 0.06 to 0.28], $p<0.01$) was observed, but only when analysis was restricted to those children consuming at least 85% of drinks with no missing data. When all children that completed the study were taken into account in 8- to 9-year old children only Mix B had a significant effect on GHA scores (effect size 0.12 [95% CI 0.03 to 0.22] $p<0.05$), while Mix A had no significant effect. Thus, the statistically significant effects were not found for the same mixture in the two age groups. Overall the increases in the GHA scores observed in the study were small, ranging from 12 to 20% increase in GHA scores for the entire sample.

A statistical reanalysis of the data from the Southampton study was undertaken by the Panel. All individual behaviour variables (minus the individual baseline value for that variable) were reanalysed without normalisation, so that each subject served as its own control. This reanalysis was undertaken both at the level of the individual behavioural variables as well as on the aggregated scores.

For the primary analysis the Global Hyperactivity Aggregate (GHA) score was recalculated following the same steps as in the original analysis, except for the omission of the final re-normalisation step.

Based on these results it is concluded that the primary analysis with the recalculated GHA score led to broadly similar conclusions to that in the original paper by McCann *et al*, except for the following:

(1) The Mix A versus Placebo comparison was not statistically significant for the three year olds when all subjects were included (entire sample), while the significance for the $\geq$ 85% consumption and complete case groups was increased slightly;

(2) for the 8- to 9-years age group, the Mix A versus Placebo comparison was no longer statistically significant in any of the three consumption groups.

In addition the data were analysed on the basis of a modified GHA score in which the parental scores were not included. The results from this analysis did no longer reveal any statistically significant effects of Mix A or Mix B versus Placebo, except for Mix B in the 8- to 9-year old completers when both sexes are pooled ($p=0.042$).

A further analysis was carried out on the whole data set, comprising analysis of the single variables of parental scores, teacher scores and observer scores, and, in the case of 8- to 9-year old children, computer-based scores. There is a suggestion from these analyses that the statistically significant effects seen in the 3-year olds (Mix A versus Placebo) and in the 8-to 9-year olds (Mix B versus Placebo) are largely driven in the data by the parental scores and, in the older males in both comparisons, by the computer score.

The Panel noted that the main contributors to the GHA scores were the parental scores. The scores from teachers and independent observers showed little positive trend and were not a major component in the overall GHA scores. The use of the GHA metric does not therefore completely overcome the criticisms of the earlier Isle of Wight study (Bateman *et al.* 2004).

The Panel thus concludes that the McCann *et al.* study (2007) provides limited evidence that the two different mixtures of synthetic colours plus sodium benzoate tested had a small but statistically significant effect on behaviour in children from the general population excluding children medicated for ADHD, as measured by an aggregated score of behavioural effects.

The Panel considers that the clinical significance of the observed effects (a) for the individual children in the study and (b) for the population as a whole remains unclear, since the effects

were small in magnitude and it is not known whether these small changes in attention and activity would interfere with schoolwork and other intellectual functioning.

The conclusions of McCann *et al.* were restricted to the hypothesis that some synthetic colours or sodium benzoate (or both) in the diet resulted in increased activity scores in children and did not implicate these agents as causative agents in ADHD. The Panel agrees that an "elevated score of activity / inattention" is by no means equivalent or even indicative of ADHD. The clinical diagnosis of this condition requires impaired social and behavioural functioning and not merely an "elevated score of activity / inattention". Furthermore it is important to stress that ADHD is a condition with a multifactorial aetiology and exclusive focus on food additives may detract from the provision of adequate treatment for children with ADHD.

The Panel noted that some, but not all, earlier studies have also reported effects of certain food colours on child behaviour, the majority of these studies being conducted on children described as hyperactive or with a clinical diagnosis of ADHD.

A recent meta-analysis of double blind Placebo-controlled trials combined fifteen individual trials that met the inclusion criteria (Schab and Trinh, 2004). Meta-analytic modelling determined the overall effect size of the artificial food colours on the hyperactivity score to be 0.283 (CI 95%, 0.079 to 0.488), falling to 0.210 (CI 95%, 0.007 to 0.414) when the smallest and lowest quality trials were excluded. Trials that screened for responsiveness before enrolment demonstrated the greatest effects. The authors concluded that despite indications of publication bias and other limitations, this study is consistent with accumulating evidence that neurobehavioural toxicity may characterise a variety of widely distributed chemicals.

On the other hand the older meta-analysis performed by Kavale and Forness (1983) integrating findings from 23 studies testing the Feingold hypothesis concluded that diet modification is not an effective intervention for hyperactivity.

In the available studies, changes in behaviour, from either addition or withdrawal of additives from the diet, were not observed in all children, suggesting there may be a subpopulation of individuals who are sensitive to food additives in general or to food colours in particular. The findings of the present study suggest that certain genetic polymorphisms, specifically two genetic polymorphisms thought to impair histamine clearance, may result in possible differential sensitivity to the particular additives used in the study. The increases in GHA scores were however not limited to individuals with the specific polymorphisms measured in the study, and there were no associations between behaviour and other genetic polymorphisms investigated in the study (COT, 2007). The observed associations between polymorphisms in the histamine N-methyltransferase gene and the difference in behaviour with Mix A in 3-year olds and Mix A and Mix B in 8- to 9-year olds compared to Placebo, even if real and not merely chance effects, were not sufficiently strong that they could usefully be applied to identify at-risk groups or individuals (COT, 2007).

If a sensitive subpopulation does exist, it is not possible, from the currently available data, to assess the overall prevalence of such sensitivity and whether particular food additives may be implicated.

Based on surveys conducted from 2002 to 2005, the target colours are more frequently used in sweets but also occur commonly in soft drinks and benzoate is frequently present in soft drinks. Children consuming brightly coloured sweets may be exposed to levels comparable to those considered in the protocol of the McCann *et al*. study for one or more of the food colours studied. Comparable levels may also be reached in those children who consume brightly coloured soft drinks. The level of exposure to sodium benzoate is also likely to occur.

There are a number of uncertainties and limitations that are apparent from this new research, some of which are echoed from earlier research. These include:

- the limited consistency of the results with respect to age and gender of the children, the effects of the two mixtures of additives tested and the type of observer (parent, teacher or independent observer);
- the unknown clinical relevance of the novel metric, i.e. the GHA score;
- the unknown relevance of the small effect size (as was also seen in the meta analysis of earlier studies by Schab and Trinh (2004);
- the fact that the study has not been designed to identify the effects of individual additives;
- a lack of information on dose-response;
- the lack of a biologically plausible mechanism for induction of behavioural effects from consumption of food additives.

In the context of the overall weight of evidence, the Panel considers that the findings from the McCann *et al.* study are not sufficiently conclusive to be used as a basis for altering the ADI of the respective food colours or sodium benzoate.

CONCLUSIONS

The Panel concludes that the McCann *et al.* study provides limited evidence that the two different mixtures of synthetic colours and sodium benzoate tested had a small and statistically significant effect on activity and attention in children selected from the general population excluding children medicated for ADHD, although the effects were not statistically significant for the two mixtures in both age groups.

Since mixtures and not individual additives were tested in the study by McCann *et al.*, it is not possible to ascribe the observed effects to any of the individual compounds.

The clinical significance of the observed effects also remains unclear, since it is not known whether these small alterations in attention and activity would interfere with schoolwork and other intellectual functioning.

In the context of the overall weight of evidence and in view of the considerable uncertainties, such as the lack of consistency and relative weakness of the effect and the absence of information on the clinical significance of the behavioural changes observed, the Panel concludes that the findings of the study cannot be used as a basis for altering the ADI of the respective food colours or sodium benzoate.

DOCUMENTATION PROVIDED TO EFSA

1. Study details provided by Prof. J. Stevenson and the Food Standards Agency.
2. Connolly, A., Boylan, E., Gibney, M., Hearty, A., McKevitt, A. & Nugent, A., 2008. Unpublished document on the pattern of usage of a selected combination of food additives in Irish children. UCD Institute of Food and Health, University College Dublin.

REFERENCES

American Psychiatric Association Task Force on DSM-IV, 2000. Diagnostic and statistical manual of mental disorders: DSM-IV-TR (4th Ed.). Washington, DC: American Psychiatric Association.

Bateman, B., Warner, J.O., Hutchinson, E., Dean, T., Rowlandson, P., Gant, C., Grundy, J., Fitzgerald, C., Stevenson, J., 2004. 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*. 89, 506-511.

Biederman, J., Faraone, S.V., 2005. Attention-deficit hyperactivity disorder. *Lancet* 366, 237-248.

Boris, M., Mandel, F.S., 1994. Foods and additives are common causes of the attention deficit hyperactive disorder in children. *Annals of Allergy* 72, 462-468.

Brookes, K., Xu, X., Chen, W., Zhou, K., Neale, B., Lowe, N. *et al*., 2006. The analysis of 51 genes in DSM-IV combined type attention deficit hyperactivity disorder: association signals in DRD4, DAT1 and 16 other genes. *Mol Psychiatry* 11(10), 934-53.

Carter, C.M., Urbanowicz, M., Hemsley, R., Mantilla, L., Strobel, S, Graham, P.J., Taylor, E., 1993. Effects of a few food diet in attention deficit disorder. *Arch Dis Child* 69, 564-568.

Conners, C.K., Goyette, C.H., Southwick, D.A., Lees, J.M., Andrulonis, P.A., 1976. Food additives and hyperkinesis: A controlled double blind experiment. *Pediatrics* 58, 154-166.

COT Committee on toxicity, 2007. Statement on research project (T07040) investigating the effect of mixtures of certain food colours and a preservative on behaviour in children. http://www.food.gov.uk/multimedia/pdfs/committee/colpreschil.pdf

Curran, S., Mill, J., Tahir, E., Kent, L., Richards, S., Gould, A. *et al*., 2001. Association study of a dopamine transporter polymorphism and attention deficit hyperactivity disorder in UK and Turkish samples. *Mol Psychiatry* 6(4), 425-8.

David T.J., 1987. Reactions to dietary Tartrazine. *Arch Dis Child*. 62, 119-122

Egger, J., Carter, C.M., Graham, P.J., Gumley, D., Soothill, J.F., 1985. Controlled trial of oligoantigenic treatment in the hyperkinetic syndrome. *Lancet* 1, 540-545.

Feingold, B.F., 1975. Hyperkinesis and learning disabilities linked to artificial food flavours and colours. *Am. J. Nurs.* 75, 797-803.

FSA (2002**).** Survey of colours in sweets**.** Food Survey Information Sheets. No. 23/02 April 2002. Available at http://www.food.gov.uk/multimedia/pdfs/23sweets.pdf

FSA (2003a). Survey of colours in soft drinks. Food Survey Information Sheets. No.37/03 May 2003. Available at http://www.food.gov.uk/multimedia/pdfs/fsiscoloursinfoods032003.pdf

FSA (2003b). Diary survey of the intake of intense sweeteners by young children from soft drinks. *Food Survey Information Sheet* no. 36, 2003.

Gilsenan, M.B., Lambe, J., Gibney, M.J., 2002. Irish National Food Ingredient Database: application for assessing patterns of additive usage in foods. *Food Addit Contam.* Dec 19(12), 1105-15.

Goyette, C.H., Connors, C.K., Petti, T.A., Curtis, L.E., 1978. Effects of artificial colours on hyperkinetic children: a double-blind challenge study. *Psychopharmacology Bulletin* 14 (2), 39-40.

Gross, M.D., Tofanelli, P.A., Butzirus, S.M., Snodgrass, E.W., 1987. The effects of diets rich in and free from additives on the behaviour of children with hyperkinetic and learning disorders. *J. Am. Acad. Child Adol. Psychiat*. 26, 53-55.

Gueorguiva, R., Krystal, J.H., 2004. Move over ANOVA: progress in analyzing repeated measures data and its reflection in papers published in the Archives of general Psychiatry 61, 310-317.

Harley, J.P., Mattews, C.G., Eichman, P., 1978b. Synthetic food colours and hyperactivity in children: A double-blind challenge experiment. *Pediatrics* 61, 975-983.

Harley, J.P., Ray, R.S., Tomasi, L., Eichman, P.L. Matthews, C.G., Chun, R., Cleeland, C.S., Traisman, E., 1978a. Hyperkinesis and food additives: testing the Feingold hypothesis. *Pediatrics* 61, 818-828.

Kaplan, B.J., McNicol, J., Conte, R.A., Moghadam, H.K., 1989. Dietary replacement in preschool-aged hyperactive boys. *Pediatrics* 83, 7-17.

Kavale, A., Forness, S.R., 1983. Hyperactivity and diet treatment: a meta-analysis of the Feingold hypothesis. *J. Learning Disabilities* 16, 324-330.

Kissling, C., Retz, W., Wiemann, S., Coogan, A.N., Clement, R.M., Hunnerkopf, R., *et al.,* 2007. A polymorphism at the 3'-untranslated region of the CLOCK gene is associated with adult attention-deficit hyperactivity disorder. *Am J Med Genet B Neuropsychiatr Genet*.

Kuntsi, J. & Stevenson, J., 2001. Psychological mechanisms in hyperactivity: II. The role of genetic factors. Journal of Child Psychology and Psychiatry 42(2), 211-219.

Levy, F., Dumbrell, S, Hobbes, G., Ryan, M., Wilton, N., Woodhill, J.M., 1978. Hyperkinesis and diet: a double blind crossover trial with Tartrazine challenge. *Med J Austr*. 1, 61-68.

Mallinckrodt, C.H., Watkin, J.G., Molenberghs, G., Carroll Raymond, J., 2004. Choice of the primary analysis in longitudinal clinical trials. *Pharm Stst* 3, 161-169.

Marshall P., 1989. Attention deficit disorder and allergy: a neurochemical model of the relation between the illnesses. *Psychol Bull* 106(3):434-46.

Mattes, J.M., Gittelman, R., 1981. Effects of artificial food colourings in children with hyperactive symptoms. *Arch Gen Psychiatry* 38, 714-718.

McCann, D., Barrett, A., Cooper, C., Crumpler, D., Dalen, L., Grimshaw, K., Kitchin, E., Lok, K., Porteous, L., Prince, E., Sonuga-Barke, E., O'Warner, J., Stevenson, J., 2007. Food additives and hyperactive behaviour in 3-year-old and 8/9-year-old children in the community a randomized, double-blinded, placebo-controlled trial. *The Lancet,* Vol. 370, Issue 9598, 1560-1567.

Mill, J., Curran, S., Richards, S., Taylor, E., Asherson, P., 2004. Polymorphisms in the dopamine D5 receptor (DRD5) gene and ADHD. *Am J Med Genet B Neuropsychiatr Genet* 125(1), 38-42.

NIH, 1982. NIH consensus development conference: defined diets and childhood hyperactivity. *Clin Pediatr* 21, 627-630.

Overmeyer, S., Taylor, E., 1999. Annotation: principles of treatment for hyperkinetic disorders: practice approaches for the UK. *J Child Psychol Psychiatry* 40, 1147-1157.

Polanczyk, G., Zeni, C., Genro, J.P., Guimaraes, A.P., Roman, T., Hutz, M.H., *et al*., 2007. Association of the adrenergic alpha2A receptor gene with methylphenidate improvement of inattentive symptoms in children and adolescents with attention-deficit/hyperactivity disorder. *Arch Gen Psychiatry* 64(2), 218-24.

Pollock, I. & Warner, J.O., 1990. Effect of artificial food colours on childhood behaviour. *Archives Dis Child* 65, 74-77.

Rose, T.L., 1978. The functional relationship between artificial food colours and hyperactivity. *J. Appl Behav Anal* 11, 439-446.

Rowe, K.S., 1988. Synthetic food colourings and 'hyperactivity': a double blind crossover study. *Aust Paediatr J* 24, 143-147.

Rowe, KS and Rowe K.J. 1994. Synthetic food coloring and behavior: a dose response effect in a double-blind placebo-controlled, repeated-measures study. *J Pediatr.* 125, 691-698.

Sagvolden, T., 2000. Behavioral validation of the spontaneously hypertensive rat (SHR) as an animal model of attention-deficit/hyperactivity disorder (AD/HD). *Neurosci Biobehav Rev* 24.1, 31-39.

Sagvolden, T., Russell, V.A., Aase, H., Johansen, E.B., Farshbaf, M., 2005. Rodent models of attention-deficit/hyperactivity disorder. *Biol Psychiatry* 57:1239-1247.

Sagvolden, T., Johansen, E.B., Aase, H., and Russell, V.A., 2005. A dynamic developmental theory of Attention-Deficit/Hyperactivity Disorder (ADHD) predominantly hyperactive/impulsive and combined subtypes. *Behavioural and Brain Sciences, 28,* 397-419.

Schab, D.W., Trinh, N.T., 2004. Do artificial food colours promote hyperactivity in children with hyperactivity in children with hyperactive syndromes? A meta analysis of double-blind placebo-controlled trials. *J Dev Behav Pediatr* 25, 423-434.

Shoukri, M.M., Donner, A., 2007. Bivariate models for co-aggregation of dichotomous traits in twins. *Stat Med* 26(2), 336-51.

Swanson, J.M., Kinsbourne, M., 1980. Food dyes impair performance of hyperactive children on a laboratory learning task. *Science* 207, 1485-1486.

Swanson, J. M., Flodman, P., Kennedy, J., Spence, M.A., Moyzis, R., Schuck, S., Murias, M., Moriarity, J., Barr, C., Smith, M., Posner, M., 2000. Dopamine genes and ADHD. *Neuroscience and Biobehavioural Reviews* 24, 21-25.

Tanaka, T., 1994. Reproductive and neurobehavioral effects of Allura Red AC administered to mice in the diet. *Toxicology*, 92, 169-177.

Tanaka, T., 1996. Reproductive and neurobehavioral effects of Sunset Yellow FCF administered to mice in the diet. *Toxicol. Ind. Health*, 12, 69-79.

Tanaka, T., 2006a. Reproductive and neurobehavioural toxicity study of tartrazine administered to mice in the diet. *Food Chem. Toxicol.*, 44 (2), 179-187.

Tanaka, T., 2006b. Reproductive and neurobehavioural toxicity study of Ponceau 4R administered to mice in the diet. *Food Chem. Toxicol.*, 44 (10), 1651-1658.

Tennant D., 2006. Screening of Colour Intakes from Non-Alcoholic Beverages. Report prepared for the Union of European Beverages Associations (UNESDA), December 2006.

Thapar, A., Holmes, J., Poulton, K., Harrington, R., 1999. Genetic basis of attention deficit and hyperactivity. *British Journal of Psychiatry* 174, 105-111.

Vorhees, C. V., Butcher, R. E., Brunner, R.L., Wootten, V., and Sobotka, T.J., 1983. Developmental toxicity and psychotoxicity of FD and C red dye No. 40 (Allura Red AC) in rats. *Toxicology*, 28, 207-217.

Weiss, B., Williams, J.H., Margen, S., Abrams B., Caan B., Citron, L.J., Cox, C., McKibben, J., Ogar, D., Schultz, S., 1980. *Science* 207, 1487-1489.

Wender E.H., & Solanto M.V., 1991. Effects of Sugar on Aggressive and Inattentive Behavior in Children With Attention Deficit Disorder With Hyperactivity and Normal Children. *Pediatrics* 88, 960-966.

WHO, 2007. International Classification of Diseases-10. Chapter 5. Behavioural and emotional disorders with onset usually occurring in childhood and adolescence (F90-Hyperkinetic disorders). http://www.who.int/classifications/apps/icd/icd10online

Williams, J.I., Cram, D.M., Tausig F.T., Webster E., 1978. Relative effects of drugs and diet on hyperactive behaviours: An experimental study. *Pediatrics* 61, 811-817.

Wolraich M. L., Lindgren S.D., Stumbo P.J., Stegink L.D., Appelbaum M.I., and Kiritsy M.C., 1994. Effects of Diets High in Sucrose or Aspartame on The Behavior and Cognitive Performance of Children. *NEJM* 330:301-307.

APPENDICES

APPENDIX A

Details on Kinetics and metabolism data for substances investigated for neurobehavioural effects in a study by McCann *et al.*, Lancet. September 2007.

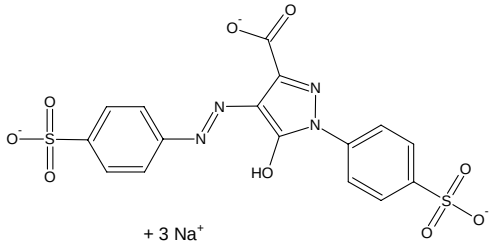
TARTRAZINE

CAS Registry number:	**1934-21-0**
EINECS number:	**217-699-5**
Colour Index number:	**19140**
E-number:	**E102**

Chemical name: Trisodium-5-hydroxy-1-(sulfonatophenyl)-4-(4-sulphonatophenylazo)-H-pyrazole-3-carboxylate

Chemical formula: $C_{16}H_9N_4Na_3O_9S_2$

Structural formula:



The JECFA evaluation (1966) describes several studies which have focussed on the toxicokinetic aspects of tartrazine. No further new literature has been published since except for some studies describing azoreduction by intestinal bacteria.

After intravenous injection of tartrazine in rats an average 1% of the dye was recovered from the bile. This low biliary excretion is believed to be associated with a free carboxyl group in the 3-position of the pyrazolone ring. After intraperitoneal administration of the colour a conjugated form of tartrazine was rapidly excreted in the urine. In both bile and urine no reduction products were detected. Based on these results the authors stated that tartrazine is a substituted phenylhydrazone rather than a true azo-compound (Ryan and Wright, 1961, 1962).

Rats were given tartrazine intraperitoneally and afterwards the urine was examined. The urine contained only the unchanged colour; conjugates or reduction products such as amines were absent. Rabbits were also given tartrazine by intraperitoneal injection and urine was examined. No details on the outcome of the study were however given.

After oral administration of tartrazine to rats, rabbits, and humans, sulfanilic acid (presumably the N-acetate) was found in the urine. As tartrazine was only reduced after oral administration it appeared that reduction was carried out by gastro-intestinal flora. Nonetheless, the authors concluded in favour of an absence of a true azo-linkage and mention that physical methods demonstrated tartrazine to be a keto-hydrazone tautomer (Wright, 1963; Jones *et al.*, 1964).

Rabbits were fed tartrazine and subsequently a 48 hours urine sample was analyzed. Besides the unchanged colour (1%) the metabolites sulfanilic acid (74%), and p-acetamido-benzenesulfonic acid (22%; the N-acetyl conjugate of sulfanilic acid) were identified in the urine sample. The percentages are the fraction of the maximal theoretical amount, which indicate that tartrazine is virtually completely reduced in the azo bond (Daniel, 1962).

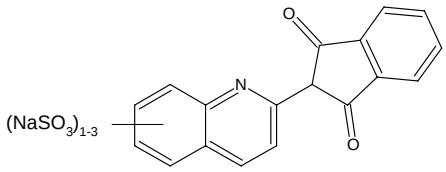
QUINOLINE YELLOW

CAS Registry number:	**8004-92-0**
EINECS number:	**305-897-5**
Colour Index number:	**47005**
E-number;	**E104**

Chemical name: disodium 2-(2-quinolyl)indane-1,3-dionedisulfonate (principle component)

Chemical formula: $C_{18}H_9NNa_2O_8S_2$ (principle component)

Structural formula:



In the JECFA evaluation of 1984 the subject of toxicokinetics is addressed for the first time.

Blood levels of radioactivity were measured in male and female rats after intra-gastric doses of 1 mg/kg bw ^{14}C-Quinoline Yellow. It appeared that the peak of radioactivity occurred between 0.5 and 1 hr. after dosing. All radioactivity was found in the plasma at a maximum concentration of 0.009% of the dose (no detail on how to interpret this percentage) and most of the radioactivity was bound to plasma proteins. No metabolites were found in the plasma. The kinetics of the blood levels fitted a two-compartment model with the following parameters: $Ta_{1/2}$ = 0.6 hour; $T1_{1/2}$ = 12 hours; and $T2_{1/2}$ = 70 hours. In a complementary study carcasses of male rats were dissected and residual tissue levels determined corresponding to 1/2, 1, 4, 8, 24 and 48 hours after dosing. The results confirmed that activity was selectively concentrated in the thyroid (LEMM, 1978).

Male rats received a single intra-gastric dose of 4 mg ^{14}C-Quinoline Yellow. About 94% of the radioactivity was recovered in the faeces (within 120 hours) and about 2% was eliminated in the urine. Retention was approximately 0.14%. The compound was found to be metabolized to only a small extent. In the urine 10-15% of the activity was associated with an unidentified metabolite. After 120 hours males were sacrificed and residual tissue levels were determined. The activity was selectively concentrated in the thyroid (no quantitative details) (LEMM, 1978).

In rats dosed with 2.85 mg/kg bw ^{14}C-Quinoline Yellow, after 31.5 hours only 1% of the dose was found to be excreted through the biliary route. No metabolites were found in the bile (LEMM, 1978).

After administering Quinoline Yellow to male rats whole body autoradiography demonstrated that after 1 hr. the activity was primarily associated with the gastro-intestinal tract and excretory organs. After 24 hours only the large intestine and, to a minor degree, the cortical zone of the kidney displayed activity (LEMM, 1978).

Tissue distribution studies after intra-gastric exposure of female rats to ^{14}C-Quinoline Yellow showed that the small proportion of the dose that was absorbed from the gastro-intestinal tract (estimated 3-4%) was primarily associated with the liver (max. 1 %), kidney (max 0.02%), and

bladder. Results expressed as concentration factors (radioactivity/g tissue) showed that a selective concentration of the thyroid persisted up to 48 hours, and a relatively high concentration was found in the ovaries in the first 24 hours (LEMM, 1978).

In dogs blood levels and excretion after intra-venous and intra-gastric administration of ^{14}C-Quinoline Yellow (0.2 and 0.44 mg/kg bw respectively) were examined. After intra-venous administration the disappearance of radioactivity corresponded to a two-compartment pharmaco-kinetic model with $T1_{1/2}$ = 4 hours and $T2_{1/2}$ = 43 hours. About 22% of the dose was excreted in the faeces. Intra-gastric administration showed that peak blood levels occurred at 1-4 hours after dosing. From 0-72 hours the urine contained 1-4% of the radio-label, 42-60% was excreted in the faeces within 72 hours. After both routes of administration there was no indication of specific tissue accumulation, particularly in the thyroid. Examination of urine, faeces, and plasma indicated that Quinoline Yellow is metabolized to only a small extend (LEMM, 1978).

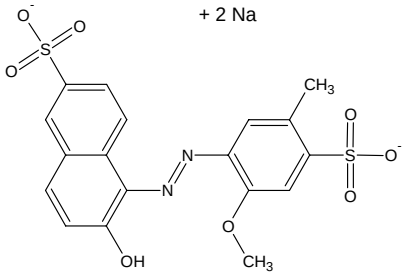
ALLURA RED AC

CAS Registry number:	**25956-17-6**
EINECS number:	**247-368-0**
Colour Index number:	**16035**
E-number:	**E129**

Chemical name: Disodium 2-hydroxy-1-(2-methoxy-5-methyl-4-sulfonatophenylazo) naphthalene-6-sulfonate

Chemical formula: $C_{18}H_{14}N_2Na_2O_8S_2$

Structural formula:



Three studies were reviewed by the JECFA (1980) to provide insight in the metabolic aspects of Allura Red AC.

Rats were fed a diet containing 5.19% of Allura Red AC (White, 1970). It was observed that 0.1% and 29% of the intact dye was excreted in the urine and faeces respectively. It was postulated that azo reduction by gut flora of the dye will yield the two components of the parent compound: 2-methoxy-5-methyl-aniline-4-sulfonic acid (cresidine-4-sulfonic acid), and 1-amino-2-naphthol-6-sulfonic acid (White, 1970)

In later studies, rats and dogs were pretreated daily with non-radioactive Allura Red AC. Subsequently, the animals were dosed with the ^{35}S-labelled compound and studied for up to 72 hours for excretion and distribution patterns of the colour. Both species showed limited absorption of the compound with the major route of excretion being via the faeces. In the dog 92 to 95% of the recovered radioactivity appeared in the faeces within 72 hours while in the rat 76 to 92% of the recovered radioactivity appeared in the faeces within this time period. Urinary recoveries of the colour varied between 5.7 and 19.8% and 2.7 and 3.6% in rats and dogs, respectively. After sacrifice, significant retention of radioactivity was located in the intestinal contents of both species and in the washed intestines of the rats. This was thought to be due to adhesion of the compound to the intestinal wall, since the total carcass and viscera of these animals contained less than 0.4% of the administered dose (Guyton & Reno, 1975).

Cresidinesulfonic acid was found to be the major metabolite of Allura Red AC in the urine of these two species, whereas the parent compound was not measurable. In addition, two other unidentifiable metabolites were found in the urine of the rats. In the rat faecal extracts, cresidinesulfonic acid was a major metabolite along with two unknowns and the parent compound. The dog faecal sample revealed an identical metabolite pattern as seen in the rat, and in addition, a third unknown was discovered. One of the urinary unknowns demonstrated an R_f value which was identical to that of the one of the faecal unknowns suggesting that they were one and the same. The other unknowns exhibited distinctive R_f values which indicated that these metabolites were different (Guyton & Stanovick, 1975).

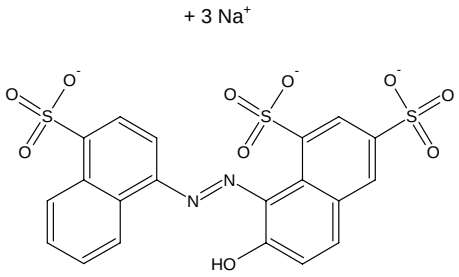
PONCEAU 4R

CAS Registry number:	**2611-82-7**
EINECS number:	**220-036-2**
Colour Index number:	**16255**
E-number:	**E124**

Chemical name: Trisodium 2-hydroxy-1-(4-Sulphonato-1-naphthylazo)-naphtalene-6,8-disulphonate.

Chemical formula: $C_{20}H_{11}N_2Na_3O_{10}S_3$

Structural formula:



The JECFA evaluation of 1983 gives a brief understanding of the biochemical fate of Ponceau 4R.

Single oral dose studies of uniformly ^{14}C-labeled Ponceau 4R of 0.5 or 50 mg/kg bw in rats, mice, and guinea-pigs show that substantially all of an orally administered dose of Ponceau 4R related material (e.g. ^{14}C-label) is excreted in the urine, bile and faeces, with the majority being accounted for in the faeces (90%; 25-35% unchanged); metabolites are found in the urine (mainly naphthionic acid) and faeces (naphthionic acid and 7-hydroxy-8-aminonaphtalene-1,3-disulfonic acid); and finally, apart from some retention in foetuses, there is no marked accumulation in any tissue. Only some Ponceau 4R was absorbed by isolated intestinal loops (Phillips *et al.*, 1982).

In a study in which rats received an intravenous dose of Ponceau 4R, 30-45% of the dye was excreted unchanged in the bile within six hours (Ryan and Wright, 1961).

Furthermore, it was found that after intraperitoneal administration of the dye the bile was coloured in mice and rats (Gaunt *et al.*, 1967).

Finally, a study by Walker (1968) indicates that Ponceau 4R is reduced *in vitro* by rat caecal contents.

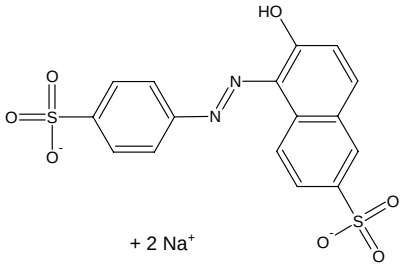
SUNSET YELLOW FCF

CAS Registry number:	**2783-94-0**
EINECS number:	**220-491-7**
Colour Index number:	**15985**
E-number:	**E110**

Chemical name: Disodium 2-hydroxy-1-(4-sulfonatophenylazo)naphthalene-6-sulfonate

Chemical formula: $C_{16}H_{10}N_2Na_2O_7S_2$

Structural formula:



The JECFA (1982) reports of five studies on the toxicokinetic aspects of Sunset Yellow FCF.

In a study in rats given a single oral dose of Sunset Yellow FCF 0.8% of the administered dose was recovered from the faeces as intact colour. In bile and urine these percentages were 3 and 0.8 respectively. In the urine of rats given large oral doses of Sunset Yellow FCF, the azo-reduction products sulfanilic acid and 1-amino-2-naphtol-6-sulfonic acid were found. No qualitative or quantitative measurement of reduction products in the faeces was carried out. From these results, in combination with observations after intravenous and intrasplenic administration, the authors concluded that breakdown of Sunset Yellow FCF to (sulfonated) aromatic amines is due to reduction by intestinal bacteria rather than by liver enzymes (Radomski and Mellinger, 1962).

Rats that received Sunset Yellow FCF by gavage excreted 0.3 % as intact colour and 37% as sulfanilic acid in the urine. In the bile 1.5% was excreted as intact colour (sulfanilic acid not measured). In the same study animals were gavage dosed with ^{14}C-Sunset Yellow FCF (labelled at the C-8 position of the naphthalene ring). As a result 94.5% of the total radioactivity was retrieved from the faeces, 8.5% from the urine. After the first 24 hours 1-2% of the total dose in urine consisted of intact dye and 40% of the dose consisted of the molar equivalents of sulfanilic acid of which 24% was *N*-acetylated. The other almost 60% of the radioactivity in the urine is unaccounted for (Honohan *et al.*, 1977).

After an intravenous injection of Sunset Yellow FCF in rats (no specification on dose) 20-30% of the dose was found in the bile after 6 hours (Ryan and Wright, 1961).

The urine of rabbits which were fed a single dose of Sunset Yellow FCF contained unchanged colour (2%), and the two azo-reduction products sulfanilic acid (54%), and 1-amino-2-naphtol-6-sulfonic acid (55% in 24 hours). In addition the *N*-acetylated form of sulfanilic acid, p-acetamidobenzene-sulfonic acid, was present in the urine (23%) (percentages indicate the ratio of the amount of the metabolite found to the theoretical amount, assuming complete breakdown) (Daniel, 1962).

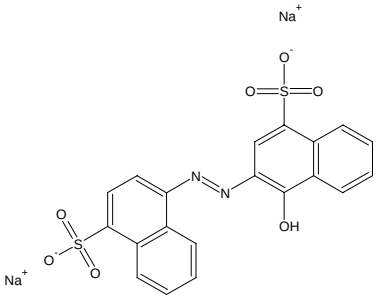
AZORUBINE (= CARMOISINE)

CAS Registry number:	**3567-69-9**
EINECS number:	**222-657-4**
Colour Index number:	**14720**
E-number:	**E122**

Chemical name: Disodium 4-hydroxy-3- (4-sulfonato-1-naphthylazo) naphthalene-1-sulfonate

Chemical formula: $C_{20}H_{12}N_2Na_2O_7S_2$

Structural formula:



JECFA (1983) describes several studies on the toxicokinetic aspects of Azorubine

Mice (CD-1) (3-6 males/group) received single doses of ^{14}C-Azorubine (5 µCi/mmol) by gavage (200 mg/kg bw, 6 µCi) or intravenously (200 mg/kg bw, 0.7 µCi).

After oral administration, peak levels of radioactivity occurred in plasma (0.08%/ml) and in the liver, lung, testes and spleen 8 hours after dosing. Radioactivity was almost completely excreted in faeces (74%) and urine (19%) within 16-32 hours.

After intravenous administration, most of the radioactivity (76%) was excreted in faeces (64%) and urine (12 %) 24 hours after dosing. The plasma ^{14}C-radioactivity decay curve indicated a very rapid distribution of the compound into the tissues ($t_{1/2}$ = 10 minutes) and an efficient excretion mostly through the gastrointestinal tract (92%) which was complete 48 hours after dosing (Galli *et al.*, 1981).

In a study in rats, animals (≥3 males/group) were given ^{14}C-Azorubine by gavage (200 mg/kg; 25 µCi), or by intravenous injection (200 mg/kg; 3 µCi). Radioactivity was measured in blood, tissue, faeces and urine 5, 10 and 30 minutes and 1, 2, 4, 8, 16, 32, 64 and 96 hours after dosing.

After gavage, no radioactivity was detected in the brain, adipose tissue, muscle, testes, spleen or lung (no time specification). After 32 hours, recovery of the administered radioactivity was 82% in faeces and 8% in urine. The radioactivity profile of the blood indicated rapid but poor absorption (maximum radioactivity content (0.01%/ml) being reached within 10 minutes).

After intravenous injection, the blood ^{14}C-radioactivity decay curve indicated rapid distribution to the tissues and could be described in terms of a two-compartment model. Radioactivity was highest in the gastrointestinal tract and liver (no time specification). However, within 24 hours after injection all radioactivity was recovered in faeces (79%) and urine (not specified). The large quantity present in the faeces was considered to indicate active excretion of Azorubine

and its metabolites in the bile and poor re-absorption from the intestine. Based on the blood-radioactivity curves after oral and intravenous administration, bioavailability of ^{14}C-Azorubine was calculated to be less than 10% (Galli *et al.*, 1982a).

Rats were given an intravenous injection with approximately 1 mg Azorubine. The 6 hours recovery of Azorubine in the bile was an average of 38% (30-40%) of the administered quantity (Ryan and Wright, 1961).

In another study, rats received 200 mg/kg bw ^{14}C-Azorubine (25 μCi) by gavage after which radioactive compounds in faeces and urine were investigated.

In addition to unmodified Azorubine, five radioactive compounds were present of which the predominant one was identified as naphthionic acid. After anaerobic incubation of ^{14}C-Azorubine with a bacterial suspension isolated from faeces from humans and rats, similar metabolic patterns were found (Marinovich *et al.*, 1983).

Wistar albino Rats (both sexes), MF-1 mice (male), and Dunkin-Hartley guinea-pigs (male) were administered a single oral dose of either 0.5 mg/kg or 50 mg/kg ^{14}C-Azorubine (20 μCi/kg).

In the first 24 hours the majority of radioactivity was excreted in the urine and faeces (18% and 73% in rats, 17% and 66% in mice, and 37% and 45% in guinea-pigs respectively). After 72 hours, substantially all of the radioactivity was recovered in the excreta, the majority being accounted for in the faeces.

In the urine of all three species 60 - 80% of the radioactivity was associated with naphthionic acid. Further, 10 - 20% of the radioactivity was identified as 2-amino-1-naphthol-4- sulfonic acid (2-ANS). In rats and mice ≥5% and in guinea-pigs 16% of the radioactivity was identified as 1,2-naphthoquinone-4-sulfonate (1,2-NQS). A fourth metabolite, accounting for 2 - 5% of the radioactivity in the urine, was not identified.

In the faeces of all three species naphthionic acid was also found; however, no 2-ANS or 1,2-NQS was detected. In addition, five unidentified metabolites (two hydrolysable by combined β-glucuronidase and sulfatase treatment) were found of which the proportions varied between species. No significant absorption of radioactivity from isolated small intestinal loops was noted in all three species.

Less than 0.03% of the administered radioactivity (50 mg/kg single dose) was recovered in the bile during 1 hour and only 0.04 - 0.7% during 5 hours. Less than 0.03% of the dose was eliminated as CO_2 (Phillips *et al.*, 1982).

Pregnant rats eliminated a single oral dose of ^{14}C-Azorubine (50 mg/kg at day 8 of pregnancy) at a similar rate to non-pregnant females. The concentration of radioactivity in the foetuses was similar to that in the other tissues (no further detail).

Pre-treatment of male rats with unlabelled Azorubine in the diet for 28 days (approximately 50 mg/kg bw/d) prior to dosing with ^{14}C-Azorubine (50 mg/kg bw), had no effect on the route of excretion or time of total elimination. The only difference compared to single dosed animals was that the proportion of metabolites extracted from the faeces differed (no further detail) (Phillips *et al.*, 1982).

In order to study the formation of (sulfonated) aromatic amines, the anaerobic reduction of Azorubine was investigated by incubating Azorubine with caecal content and hepatic microsomal fraction of rats.
Caecal suspension exhibited higher azo reductase activity than that of hepatic microsomal fraction. The researchers consider that the reductive ability through caecal flora signifies the formation of sulfonated aromatic amines which may be re-absorbed through the intestine to be

either eliminated through urine as conjugates or retained in the target tissues (Singh *et al.*, 1997).

Pregnant rats received 200 mg ^{14}C-Azorubine/kg bw by gavage on gestational days (GD) 16-19 and were sacrificed to analyse maternal tissues, amniotic fluid, placentas, foetal membranes and foetuses for radioactivity. Male rats were given a single oral dose of 200 mg ^{14}C-Azorubine/kg bw and sacrificed at different times after dosing.

In animals of both sexes, over 90% of radioactivity was excreted in faeces and urine within 64 h. This suggested that absorption of Azorubine is limited and that no significant accumulation occurred in any particular tissue. Of 5 metabolites determined, the principle one was identified as naphthionic acid. There was no evidence of transplacental transfer of ^{14}C-Azorubine or its metabolites. Results demonstrated that pregnancy does not affect the toxicokinetic profile of Azorubine (Tragni *et al.*, 1985).

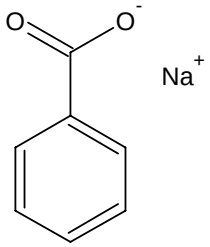
NA-BENZOATE / BENZOIC ACID

CAS Registry number:	**532-32-1**	**65-85-0**
EINECS number:	**208-534-8**	**200-618-2**
E-number:	**E211**	**E210**

Chemical name: sodium benzoate / benzoic acid

Chemical formula: $C_7H_6O_2$ / $NaC_7H_5O_2$

Structural formula:



(TAKEN FROM WHO, 2000; only information on oral absorption is included)

After oral ingestion of benzoic acid and sodium benzoate, there is a rapid absorption (of undissociated benzoic acid) from the gastrointestinal tract in experimental animals or humans (US FDA, 1972a, 1973). From the figures on excretion given below, 100% absorption can be assumed. In humans, the peak plasma concentration is reached within 1-2 h (Kubota *et al*., 1988; Kubota & Ishizaki, 1991).

In vivo dermal studies with benzoic acid in experimental animals (e.g., guinea-pigs, mice, rats, pigs, dogs, rhesus monkeys) confirm the results with humans (Hunziker *et al*., 1978; Andersen *et al*., 1980; Wester & Noonan, 1980; Bronaugh *et al*., 1982a; Reifenrath *et al*., 1984; Carver & Riviere, 1989; Maibach & Wester, 1989; Bucks *et al*., 1990). Absorption ranged from 25% in pigs (Reifenrath *et al*., 1984; Carver & Riviere, 1989) to 89% in rhesus monkeys (Wester & Noonan, 1980; Maibach & Wester, 1989; Bucks *et al*., 1990).

After oral and dermal uptake, benzoate is metabolized in the liver by conjugation with glycine, resulting in the formation of hippuric acid (Feldmann & Maibach, 1970; US FDA, 1972a; WHO, 1996; Feillet & Leonard, 1998). The rate of biotransformation in humans is high: after oral doses of 40, 80 or 160 mg sodium benzoate/kg body weight, the transformation to hippuric acid was independent of the dose -- about 17-29 mg/kg body weight per hour, corresponding to about 500 mg/kg body weight per day (Kubota & Ishizaki, 1991). Other authors obtained higher values of 0.8-2 g/kg body weight per day (US FDA, 1972a, 1973; WHO, 1996). Hippuric acid is rapidly excreted in urine. In humans, after oral doses of up to 160 mg/kg body weight, 75-100% of the applied dose is excreted as hippuric acid within 6 h after administration, and the rest within 2-3 days (Kubota *et al.,* 1988; Fujii *et al.,* 1991; Kubota & Ishizaki, 1991).

The limiting factor in the biosynthesis of hippuric acid is the availability of glycine. The utilization of glycine in the detoxification of benzoate results in a reduction in the glycine level of the body. Therefore, the ingestion of benzoic acid or its salts affects any body function or metabolic process in which glycine is involved; for example, it leads to a reduction in

creatinine, glutamine, urea, and uric acid levels (US FDA, 1972a, 1973; Kubota & Ishizaki, 1991; WHO, 1996).

Another metabolite of benzoate is the benzoyl glucuronide. For example, the dog excretes considerable amounts of this metabolite in the urine (20% after a single dose of 50 mg/kg body weight; Bridges *et al*., 1970). In other species, this metabolite appears only after higher doses of about 500 mg/kg body weight (see above) of benzoic acid or sodium benzoate, after depletion of the glycine pool (Bridges *et al*., 1970; US FDA, 1972a; Kubota *et al*., 1988). In cats, glucuronidation is generally very low (Williams, 1967).

In some species, including humans, minor amounts of benzoic acid itself are also excreted in the urine (Bridges *et al*., 1970; Kubota & Ishizaki, 1991).

Experiments on the distribution and elimination of ^{14}C-benzoate in the rat have shown no accumulation of sodium benzoate or benzoic acid in the body (US FDA, 1972a, 1973).

In the acid conditions of the stomach, the equilibrium moves to the undissociated benzoic acid molecule, which should be absorbed rapidly. Benzoate from sodium benzoate would change from the ionized form to the undissociated benzoic acid molecule. As a result, the metabolism and systemic effects of benzoic acid and sodium benzoate can be evaluated together.

(TAKEN FROM EFSA 2006)

Ring labelled ^{14}C-benzoic acid was given orally at doses in the range of 1 – 400 mg/kg bw to various species including primates, pigs, rabbits, rodents, cats, dogs, hedgehogs, bats, birds, and reptiles. Hippuric acid was the primary urinary metabolite in most species. The ornithine conjugate of benzoic acid, ornithic acid, was the major urinary metabolite excreted within 24 hours in chickens and reptiles. Benzoyl glucuronide was predominant in bats. In humans, >99% of ^{14}C was excreted as hippuric acid within 24 hours (Bridges *et al*., 1970).

Following oral administration of 375 mg [^{14}C]-benzoic acid/kg bw to rats, 91 – 94% of the radioactivity was recovered in the urine of rats after 72 hours, and only 1 – 6% was present in the faeces. The following metabolites were identified: hippuric acid (70.2 – 84.2%), benzoyl glucuronide (0.7 – 1.8%), benzoic acid (0.4 – 12.8%), and 3-hydroxy-3-phenyl propionic acid (0.1- 0.2%) (Nutley, 1990).

In order to investigate the types and quantities of beverages that increase urinary hippuric acid excretion, 137 healthy students were recruited and divided into quintiles based on their consumption of non-alcoholic beverages containing benzoic acid. HPLC was used to determine benzoic acid intake from beverages and urinary hippuric acid before, and 1.5 and 3 hours after consumption of various beverages. The range of benzoic acid in 13 beverages was 0 – 1.02 mg/ml and benzoic acid intakes from the beverages for groups 1 – 5, respectively, were: 0.4 mg ± 0.5; 23.4 mg ± 9.8; 55.2 mg ± 2.3; 76.3 mg ± 4.0; and 116.5 mg ± 16.5. Urinary hippuric acid geometric mean concentrations before consuming beverages in the five groups, respectively, were 0.276, 0.270, 0.207, 0.262, and 0.316 g/l; 1.5 hours after beverage consumption they were 0.210, 0.603, 1.026, 1.066, and 1.688 g/l and significantly increased ($p<0.001$) after adjustment for urinary hippuric acid before ingestion. Three hours after beverage consumption, urinary hippuric acid geometric mean concentrations in the five groups, respectively, were 0.160, 0.232, 0.306, 0.287, and 0.337 g/l ($p<0.001$). The authors concluded that beverages containing more than 100 mg benzoic acid may increase urinary hippuric acid significantly (Chang *et al.,* 2000).

Sodium benzoate

Male volunteers were given oral doses of 2000 to 5000 mg sodium benzoate. The 5000 mg dose group was given a 5000 mg dose of glycine one hour later and 2000 mg doses every two hours thereafter. Benzoate was excreted mainly as hippuric acid. No free benzoic acid was detected. Minor amounts of benzoyl glucuronide were detected at both doses. Co-administration of glycine with benzoate increased the rate of hippuric acid excretion, indicating that at high dose levels, glycine is rate limiting for formation of hippuric acid (Amsel & Levy, 1969).

After administration of oral doses of 40, 80, and 160 mg/kg bw of sodium benzoate to humans, the mean plasma AUCs of benzoic acid increased disproportionately to the dose, 3.7 and 12.0 times greater respectively for the higher dosages than for the lowest dose, while the mean AUCs for hippuric acid was proportional to dose. Peak plasma concentrations of benzoic acid increased with increasing dose, while peak hippuric acid concentrations did not change. The data suggest that the conjugation with glycine to form hippuric acid is a saturable process in humans (Kubota *et al*., 1988; Kubota & Ishizaki, 1991).

(TAKEN FROM SCF 1996; opinion released February 25, 1994)

Benzoate is a normal product of intermediary metabolism of phenylalanine and tyrosine and this results in human urinary excretion of a few tens of milligrams of benzoate/kg bw/day. Benzoate administered orally to man is rapidly absorbed and excreted in the urine within 14 hours. The main metabolite is its glycine conjugate, hippuric acid, with the glucuronyl conjugate and free benzoic acid as minor pathways of excretion. The rate limiting step in excretion of hippuric acid is the availability of glycine and this accounts for the glycine depletion which can occur when high doses of benzoate are administered. For example, in man the bolus dose of sodium benzoate causing 80% saturation of the maximal rate of hippuric acid secretion was found to be 28 mg/kg bw (expressed as benzoic acid).

REFERENCES TO CITED EVALUATIONS

For further data on the primary sources, the reader is referred to these reviews.

TARTRAZINE

JECFA (1966). WHO/FAO Joint Expert Committee on Food Additives. Specifications for identity and purity and toxicological evaluation of food colours 8th report. *WHO Food additives series***,** 66.25, 88-92

QUINOLINE YELLOW

JECFA (1984). WHO/FAO Joint Expert Committee on Food Additives 28th report. Toxicological evaluation of certain food additives and contaminants. *WHO Food Additives Series*, No. 19.

ALLURA RED AC

JECFA (1980). WHO/FAO Joint Expert Committee on Food Additives 24th report. Toxicological evaluation of certain food additives. *WHO Food Additive Series*, No. 15.

PONCEAU 4R

JECFA (1983a). WHO/FAO Joint Expert Committee on Food Additives 27th report. Toxicological evaluation of certain food additives and contaminants. *WHO Food Additives Series*, No. 18.

SUNSET YELLOW FCF

JECFA (1982). WHO/FAO Joint Expert Committee on Food Additives 26th meeting. Toxicological evaluation of certain food additives. *WHO Food Additives Series*, No. 17.

CARMOISINE / AZORUBINE

JECFA (1983b). WHO/FAO Joint Expert Committee on Food Additives. Toxicological evaluation of certain food additives and contaminants. *WHO Food Additives Series*, No. 18.

NA-BENZOATE / BENZOIC ACID

WHO (2000) Concise International Chemical Assessment Document 26 (CICAD). Benzoic acid and sodium benzoate. World Health Organization, Geneva, 2000.

SCF (1996).Opinion on Benzoic acid and its salts, expressed on 25 February, 1994. Reports of the Scientific Committee for Food (Thirty-fifth series).Office for Official Publications of the European Communities, Luxembourg

EFSA (2006) Opinion of the Scientific Panel on food additives, flavourings, processing aids and materials in contact with food (AFC) on a request from the Commission related to Flavouring Group Evaluation 20 (FGE.20): Benzyl alcohols, benzaldehydes, a related acetal, benzoic acids, and related esters from chemical group 23. Published 09/01/2006.

GLOSSARY / ABBREVIATIONS

ADI	Acceptable Daily Intake
AFC	Scientific Panel on Food Additive, Flavourings, Processing Aids and Materials in Contact with Food
bw	Body Weight
CNS	Central Nervous System
EFSA	European Food Safety Authority
FSA	UK Food Standards Agency
JECFA	Joint FAO/WHO Expert Committee on Food Additives
MPL	Maximum Permitted Levels
SCF	Scientific Committee for Food

European Food Safety Authority

ASSESSMENT METHODOLOGY UNIT

Parma, 11th March 2008
EFSA/ (2008)
EFSA/AMU/AFCA3(2008)

Statistical Report on the study by McCann et al. (2007) on the effect of some colours and sodium benzoate on children's behaviour

Assessment Methodology Unit
SCA Directorate

European Food Safety Authority

Table of contents

1 Introduction

On the 6th of September 2007, the Lancet published a paper by McCann *et al*. entitled "Food additives and hyperactive behaviour in 3 and 8/9 year old children in the community", reporting the results from a randomised, double-blind, placebo controlled, within subject cross-over trial testing for behavioural adverse effects of food colorant additives on children. This paper is based on a full study report produced by the School of Psychology (University of Southampton) for the U.K. FSA (Food Standard Agency).

The full study report was made available to EFSA (European Food Safety Authority) by the UK FSA on 19th September. At that time, an initial review was performed by the Assessment Methodology Unit in EFSA and a preliminary internal Statistical Review report was issued on the 26th of September 2007 which was presented to the working group during its first meeting on November 15th. Consequently and, to allow a more detailed evaluation of the study, additional clarification, information and data were requested. This information (the output of the analysis, the protocols of the experiment and the data and metadata) arrived at EFSA during October and November (last files received on November 30th).

The purpose of this Statistical Report is to provide a new and additional analysis of the data including descriptive statistics and model-based analyses, and finally to provide clear statistical conclusions to underpin the biological interpretation.

2 Overview of the methodology used in the McCann *et al.* paper

2.1 Experimental design of the study

The study design consisted of a randomised, double-blind, placebo controlled, within subject crossover trial with 153 children in a first group of 3 year old children and 144 in a second group of 8 to 9 year olds.

The treatments to be investigated consisted of two mixes of artificial food colours and additives (AFCA) plus sodium benzoate (treatment groups A and B) which were compared with a placebo (treatment group C). The composition of mixes A and B was as shown in **Table 2-1**.

Table 2-1 AFCA's used in the trial

AFCA	Mix A	Mix B
E110 (sunset yellow FCF)	5 mg	7.5 mg
E122 (carmoisine)	2.5 mg	7.5 mg
E102 (tartrazine)	7.5 mg	
E124 (ponceau 4R)	5 mg	
E211 (sodium benzoate)	45 mg	45 mg
E104 (quinoline yellow)		7.5 mg
E129 (allura red AC)		7.5 mg

The study duration was 7 weeks. During the study, each subject received each of treatments A, B and C for one week. The treatment allocation is shown in **Table 2-2**. Between each treatment period, there was a one-week Wash Out (WO) period. During the WO periods, the placebo (treatment C) was administered to all the subjects enrolled in the study. At week 0 of the trial all children received their normal diet to assess their baseline behaviour levels. From then on, all artificial food colours used during the trial as well as sodium benzoate were withdrawn from their diet for the duration of the study.

Children of each age group were randomly distributed over the 6 treatment sequences (**Table 2-2**).

Table 2-2 Schematic overview of the crossover trial

	Week 0	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
Sequence 1	Normal diet* (27,25***)	WO**	Mix A (27,24)	WO	Mix B (23,24)	WO	Mix C (21,23)
Sequence 2	Normal diet (25,24)	WO	Mix A (24,22)	WO	Mix C (23,21)	WO	Mix B (22,21)
Sequence 3	Normal diet (26,23)	WO	Mix B (25,22)	WO	Mix A (25,23)	WO	Mix C (20,21)
Sequence 4	Normal diet (24,24)	WO	Mix B (24,24)	WO	Mix C (21,21)	WO	Mix A (20,21)
Sequence 5	Normal diet (27,25)	WO	Mix C (24,21)	WO	Mix A (21,22)	WO	Mix B (23,22)
Sequence 6	Normal diet (24,23)	WO	Mix C (23,20)	WO	Mix B (19,20)	WO	Mix A (18,20)

*Normal diet to set baseline levels
** Wash Out period
*** Number of children with a GHA score (Number children in 3 Year Group, Number of children in 8-9 group)

2.2 The outcome measure used in the model

Behaviour, the outcome measure, was assessed using a Global Hyperactivity Aggregate (GHA) which was a combination of different component scores. The components measured, and hence the resultant GHA scores, differed between the two age groups.

2.2.1 Calculation of outcome variable for 3 Years Group

For every child for week i=0 to 6 there are three measured variables:

- Teacher score $week_i$
- Parent score $week_i$
- Observer score $week_i$

These were used to calculate following variables:

Z score Teacher $week_i$ = (Teacher score $week_i$ – Mean(Teacher score $week_0$))/Stdev(Teacher score $week_0$)

Z score Parent $week_i$ = (Parent score $week_i$ – Mean(Parent score $week_0$))/Stdev(Parent score $week_0$)

Z score Observer $week_i$ = (Observer score $week_i$ – Mean(Observer score $week_0$))/Stdev(Observer score $week_0$)

From there, a total Z score and then a GHA for $week_i$ was derived:

Total Z score $week_i$ = Mean(Z score Teacher $week_i$, Z score Parent $week_i$, Z score Observer $week_i$)

In order to calculate the GHA score each child must have an Observer score in $week_i$ plus either a Teacher score $week_i$ and/or a Parent score $week_i$

GHA score $week_i$ = (Total Z score $week_i$ – Mean(Z score $week_0$))/Stdev(Z score $week_0$)

2.2.2 Calculation of outcome variable for 8-9 Years Group

For every 8-9 year old child, for week i=0 to 6, there were seven measured variables:

- The three variables recorded for 3 year olds:
  - Teacher score $week_i$
  - Parent score $week_i$
  - Observer score $week_i$

- And four computer scores:
  - Comm score $week_i$
  - Hitrate score $week_i$
  - Dprime score $week_i$
  - Beta score $week_i$

These were used to calculate following variables:

Z score Teacher $week_i$ = (Teacher score $week_i$ – Mean(Teacher score $week_0$))/Stdev(Teacher score $week_0$)

Z score Parent $week_i$ = (Parent score $week_i$ – Mean(Parent score $week_0$))/Stdev(Parent score $week_0$)

Z score Observer week$_i$ = (Observer score week$_i$ – Mean(Observer score week$_0$))/Stddev(Observer score week$_0$)

Z score Comm week$_i$ = (Comm score week$_i$– Mean (Comm score week$_0$))/Stdev(Comm score week$_0$)

Z score Hitrate week$_i$ = (Hitrate score week$_i$ – Mean (Hitrate score week$_0$))/Stdev(Hitrate score week$_0$)

Zscore Dprime week$_i$ = (Dprime score week$_i$ – Mean (Dprime score week$_0$))/Stdev(Dprime score week$_0$)

Zscore Beta week$_i$ = (Beta score week$_i$ – Mean(Beta score week$_0$))/Stdev(Beta score week$_0$)

Total computer score at week$_i$:

Total comp score week$_i$ = (Z score Comm week$_i$ + Z score Hitrate week$_i$ + Zscore Dprime week$_i$ + Zscore Beta week$_i$)/4

Z total computer score week$_i$ = (Total comp score week$_i$ – Mean(Total comp score week$_0$))/Stdev(Total comp score week$_0$)

Total Z score and GHA at week$_i$:

Total Z score week$_i$ = Mean(Z score Teacher week$_i$, Z score Parent week$_i$, Z score Observer week$_i$, Z total computer score week$_i$)

For a valid GHA score each child must have an Observer score in week$_i$ plus at least two scores from Teacher score week$_i$, Parent score week$_i$ or Computer score week$_i$

GHA score week$_i$ = (Total Z score week$_i$ –Mean(Total Z score week$_0$))/ Stdev(Total Z score week$_0$)

2.3 Statistical approach used in the McCann et al. study

The data from the two age groups were analysed separately. Two linear mixed models were fitted for each age group. Although in the original Lancet paper (from the McCann *et al.* study) no details were given on the parameterization of the model, the rough output of the analysis requested and provided to the EFSA gave detailed information. The subject identification variable was set as a normal random effect,

whereas treatment was set as a fixed effect. All variables included in the different models and a data dictionary are listed in Appendix 1.

The compound symmetry covariance matrix was found to provide an appropriate fit for the models fitted to the data of the first age group, while an unstructured covariance matrix was chosen as appropriate for the second age group. The choice of covariance structure was based on conventional log-likelihood ratio comparisons under the fixed effects of Model 1.

The analyses were replicated for three sets of data: the full sample randomised and two subgroups: a high consumption group (data included if the child consumed ≥ 85% of drinks in each treatment week) and a high consumption completers group (high consumption and no missing GHA).

A statistical test for simple first-order carryover was done. No effect due to the type of challenge in the previous period on the current GHA could be demonstrated. From this, it was concluded that the wash out periods were sufficiently long to have prevented carry over effects.

Several issues concerning the statistical methodology used were noted. In particular, the approach used for score normalization is not completely appropriate as each subject anyway serves as his/her own control in such a design. For the same reason, the adjustment by between-subject covariates in Model 2 is completely redundant in this setting.

3 Methodology used for the re-evaluation of the data

3.1 General overview of the methodology

A new evaluation of the data and information collated in the McCann *et al.* study was necessary in order to address some of the issues identified. The re-analysis consisted of two parts. First, the authors' primary analysis was repeated, with minor changes to reflect a more appropriate statistical treatment and, second, a set of supplementary analyses were carried out.

For the primary analysis the Global Hyperactivity Aggregate (GHA) was recalculated following the same steps as in the original analysis, except for the omission of the final re-normalisation step, which was not seen as required. It was noted that some computer scores like the Commission Computer score, were negatively correlated with other individual scores (See **Figure 7-4**). So plausibly, this Commission Computer score or others like the Computer DPrime score should enter the aggregate score with a negative sign, but since there was no strong clinical evidence to do so, the signs were not reversed.

The remainder of the supplementary analyses consisted of the calculation of various descriptive statistics and formal analysis of each of the individual component measures, with the aim of understanding better the results observed from the overall GHA scores.

For all formal analyses, both primary and supplemental, a linear mixed model was used that was similar to that of the first analysis reported in the McCann *et al.* study. The model included only within-subject effects, namely those associated with the experimental intervention and periods in addition to treatment. A random subject effect was also included, and in this setting implies an analysis identical to that with a compound symmetry covariance structure. Subgroup analyses matched those of the original paper.

3.2 Dataset and results validation

The SPSS data files received from the authors of the McCann *et al.* study were converted to excel sheets for import into SAS, Access and S-Plus. To ensure high quality of all results presented in this report, several checking and validation procedures were performed.

- Standard checks were performed to ensure high quality of data import from SPSS to SAS and S-Plus formats used for the re-analysis.
- Recalculation of Z-scores to ensure that formulas were properly reported.
- Recalculation of summary statistics if the data as provided in the FSA report, in order to check the exclusion criteria and filters used to define consumption groups.
- A subset of the graphs was produced twice.

3.3 Definition of aggregated scores

To match as closely as possible the original authors' stated intentions, but at the same time conforming to acceptable statistical reasoning, the aggregate scores were defined as similarly as possible to those used originally in the McCann *et al.* study. The only modifications were the minor ones described in Section 3.1, *i.e.* using the following procedure:

- Each single variable was first adjusted to the individual baseline and then divided by the corresponding baseline STD.

- The overall computer score was defined as the average of 4 computer sub-scores (only for 8/9 year old children).

- The final aggregated scores were derived by averaging over the standardized single scores.

- No imputation was performed on missing component data.

- The same exclusion criteria were applied: each child required the Observer score, and only one missing score was allowed. As in the McCann *et al.* analysis, the final score was still computed as the average over the remaining single scores.

- The final re-normalization step performed on the Z scores in the Lancet paper was not reproduced.

3.4 Methodology for the descriptive re-analysis

For each age group, each consumption group, and each single (non-normalised) score, box plots were first displayed to compare responses between treatment groups and to visualize intra-group variability. In the box plots comparing treatment groups, all variables were adjusted to baseline for each child, meaning that the baseline score was subtracted from the behaviour scores for each child at each treatment. In addition, summary statistics of all data are provided.

Subsequently, and in a similar way to the single scores, summary statistics, histograms and box plots were displayed for both aggregated scores for each age group and each consumption group. For the sake of comparison, they were also derived for the aggregated score as calculated in the Lancet paper.

Graphs and plots were made using various routines from SAS and S-Plus software packages.

3.5 Methodology for the model-based re-analysis

3.5.1 Analysis of single scores

We first re-analysed all single component variables (minus the individual baseline value for that variable) without normalization, so that each subject serves as his own reference baseline. A similar mixed-effect model was fitted as in the McCann *et al.* paper, with the similar assumptions:

- A linear mixed model with random subject effects and fixed effects for treatment and week of treatment (being either 2, 4 or 6).
- No carry over effect assumed.
- Normality assumed for random effects and errors.
- No additional within-subject dependence.

Similarly to the model used in the McCann *et al.* paper, treatment contrasts were calculated with respect to the placebo mix (i.e. Mix A - Placebo and Mix B - Placebo). The period or "week" effect was regarded as a categorical variable and calculated with respect to the final week (week 6).

The tables displayed in the results sections report the two-sided P-values for each contrast (without any multiplicity corrections) and the corresponding fixed-effect estimates. Like in the McCann *et al.* paper, statistical significance means a P-value of less than 0.05.

Analyses were made:

- Independently for each age group and for all ages pooled together.
- Independently for each sex group and for both sexes pooled together.
- Note that, although pooling age groups leads to an increase of statistical power, it may also induce results which are difficult to interpret as the experimental protocols were not identical for the two age groups (e.g. different dosages in the mixes).

For all statistically significant comparisons (when P-value<0.05), the related data were plotted (treatment vs. placebo), where each data point corresponds to one child, in order to visualize how the children are scattered around the diagonal line.

Models were fitted using SAS Proc MIXED, and results were validated using S-plus.

3.5.2 Analysis of aggregated scores

The aggregated scores were analysed using the same linear mixed model as described in the previous section.

3.5.3 Additional analyses

In order to support further the AFC Panel Opinion, additional (*post hoc*) analyses have also been performed. They are shortly reported here, and include:

- Histograms by treatment group of Parent and Computer Commission scores, which were the two main variables that showed significant effects.
- The analysis of the new aggregate score but without the parental score, using the same modelling assumptions and method (3 years olds must have an Observer and Teacher score and 8-9 years must additionally have at least one Computer score to be included).
- The investigation of the interaction effect Baseline*Treatment, for each single variable. This was performed by adding as fixed effects the Baseline and Baseline*Treatment factors in the model used for the analysis of the single variables.

4 Results of datasets and results validation

This section reports on the results of the datasets validation. It was noted that:

- All aggregated scores as computed in the McCann *et al.* study report could be re-produced.
- Import and data manipulation were checked and validated.
- However summary tables provided by the UK FSA could not be fully reproduced as described in **Table 4-1**.

Table 4-1 Mean GHA for 3 year olds and for 8/9 year olds by challenge type

		Mix A			Mix B			Placebo		
		N	**mean**	**Sd**	**N**	**mean**	**sd**	**N**	**mean**	**sd**
3 Year olds										
Entire sample	Report	131	-0.11	1.03	134	-0.14	1.03	129	-0.32	1.11
	EFSA	**135**	**-0.1**	**1.04**	**136**	-0.14	1.03	**132**	**-0.33**	**1.12**
	EFSA									
>85%	Attrition =1	131	-0.11	1.03	**133**	-0.14	1.03	**127**	-0.32	**1.1**
consumption	Report	104	-0.11	1.03	108	-0.15	1.07	99	-0.39	1.07
	EFSA	104	**-0.12**	**1.09**	**106**	**-0.2**	**1**	**106**	**-0.34**	**1.1**
	EFSA									
	Attrition =1	**102**	**0.11**	**1.09**	**106**	**-0.2**	**1**	**105**	**-0.35**	**1.09**
Complete case	Report	73	-0.14	1.04	73	-0.26	1.05	73	-0.44	0.98
	EFSA	73	-0.14	1.04	73	-0.26	1.05	73	-0.44	0.98
8/9 Year olds										
Entire sample	Report	132	0.25	0.97	133	0.33	1.1	127	0.19	1.03
>85%	EFSA	132	0.25	0.97	133	0.33	1.1	127	0.19	1.03
consumption	Report	104	0.26	0.93	112	0.32	1.09	103	0.19	1.04
	EFSA	104	0.26	0.93	112	0.32	1.09	103	0.19	1.04
Complete case	Report	91	0.27	0.92	91	0.35	1.08	91	0.19	1.06
	EFSA	91	0.27	0.92	91	0.35	1.08	91	0.19	1.06

This issue was reported to the authors' of the Southampton study who confirmed the typographic mistake in their report. This had not any consequence on the main results and conclusion of the study.

5 Results of the descriptive analysis of the aggregated scores

In the following, we used the following abbreviations for aggregated scores:

'newGHA' = the newly computed aggregated score
'Validagg' = the final score as computed in the McCann *et al.* study

5.1.1 Comparative summary statistics and histograms of aggregated scores

Summary statistics are first reported all GHA scores, by age group and by treatment.

Table 5-1 Mean, standard deviation and number of scores for children in the 3 year age group by treatment

		Treatment				
		Mix A	**Mix B**	**Placebo**	**no mix**	**washout**
Mean	newGHA	-0.06	-0.10	-0.18	0.00	-0.03
StdDev	newGHA	0.85	0.77	0.85	0.00	0.81
N	newGHA	135	136	132	153	408
Mean	validagg	-0.10	-0.14	-0.33	0.00	-0.04
StdDev	validagg	1.04	1.03	1.12	1.00	1.10
N	validagg	135	136	132	153	408

Table 5-2 Mean, standard deviation and number of scores for children in the 8-9 year age group by treatment

		Treatment				
		Mix A	**Mix B**	**Placebo**	**no mix**	**washout**
Mean	newGHA	0.04	0.07	0.01	0.00	0.04
StdDev	newGHA	0.34	0.37	0.35	0.00	0.33
N	newGHA	130	130	124	144	389
Mean	validagg	0.25	0.33	0.19	0.00	0.25
StdDev	validagg	0.97	1.10	1.03	1.00	1.09
N	validagg	132	133	127	144	394

In addition, for a better visualisation of the data, histograms of such aggregated scores are reported below, by treatment groups.

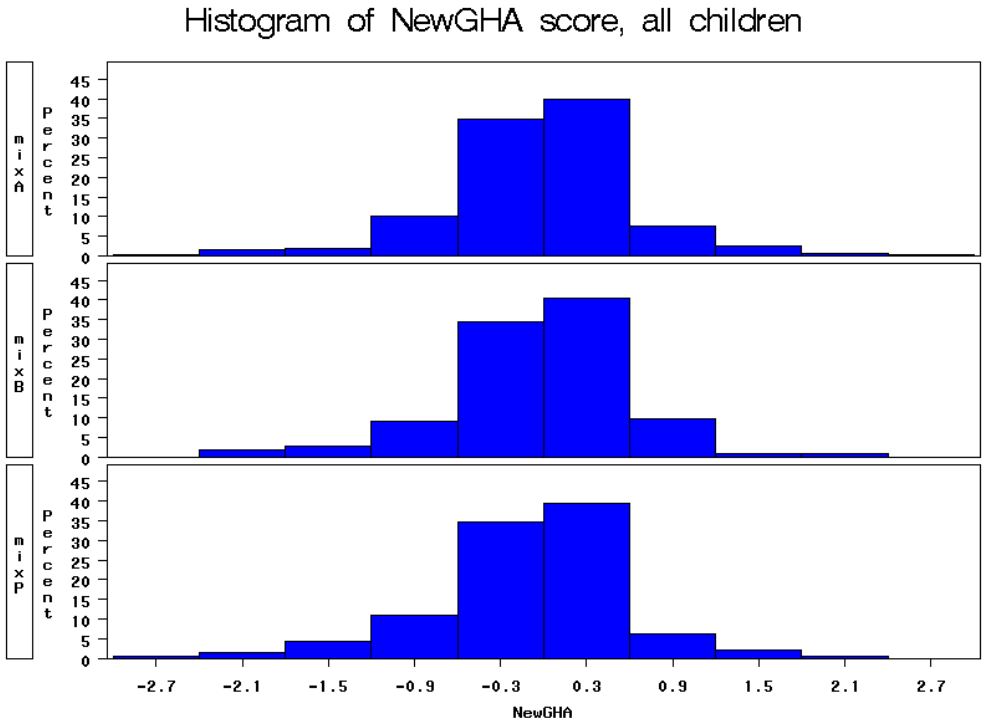


Figure 5-3 Histograms for the new aggregated score

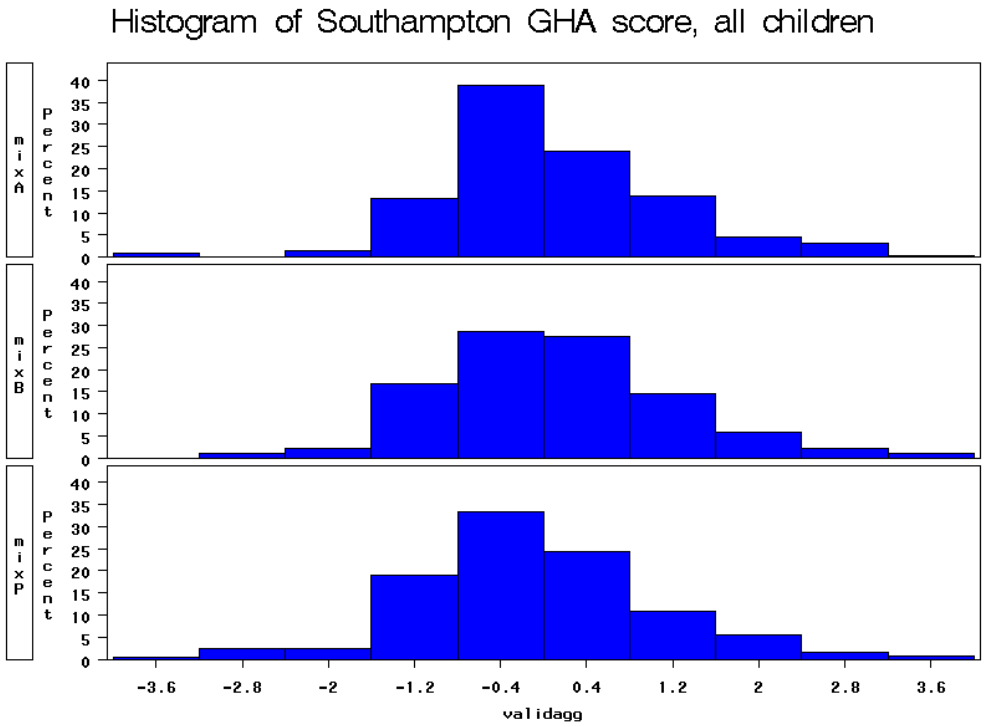


Figure 5-4 Histogram of original GHA as computed in the McCann *et al.* study

5.1.2 Box plots of the new aggregated GHA score

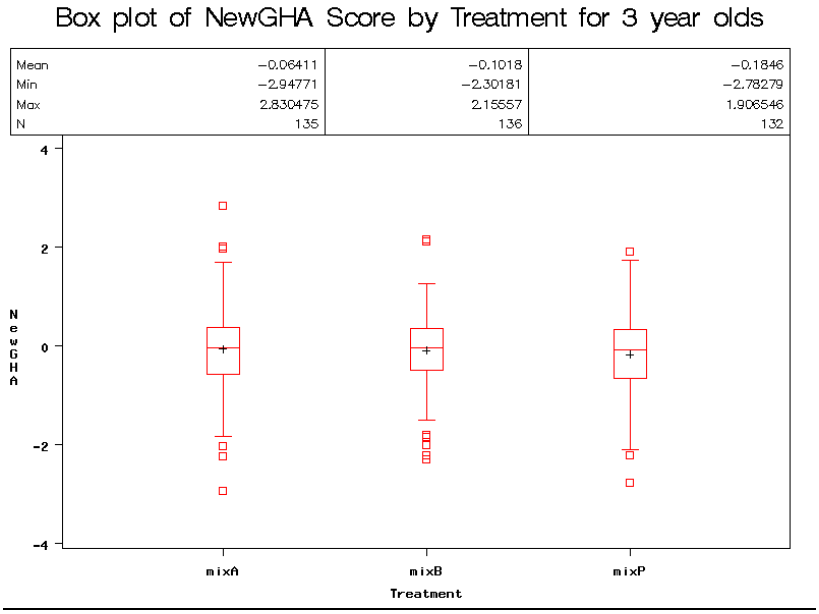


Figure 5-5 Box plot of the new aggregated scores (newGHA) by Treatment for 3 Year Olds

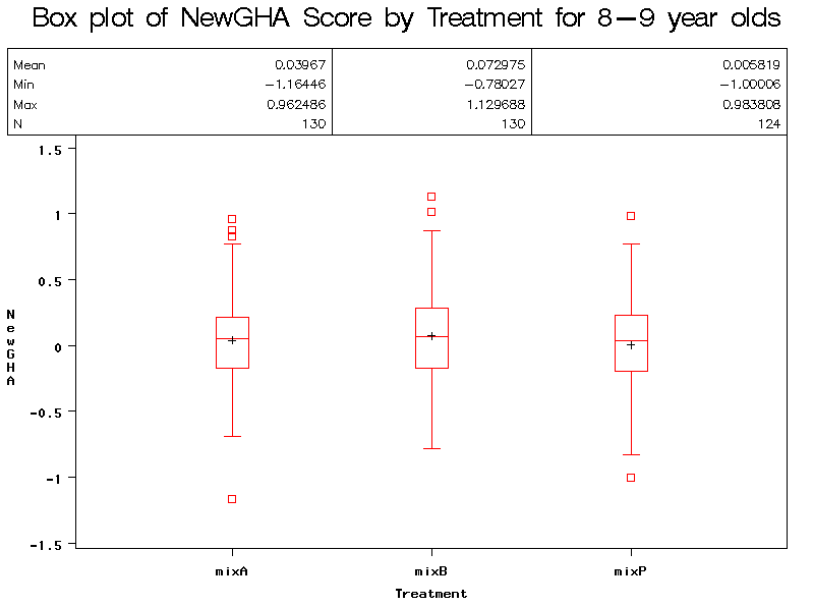


Figure 5-6 Box plot of the new aggregated GHA by Treatment for 8-9 Year Olds

6 Results of the model-based analysis of the aggregated scores

6.1 Results for the new aggregated scores

The tables below summarize all significant treatment group comparisons (**Table 6-1** and **Table 6-2**) and the treatment period (week) comparisons (**Table 6-3**) observed in the analysis of the newly computed GHA scores, by consumption groups. The complete list of outputs is reported in Appendix 2.

Table 6-1 Summary of all significant cases found in treatment group comparisons, for 'newGHA' for each age group (Alpha=0.05):

Test	YearGroup	sex	Estimate	StdErr	P-Value
Entire sample					
Mix B vs. Placebo	8/9Y	M	0.1115	0.04394	0.0124
Mix B vs. Placebo	8/9Y	both	0.05963	0.02981	0.0466
≥85% consumption					
Mix B vs. Placebo	8/9Y	both	0.08348	0.03276	0.0116
Mix A vs. Placebo	3Y	both	0.1962	0.08074	0.0161
Mix B vs. Placebo	8/9Y	M	0.1116	0.04644	0.0179
Complete case					
Mix B vs. Placebo	8/9Y	both	0.08546	0.03536	0.0167
Mix A vs. Placebo	3Y	both	0.2359	0.09764	0.0169
Mix B vs. Placebo	8/9Y	M	0.1118	0.04993	0.0273

Table 6-2 Summary of all significant cases found in treatment group comparisons, for 'newGHA' when age groups are pooled (Alpha=0.05):

Test	YearGroup	sex	Estimate	StdErr	P-Value
Entire sample					
Mix B vs. Placebo	both	both	0.07429	0.03765	0.0491
≥85% consumption					
Mix A vs. Placebo	both	both	0.1096	0.04304	0.0113
Mix B vs. Placebo	both	both	0.1059	0.04249	0.0131
Mix A vs. Placebo	both	F	0.1312	0.05972	0.0294
Mix B vs. Placebo	both	F	0.1224	0.05935	0.0407
Complete case					
Mix A vs. Placebo	both	both	0.1211	0.04769	0.0115
Mix B vs. Placebo	both	both	0.1093	0.04771	0.0226
Mix A vs. Placebo	both	F	0.1443	0.06836	0.0365

Table 6-3 Summary of all significant cases found in Period (week) comparisons, for 'newGHA' (Alpha=5%):

Test	Year Group	sex	Estimate	StdErr	P-Value
Entire sample					
Week 4 vs. Week 6	both	M	0.1824	0.05636	0.0014
Week 4 vs. Week 6	8/9Y	both	0.09192	0.02972	0.0022
Week 4 vs. Week 6	both	both	0.1016	0.03795	0.0077
Week 4 vs. Week 6	8/9Y	M	0.108	0.04398	0.0154
Week 4 vs. Week 6	3Y	M	0.2467	0.1057	0.0212
≥85% consumption					
Week 4 vs. Week 6	both	M	0.2121	0.062	0.0008
Week 4 vs. Week 6	both	both	0.1225	0.04289	0.0045
Week 4 vs. Week 6	8/9Y	both	0.0938	0.03282	0.0047
Week 4 vs. Week 6	3Y	M	0.3267	0.1274	0.012
Week 4 vs. Week 6	8/9Y	M	0.1108	0.04722	0.0207
Week 4 vs. Week 6	3Y	both	0.1621	0.08156	0.0483
Complete case					
Week 4 vs. Week 6	both	M	0.1812	0.06638	0.007
Week 4 vs. Week 6	8/9Y	both	0.08335	0.0353	0.0193
Week 4 vs. Week 6	both	both	0.101	0.04771	0.0351

6.2 Summary of conclusions from the formal analysis of the GHA scores

The primary analysis with the recalculated GHA score led to broadly similar conclusions to that in the original paper by McCann *et al*, except:

- The Mix A versus Placebo comparison was not statistically significant for the 3 year olds when all subjects were included (entire sample), while the significance for the higher consumption and complete case groups was increased slightly, but not beyond 0.01.

- For the 8-9 years age group, the Mix A versus Placebo comparison was no longer statistically significant in any of the consumption group.

The results for pooled age groups are tricky to interpret since the protocol and dosage were different between the groups. Nonetheless, they are reported here for information. The only consistency across consumption groups observed was for 8-9 year olds, for the Mix B vs. Placebo comparison, when both sexes are pooled. Finally, it should also be noted that the Placebo group was consistently showing a decreased GHA compared to Week 0 (baseline), such a decrease was significant in about half of the cases

7 Results of the descriptive re-analysis of the single scores

7.1 Global view of all data

7.1.1 Summary statistics of raw single scores

The summary statistics for each single variable are first reported for all groups pooled together, including values at wash-out weeks.

Table 7-1 Entire sample (both age groups and sexes including scores from washout weeks)

Variable	N	N Miss	Mean	Std Dev	Median	Minimum	Maximum
Teacher	1890	189	6.51	6.52	5.00	0	29.00
Parent	1911	168	15.11	8.19	18.00	0	35.00
Observer	1906	173	32.75	20.09	29.00	0	197.00
CPTcommision	902	1177	73.50	19.56	77.78	2.78	100.00
CPThitrate	902	1177	14.87	9.71	11.81	3.60	73.73
CPTdprime	902	1177	0.198	0.325	0.163	-0.613	2.12
CPTbeta	902	1177	1.02	2.70	0.834	1.92 E-7	79.14

Summary statistics are then displayed by age and treatment groups:

Table 7-2 Mean, standard deviation and number of scores for children in the 3 year age group by treatment

		Treatment				
		Mix A	Mix B	Placebo	no mix	washout
Mean	Teacher	6.14	6.41	5.81	7.34	6.40
StdDev	Teacher	6.09	6.14	5.73	6.78	6.21
N	Teacher	133	135	129	153	409
Mean	Parent	21.05	20.89	20.40	21.17	21.20
StdDev	Parent	4.01	3.84	3.85	2.27	3.76
N	Parent	136	138	137	153	423
Mean	Observer	26.18	24.87	24.87	25.87	25.76
StdDev	Observer	12.26	12.07	12.98	13.86	12.57
N	Observer	136	136	135	153	411

Table 7-3 Mean, standard deviation and number of scores for children in the 8-9 year age group by treatment

		Treatment				
		Mix A	Mix B	Placebo	no mix	washout
Mean	Teacher	6.25	6.59	6.52	6.74	6.68
StdDev	Teacher	6.64	6.73	6.75	7.19	6.83
N	Teacher	131	133	127	144	396
Mean	Parent	8.17	8.93	8.11	9.88	8.77
StdDev	Parent	6.40	7.15	6.72	6.78	7.02
N	Parent	129	131	127	144	393
Mean	Observer	40.98	41.76	40.30	37.24	40.42
StdDev	Observer	21.18	23.74	24.69	21.84	24.07
N	Observer	132	133	127	144	399
Mean	CPTcommission	73.60	73.52	72.09	75.52	73.31
StdDev	CPTcommission	18.31	19.82	21.19	17.79	19.88
N	CPTcommission	132	132	126	117	395
Mean	CPThitrate	16.21	16.50	15.42	10.42	15.00
StdDev	CPThitrate	9.89	10.98	9.61	5.19	9.92
N	CPThitrate	132	132	126	117	395
Mean	CPTdprime	0.19	0.20	0.21	0.16	0.20
StdDev	CPTdprime	0.30	0.31	0.37	0.29	0.33
N	CPTdprime	132	132	126	117	395
Mean	CPTbeta	0.93	0.96	0.92	0.82	1.16
StdDev	CPTbeta	0.58	0.59	0.52	0.59	4.03
N	CPTbeta	132	132	126	117	395

7.1.2 Scatter plot matrix of raw single scores

In order to visualize the 2-by-2 correlations between single scores, a multiple scatter plot is displayed, using the raw single scores for pooled age groups.

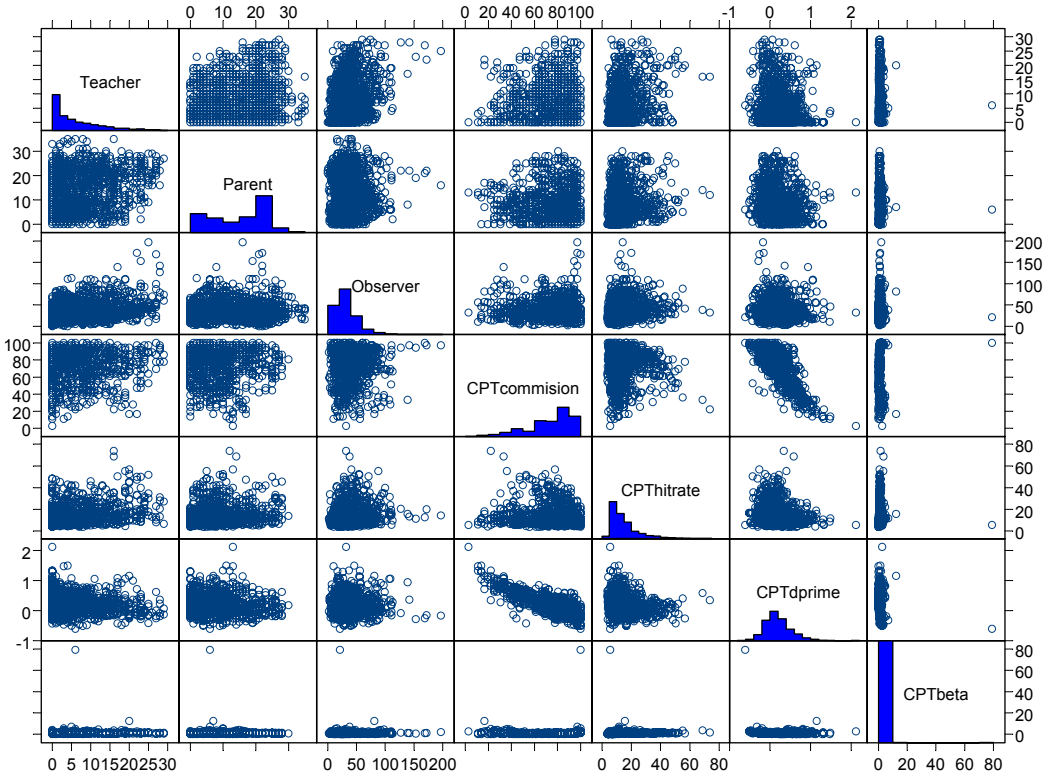


Figure 7-4 Scatter plot matrix of raw single scores

7.2 Box plots comparison of treatment groups

To visualize global group comparisons, box plots of single scores (adjusted to individual baseline) are displayed, by age group and consumption groups.

7.2.1 Box plots of single scores for 3-year-old children

7.2.1.1 Entire sample

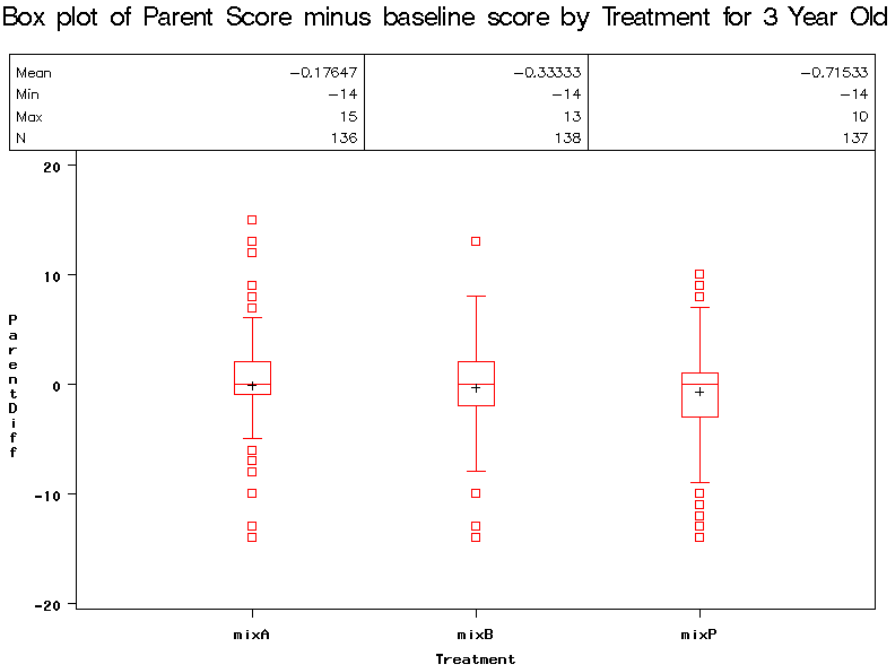


Figure 7-5 Box plot of Parent Score minus baseline score by Treatment for 3 Year Olds

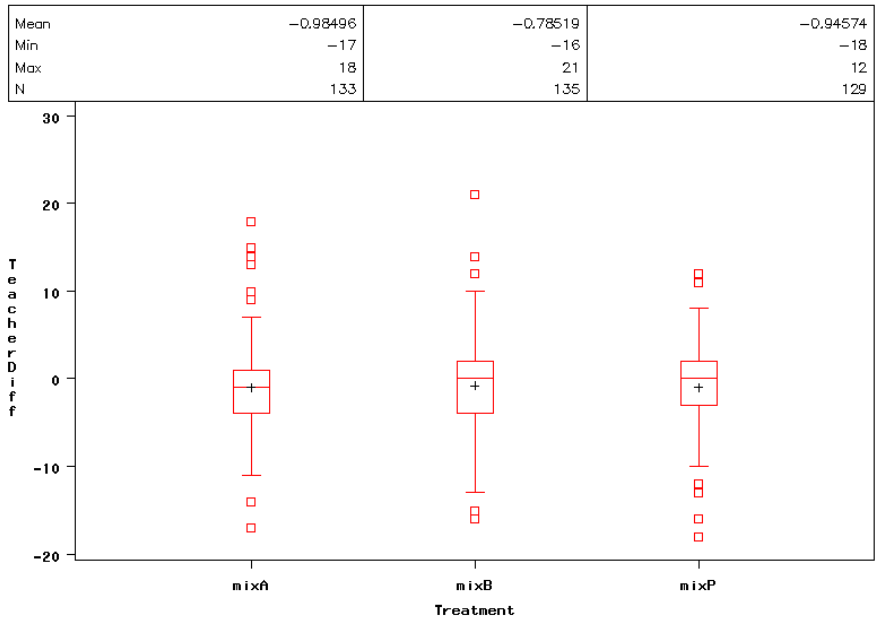


Figure 7-6 Box plot of Teacher Score minus baseline score by Treatment for 3 Year Olds

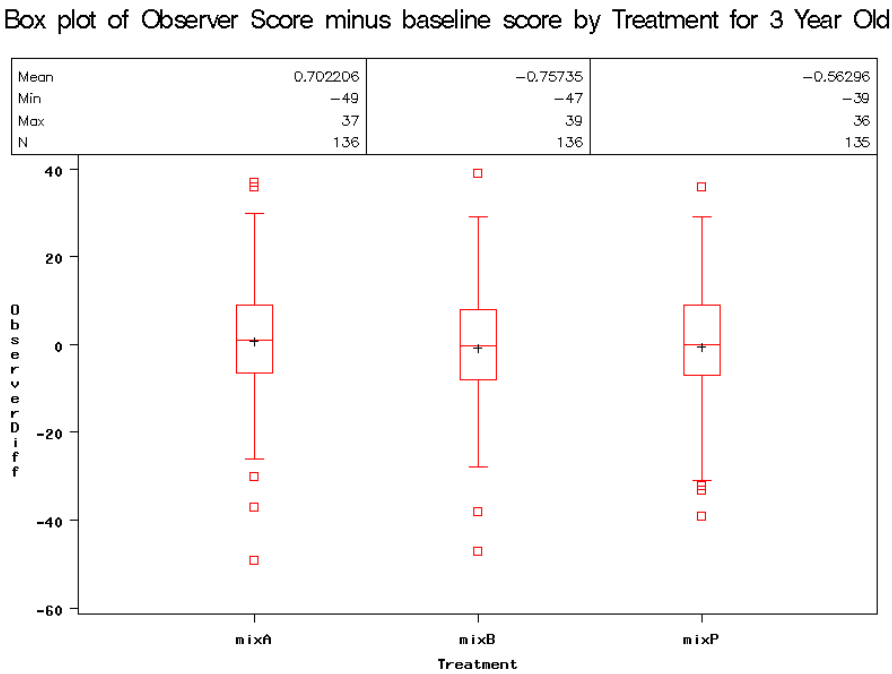


Figure 7-7 Box plot of Observer Score minus baseline score by Treatment for 3 Year Olds

7.2.1.2 High Consumers (Consumption ≥ 85%)

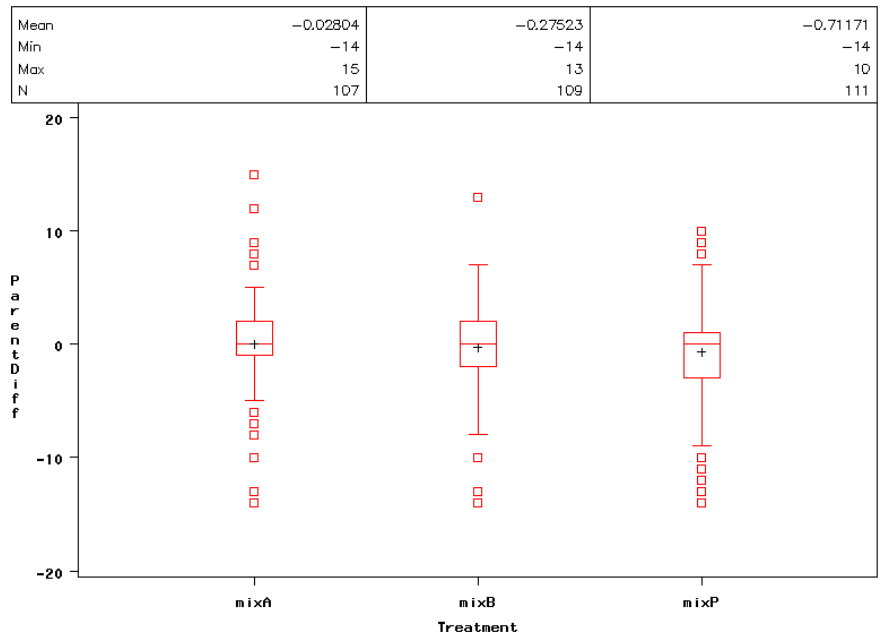


Figure 7-8 Box plot of Parent Score minus baseline score by Treatment for 3 Year Olds with >=85% consumption

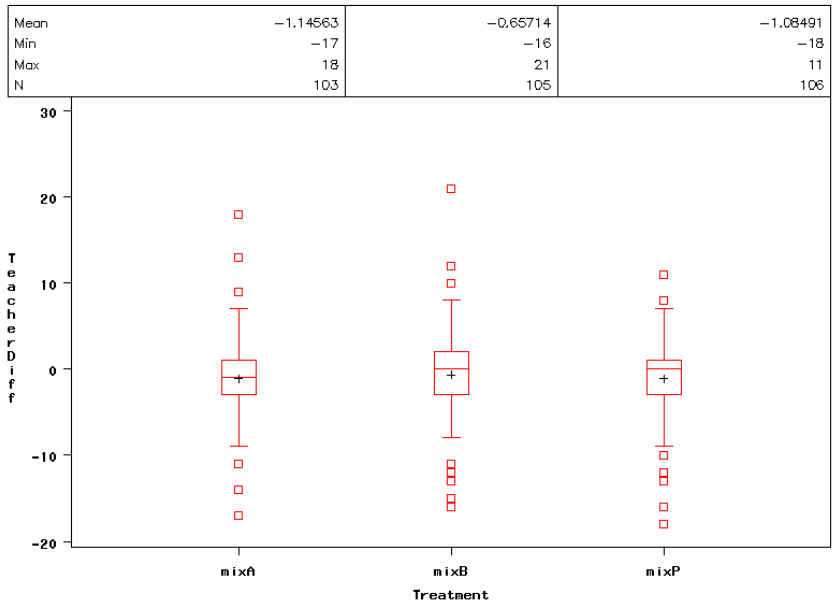


Figure 7-9 Box plot of Teacher Score minus baseline score by Treatment for 3 Year Olds with >=85% consumption

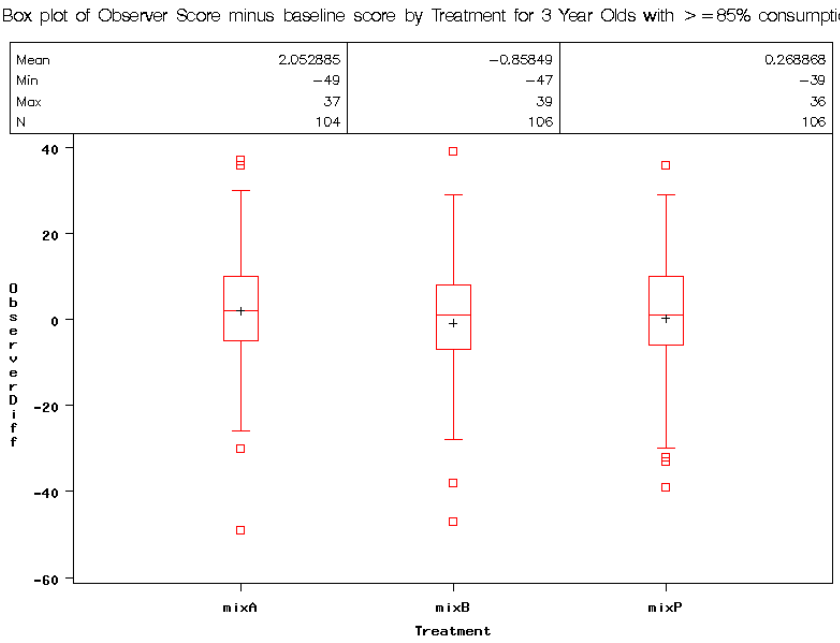


Figure 7-10 Box plot of Observer Score minus baseline score by Treatment for 3 Year Olds with >=85% consumption

7.2.1.3 Complete cases (case included if >=85% consumption in all challenge weeks and no missing GHA score)

Box plot of Parent Score minus baseline score by Treatment for 3 Year Olds , complete cases

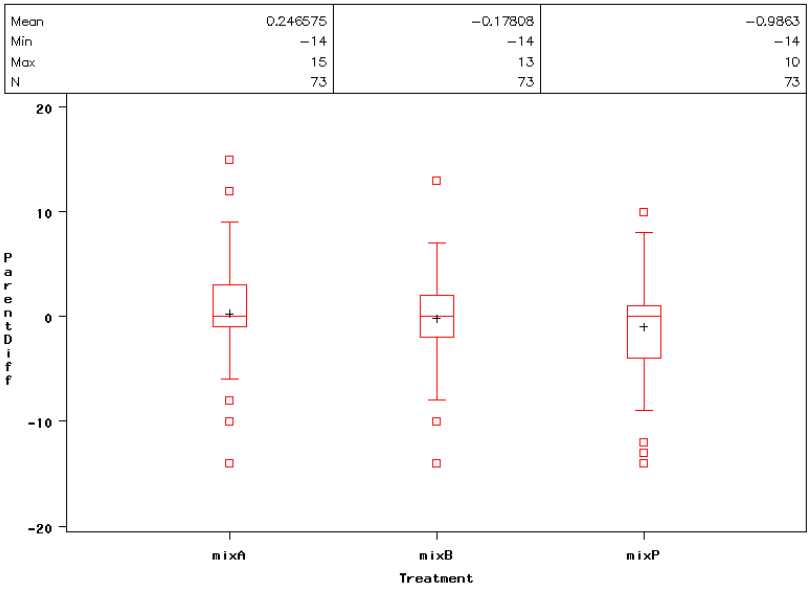


Figure 7-11 Box plot of Parent Score minus baseline score by Treatment for 3 Year Olds complete cases

Box plot of Teacher Score minus baseline score by Treatment for 3 Year Olds, complete cases

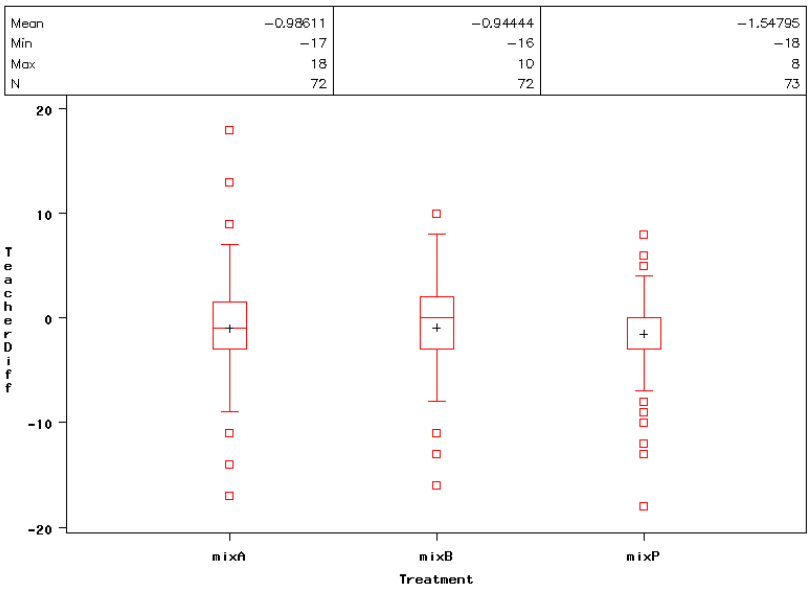


Figure 7-12 Box plot of Teacher Score minus baseline score by Treatment for 3 Year Olds complete cases

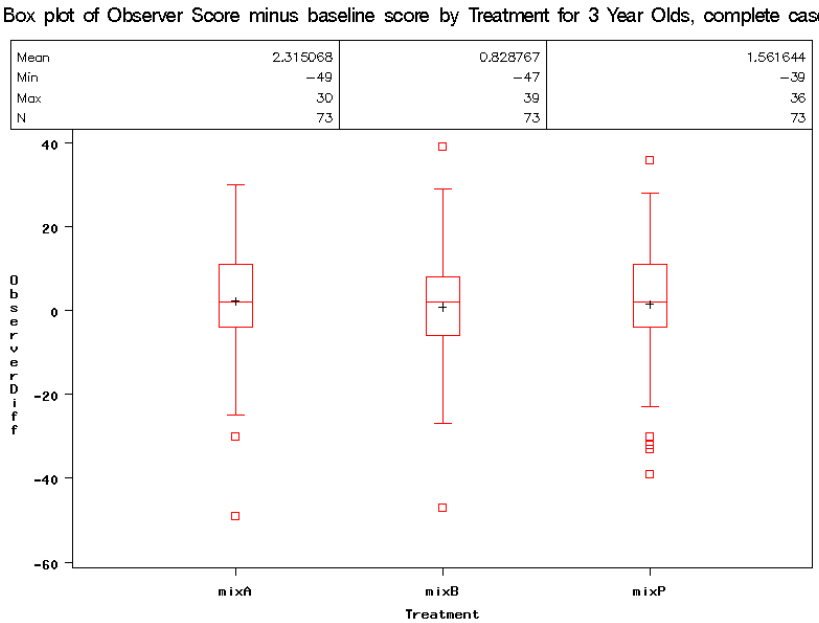


Figure 7-13 Box plot of Observer Score minus baseline score by Treatment for 3 Year Olds complete cases

7.2.2 Box plots of single scores for 8/9-year-old children

7.2.2.1 Entire sample

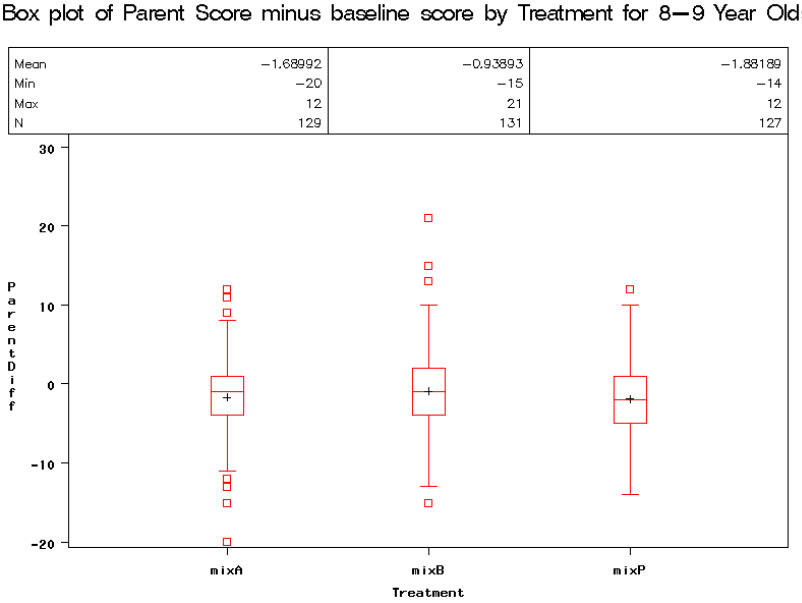


Figure 7-14 Box plot of Parent Score minus baseline score by Treatment for 8-9 Year Olds

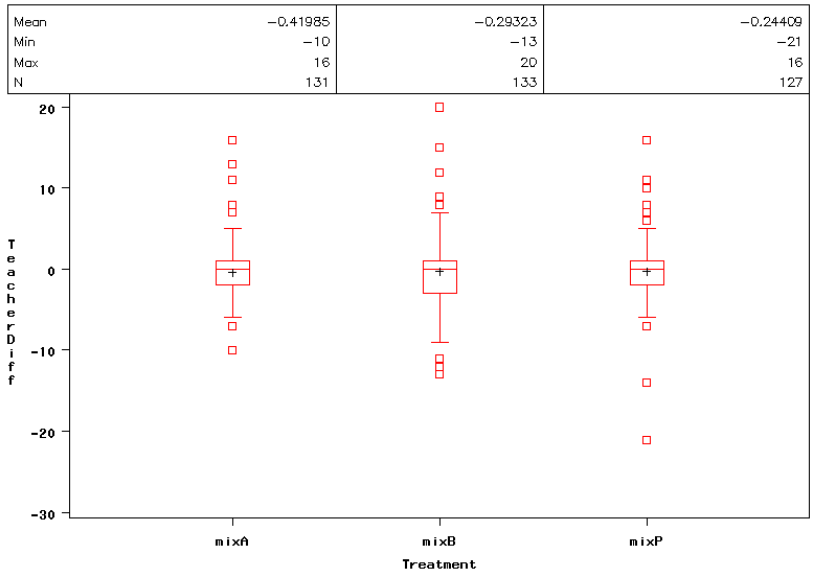


Figure 7-15 Box plot of Teacher Score minus baseline score by Treatmentfor 8-9 Year Olds

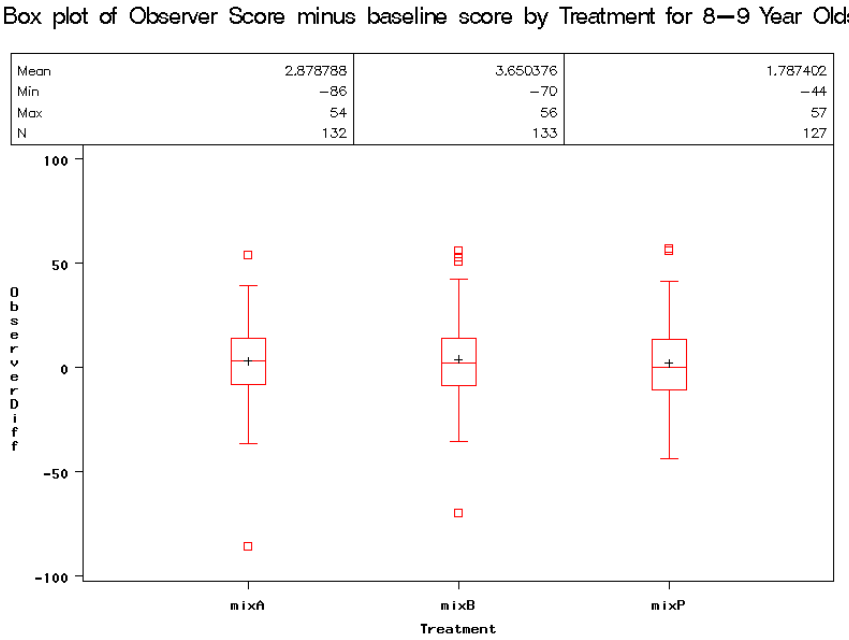


Figure 7-16 Box plot of Observer Score minus baseline score by Treatment for 8-9 Year Olds

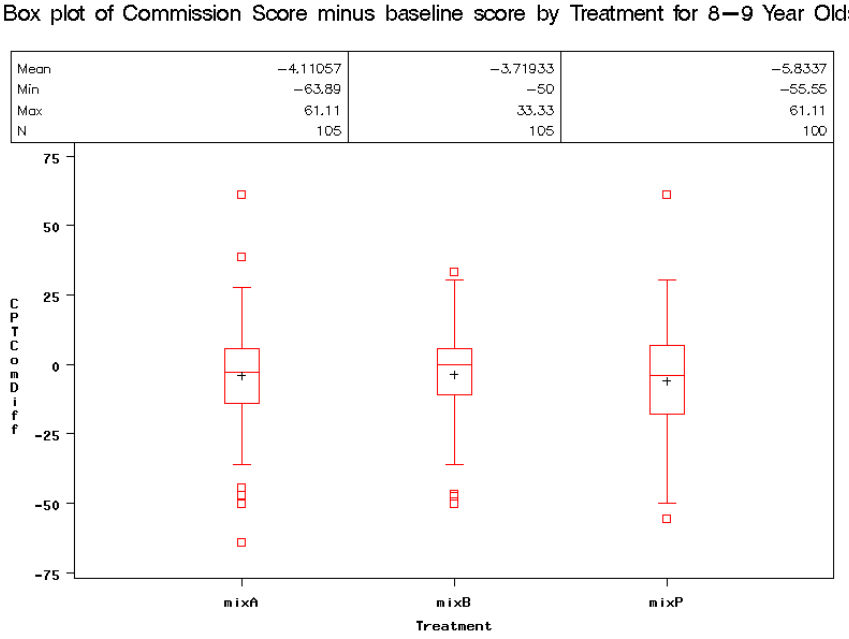


Figure 7-17 Box plot of Commission Score minus baseline score by Treatment for 8-9 Year Olds

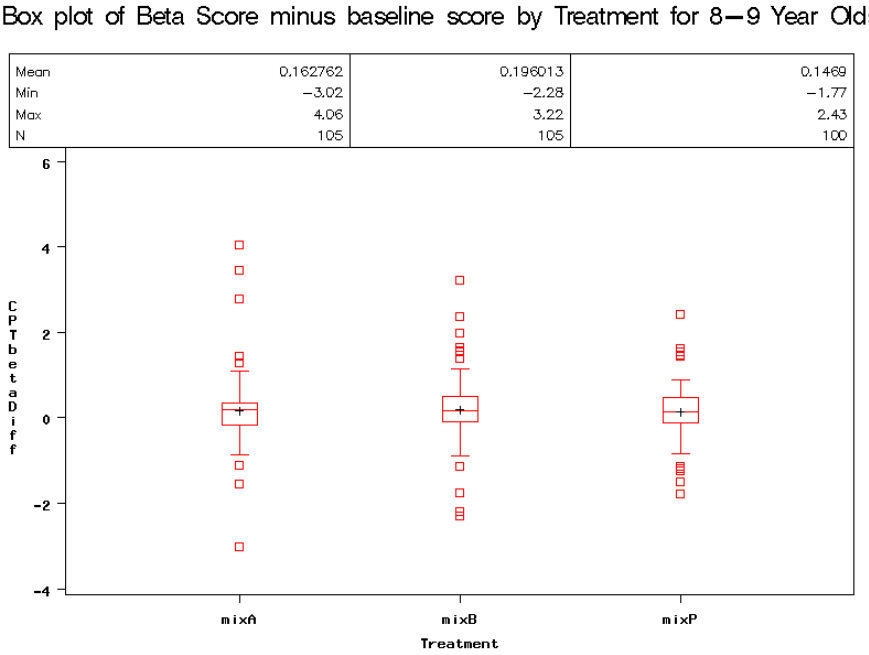


Figure 7-18 Box plot of Beta Score minus baseline score by Treatment for 8-9 Year Olds

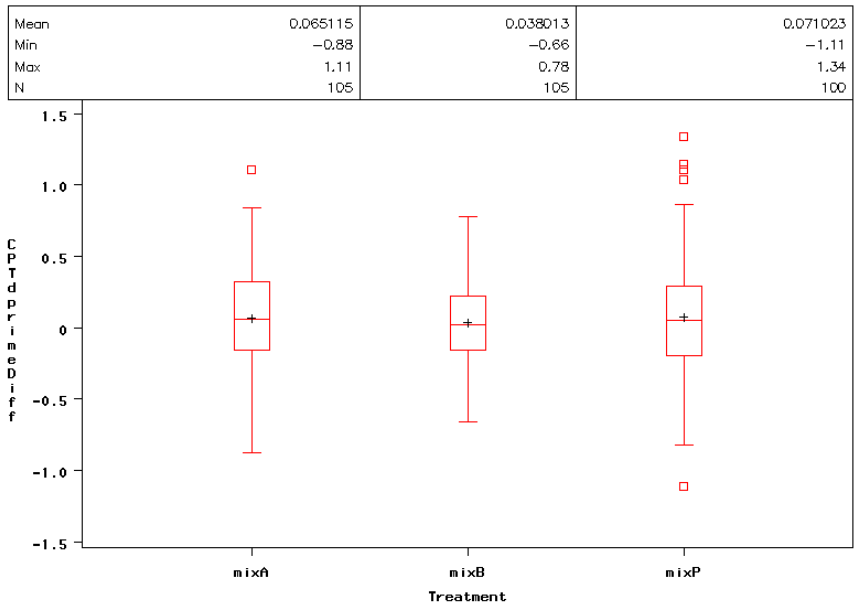


Figure 7-19 Box plot of Dprime Score minus baseline score by Treatment for 8-9 Year Olds

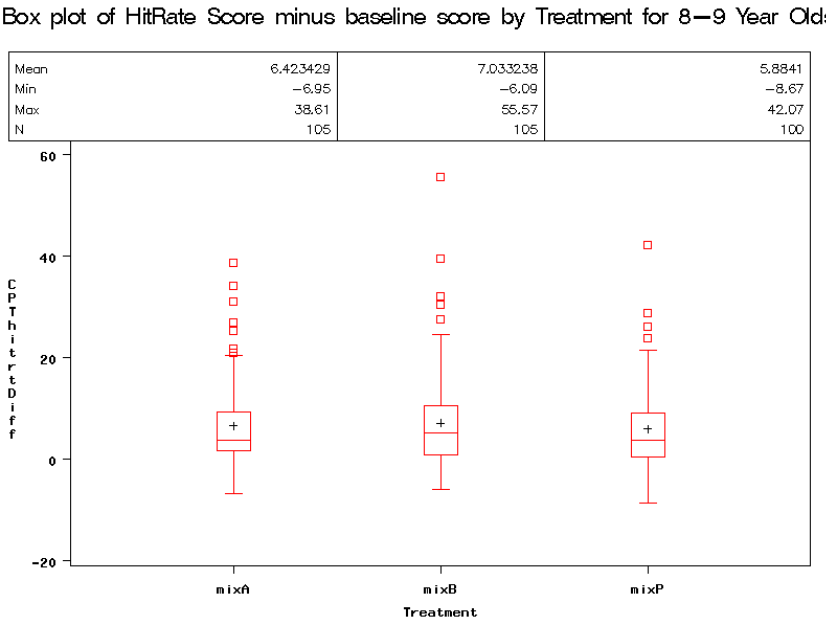


Figure 7-20 Box plot of HitRate Score minus baseline score by Treatment for 8-9 Year Olds

7.2.2.2 High Consumers (Consumption ≥ 85%)

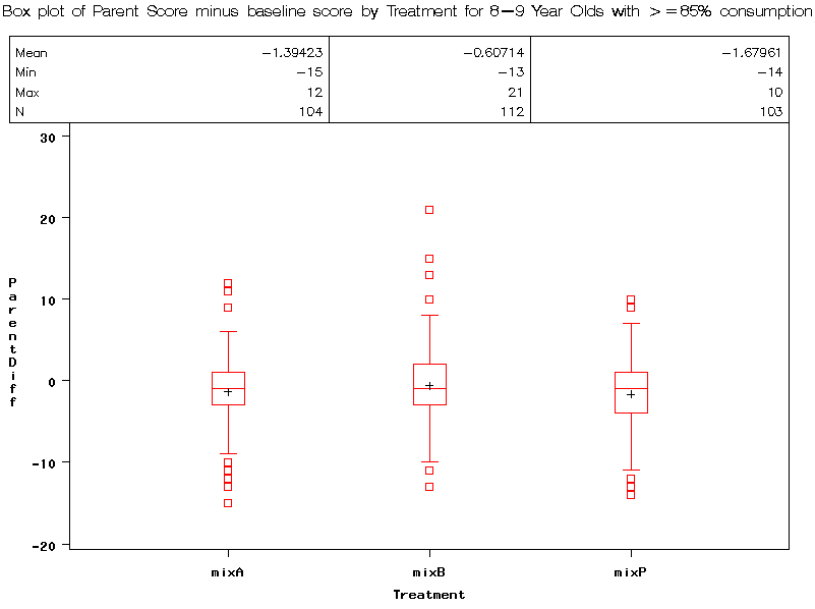


Figure 7-21 Box plot of Parent Score minus baseline score by Treatment for 8-9 Year Olds with > = 85% consumption

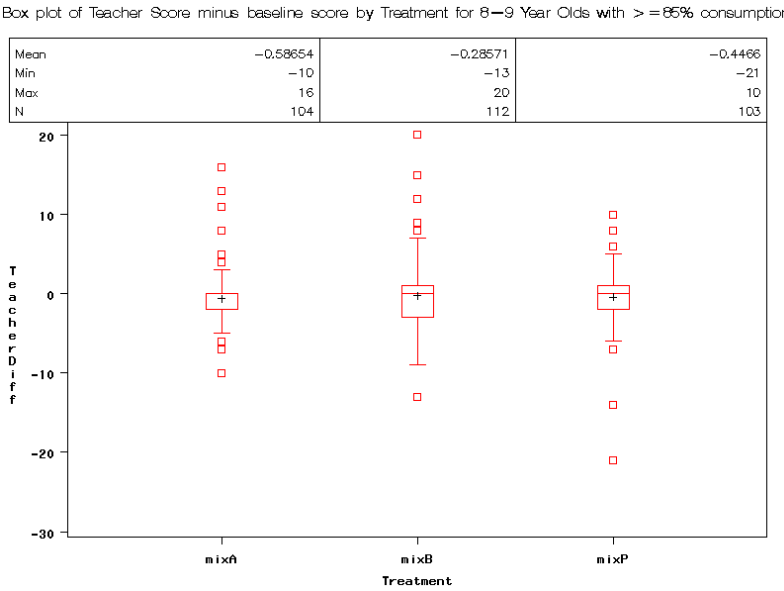


Figure 7-22 Box plot of Teacher Score minus baseline score by Treatment for 8-9 Year Olds with > = 85% consumption

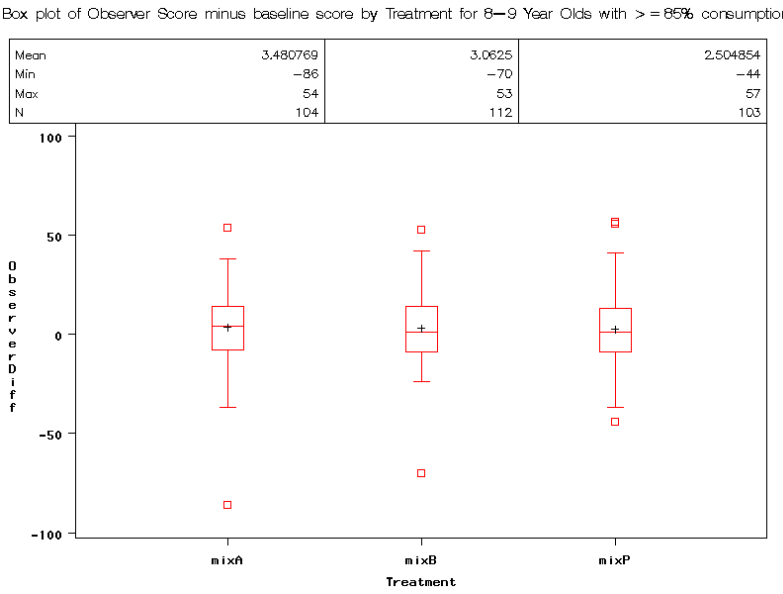


Figure 7-23 Box plot of Observer Score minus baseline score by Treatment for 8-9 Year Olds with > = 85% consumption

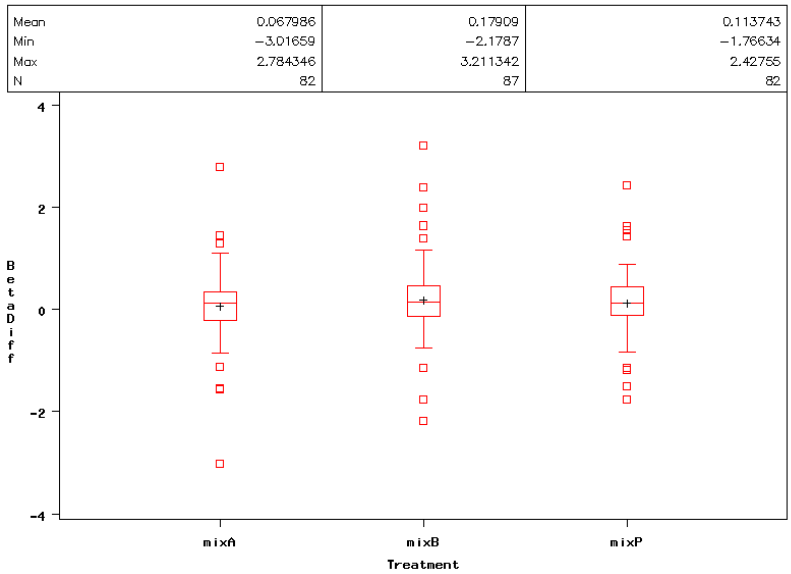


Figure 7-24 Box plot of Beta Score minus baseline score by Treatment for 8-9 Year Olds with > = 85% consumption

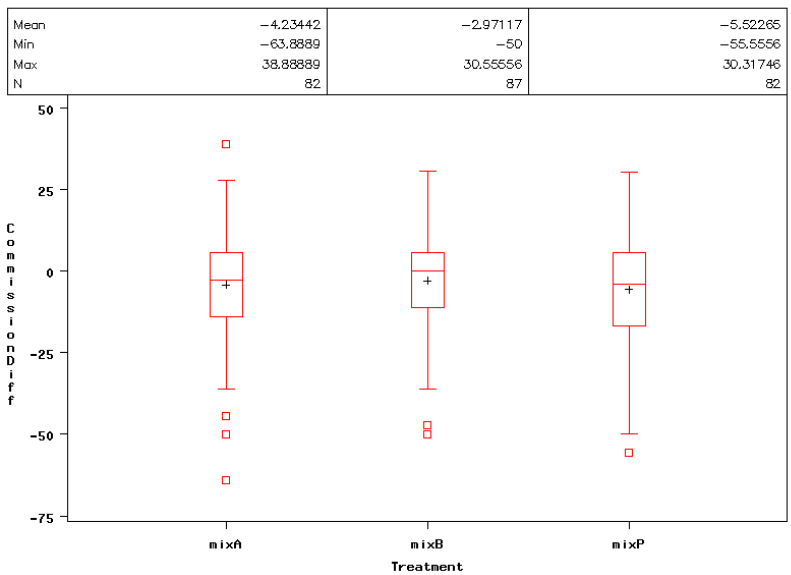


Figure 7-25 Box plot of Commission Score minus baseline score by Treatment for 8-9 Year Olds with > = 85% consumption

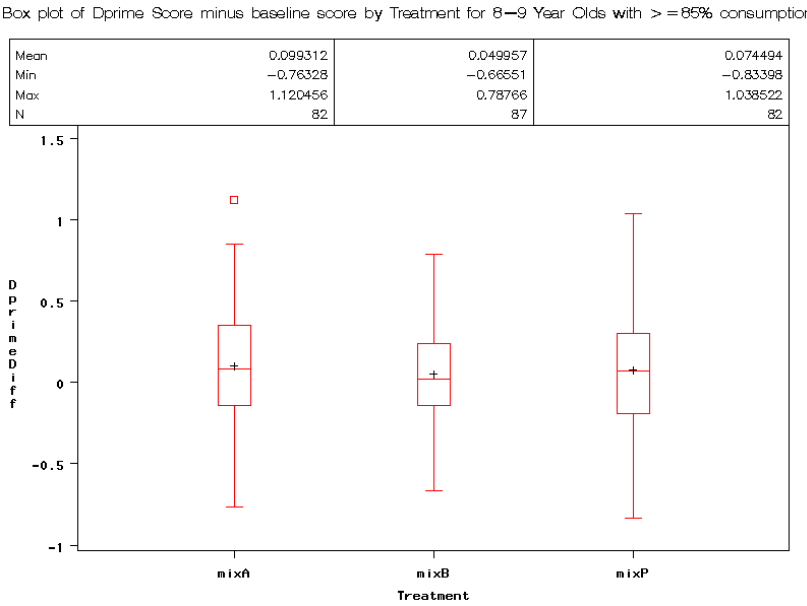


Figure 7-26 Box plot of Dprime Score minus baseline score by Treatment for 8-9 Year Olds with > = 85% consumption

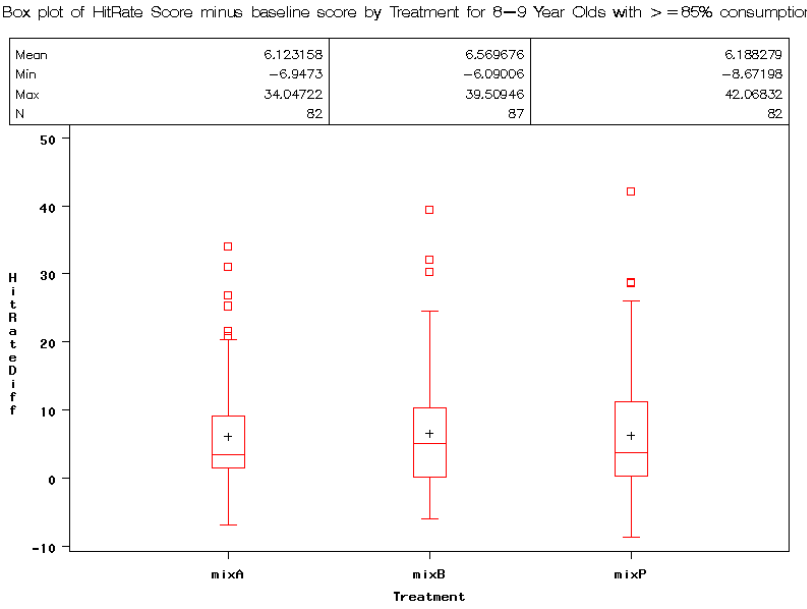


Figure 7-27 Box plot of HitRate Score minus baseline score by Treatment for 8-9 Year Olds with > = 85% consumption

7.2.2.3 Complete cases (case included if >=85% consumption in all challenge weeks and no missing GHA score)

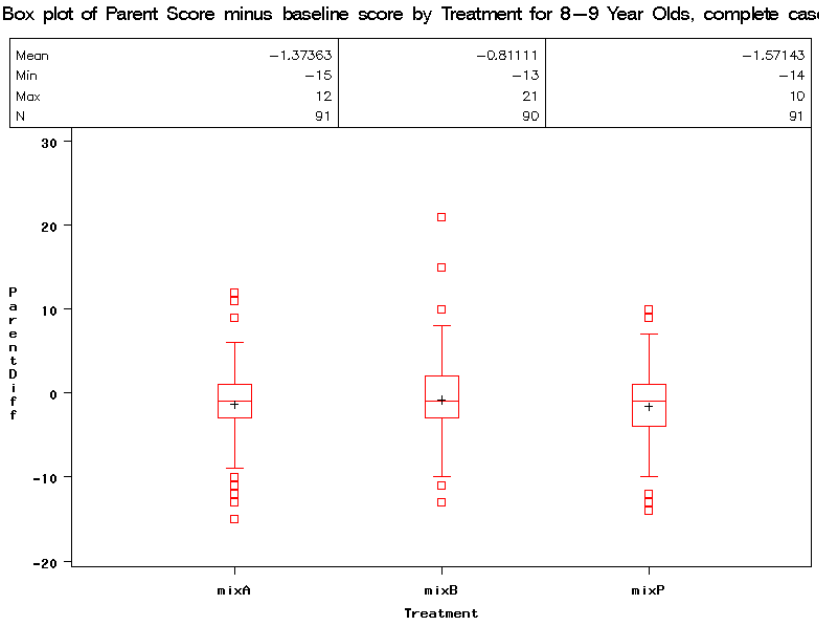


Figure 7-28 Box plot of Parent Score minus baseline score by Treatment for 8-9 Year Olds complete case

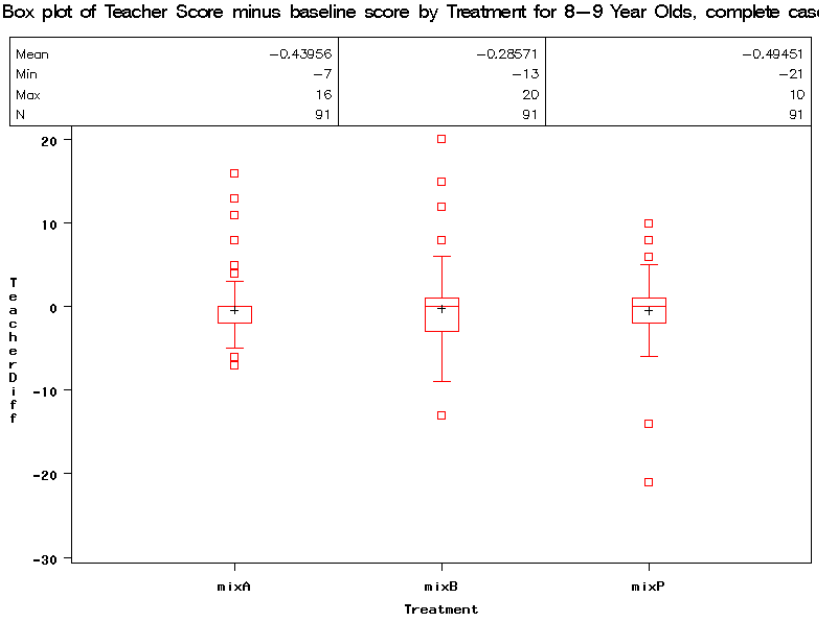


Figure 7-29 Box plot of Teacher Score minus baseline score by Treatment for 8-9 Year Olds complete case

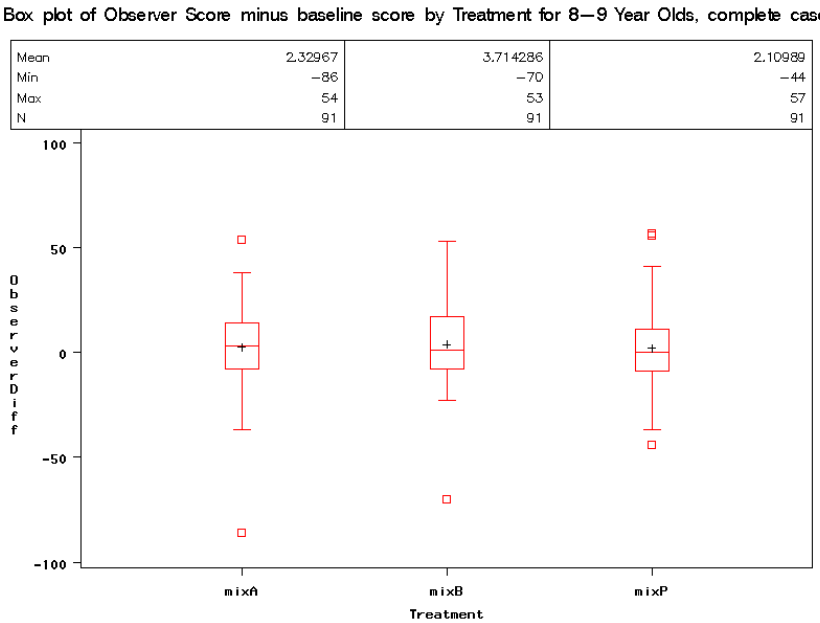


Figure 7-30 Box plot of Observer Score minus baseline score by Treatment for 8-9 Year Olds complete case

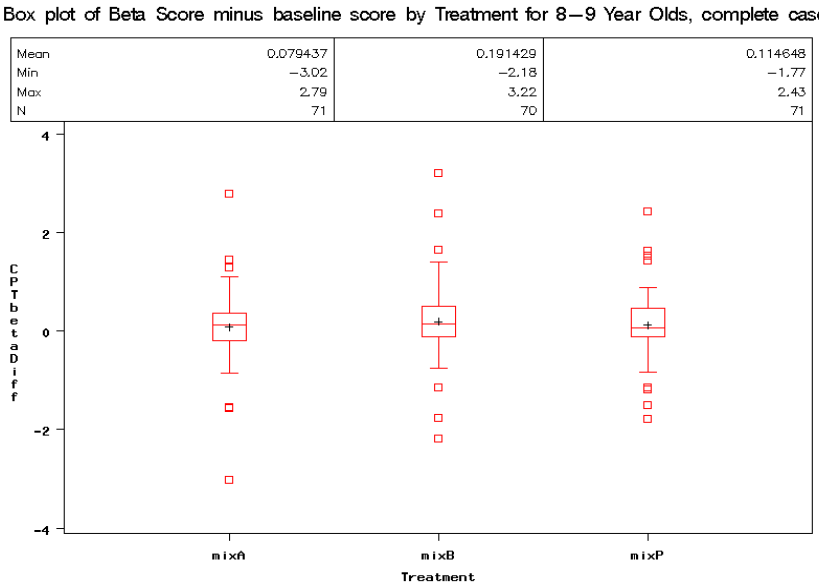


Figure 7-31 Box plot of Beta Score minus baseline score by Treatment for 8-9 Year Olds complete case

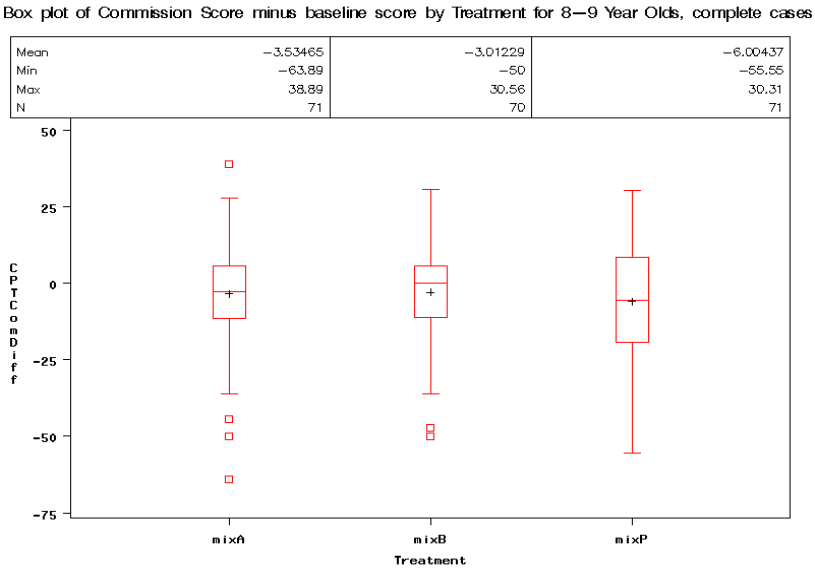


Figure 7-32 Box plot of Commission Score minus baseline score by Treatment for 8-9 Year Olds complete case

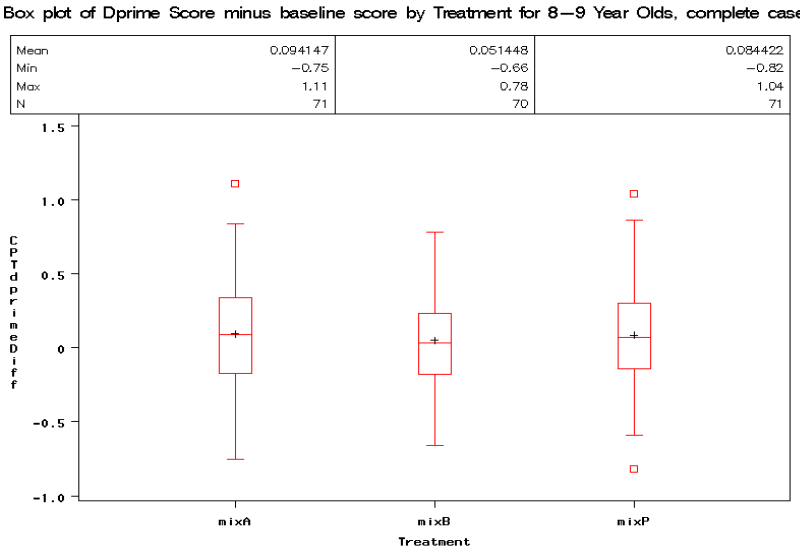


Figure 7-33 Box plot of Dprime Score minus baseline score by Treatment for 8-9 Year Olds complete case

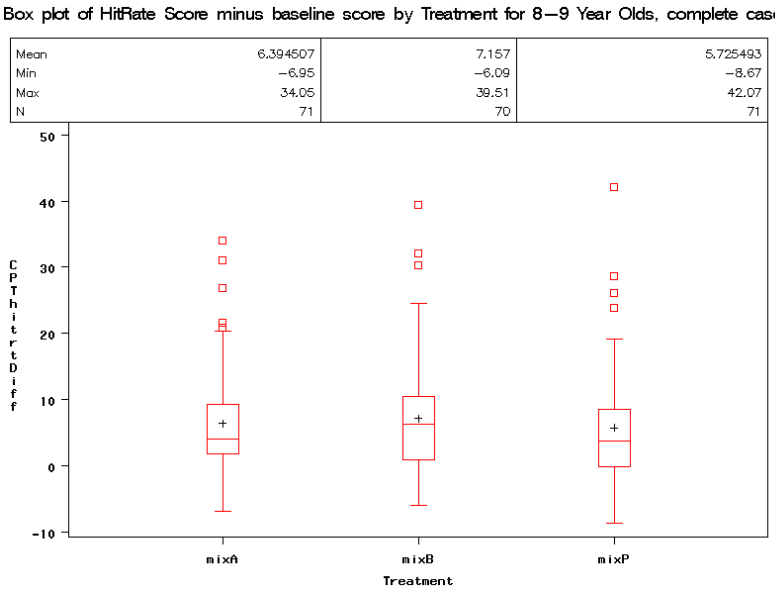


Figure 7-34 Box plot of Computer Hit Rate Score minus baseline by Treatment for 8-9 Year Olds, complete case

7.3 Conclusion and comments

No outstanding treatment effects are apparent from an inspection of the descriptive plots and statistics, although this does not of course rule out effects that can be identified only through formal statistical analysis. This does suggest however that any effect that is identified this way will have a comparatively small absolute size. There is a suggestion however that variability may, in some cases, differ between treatments, for example with the Computer Commission score.

8 Results of the model-based analysis of the single scores

8.1 Tests of the treatment effect

In **Table 8-1** below the P-values are reported of the contrast 'Mix A vs. Placebo' and 'Mix B vs. Placebo' respectively for the entire sample, for the >= 85% -consumers and the completers. Models were run for all combinations of sexes and age groups. Treatment by treatment scatter plots are provided for the significant contrasts (P<=0.05). The significant cases are also highlighted in the tables. The complete list of all P-values is reported in Appendix 2.

Table 8-1 Entire sample, both year groups[1], both sexes:

Test	score	Estimate	StdErr	P-Value
Mix A vs. Placebo	Parent	0.3552	0.2933	0.2265
Mix B vs. Placebo	**Parent**	**0.6385**	**0.2925**	**0.0295**
Mix A vs. Placebo	Teacher	-0.1729	0.2655	0.5152
Mix B vs. Placebo	Teacher	0.06112	0.2643	0.8172
Mix A vs. Placebo	Observer	0.947	0.8911	0.2884
Mix B vs. Placebo	Observer	0.6009	0.8897	0.4997
Mix A vs. Placebo	CPTCom	2.1591	1.6892	0.2027
Mix B vs. Placebo	CPTCom	2.208	1.6884	0.1925
Mix A vs. Placebo	CPThit	0.4611	0.7875	0.5589
Mix B vs. Placebo	CPThit	1.0339	0.7871	0.1905
Mix A vs. Placebo	CPTdpr	-0.01725	0.03306	0.6025
Mix B vs. Placebo	CPTdpr	-0.03695	0.03304	0.2648
Mix A vs. Placebo	CPTbet	0.01165	0.07102	0.8699
Mix B vs. Placebo	CPTbet	0.05985	0.07099	0.4002

[1] In the case of Computer scores, the pool of both age groups is reduced to the 8/9 year-old children, this applies to all the following tables.

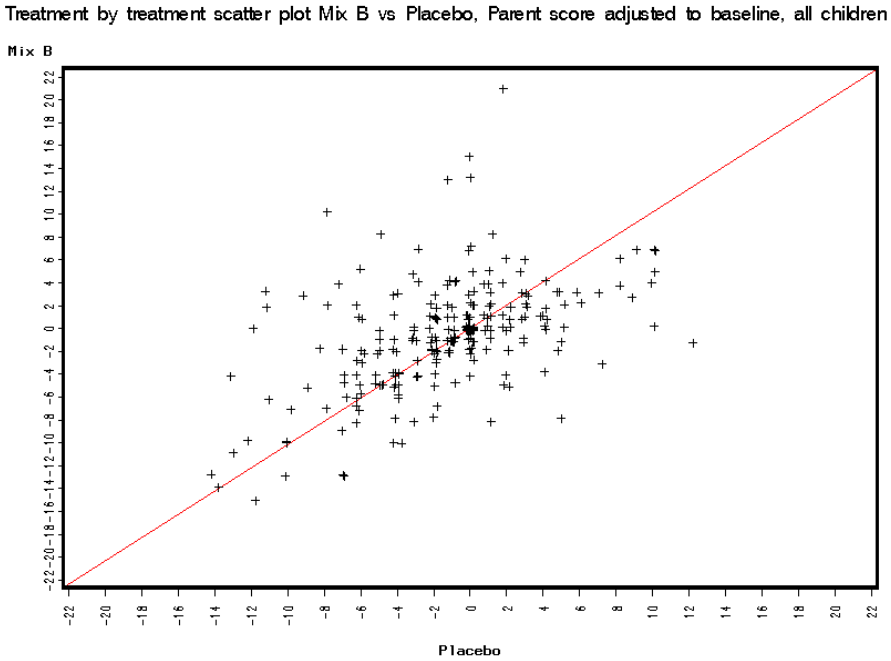


Figure 8-2 Treatment by treatment scatter plot Mix B vs. Placebo, Parent score, both year groups, both sexes

Table 8-3 Entire sample, both year groups, males:

Test	score	Estimate	StdErr	P-Value
Mix A vs. Placebo	Parent	0.452	0.4237	0.287
Mix B vs. Placebo	Parent	0.7304	0.4237	0.0859
Mix A vs. Placebo	Teacher	-0.2474	0.4163	0.5529
Mix B vs. Placebo	Teacher	0.22	0.4105	0.5926
Mix A vs. Placebo	Observer	-0.1714	1.4212	0.9041
Mix B vs. Placebo	Observer	0.5354	1.4107	0.7046
Mix A vs. Placebo	**CPTCom**	**5.0607**	**2.1497**	**0.0206**
Mix B vs. Placebo	CPTCom	4.0152	2.1326	0.0627
Mix A vs. Placebo	CPThit	0.1799	1.2323	0.8842
Mix B vs. Placebo	CPThit	1.5217	1.2228	0.2163
Mix A vs. Placebo	CPTdpr	-0.01613	0.04248	0.7049
Mix B vs. Placebo	CPTdpr	-0.0266	0.04216	0.5296
Mix A vs. Placebo	CPTbet	0.00444	0.1022	0.9654
Mix B vs. Placebo	CPTbet	0.08054	0.1014	0.4291

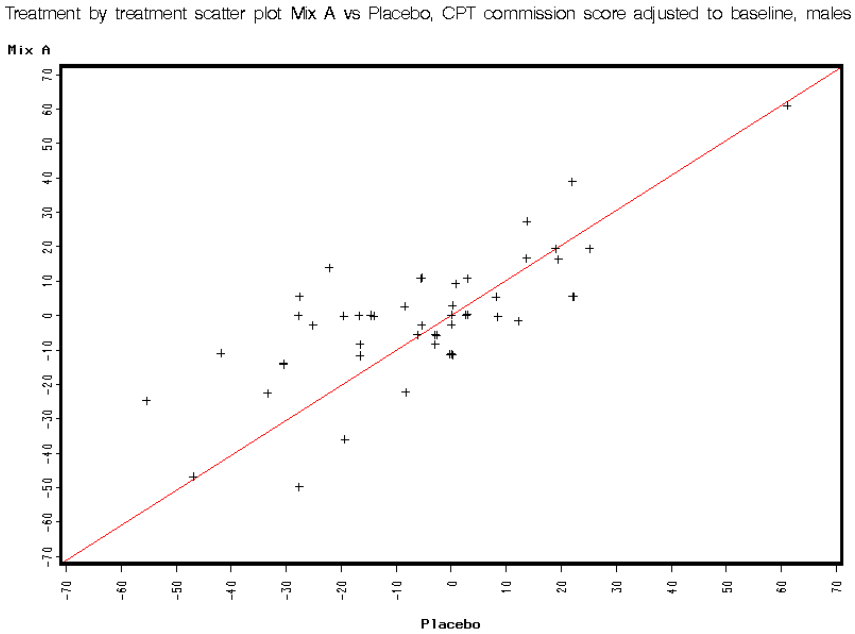


Figure 8-4 Treatment by treatment scatter plot Mix A vs. Placebo CPTCom score, both year groups, males

Table 8-5 Entire sample, both year groups, females:

Test	score	Estimate	StdErr	P-Value
Mix A vs. Placebo	Parent	0.2334	0.401	0.5612
Mix B vs. Placebo	Parent	0.4571	0.4001	0.2543
Mix A vs. Placebo	Teacher	-0.1104	0.3259	0.7352
Mix B vs. Placebo	Teacher	-0.1352	0.3289	0.6814
Mix A vs. Placebo	**Observer**	**2.1135**	**1.0503**	**0.0453**
Mix B vs. Placebo	Observer	0.5275	1.0592	0.619
Mix A vs. Placebo	CPTCom	-0.8642	2.6699	0.7469
Mix B vs. Placebo	CPTCom	-0.2705	2.6898	0.9201
Mix A vs. Placebo	CPThit	0.4012	1.0358	0.6994
Mix B vs. Placebo	CPThit	0.456	1.0431	0.663
Mix A vs. Placebo	CPTdpr	0.00188	0.05236	0.9714
Mix B vs. Placebo	CPTdpr	-0.03838	0.05275	0.4686
Mix A vs. Placebo	CPTbet	0.009068	0.1057	0.9318
Mix B vs. Placebo	CPTbet	0.02831	0.1062	0.7905

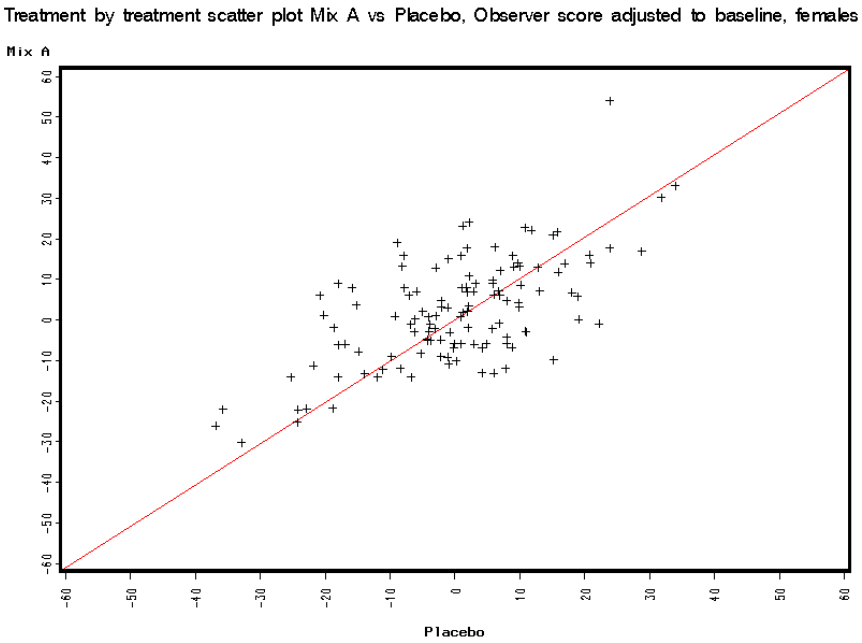


Figure 8-6 Treatment by treatment scatter plot Mix A vs. Placebo, Observer score, both year groups, females

Table 8-7 Entire sample, 3 year old, both sexes:

Test	score	Estimate	StdErr	P-Value
Mix A vs. Placebo	Parent	0.5615	0.4097	0.1717
Mix B vs. Placebo	Parent	0.3825	0.4081	0.3494
Mix A vs. Placebo	Teacher	-0.05585	0.4009	0.8893
Mix B vs. Placebo	Teacher	0.2279	0.399	0.5684
Mix A vs. Placebo	Observer	1.0569	0.9818	0.2827
Mix B vs. Placebo	Observer	-0.2815	0.982	0.7746

Table 8-8 Entire sample, 3 year old males:

Test	score	Estimate	StdErr	P-Value
Mix A vs. Placebo	Parent	0.5895	0.6036	0.3305
Mix B vs. Placebo	Parent	0.366	0.5988	0.5421
Mix A vs. Placebo	Teacher	-0.2682	0.6009	0.6562
Mix B vs. Placebo	Teacher	0.1665	0.5879	0.7775
Mix A vs. Placebo	Observer	0.8724	1.4954	0.5606
Mix B vs. Placebo	Observer	-1.0092	1.4719	0.4942

Table 8-9 Entire sample, 3 year old females

Test	score	Estimate	StdErr	P-Value
Mix A vs. Placebo	Parent	0.2933	0.5538	0.5972
Mix B vs. Placebo	Parent	0.3998	0.5448	0.4644
Mix A vs. Placebo	Teacher	0.2744	0.5406	0.6127
Mix B vs. Placebo	Teacher	0.301	0.5403	0.5785
Mix A vs. Placebo	Observer	1.3641	1.3085	0.2992
Mix B vs. Placebo	Observer	0.7309	1.3107	0.5781

Table 8-10 Entire sample, 8-9 year old, both sexes:

Test	score	Estimate	StdErr	P-Value
Mix A vs. Placebo	Parent	0.1391	0.4193	0.7404
Mix B vs. Placebo	**Parent**	**0.9017**	**0.4186**	**0.0322**
Mix A vs. Placebo	Teacher	-0.2868	0.3503	0.4136
Mix B vs. Placebo	Teacher	-0.1082	0.3489	0.7566
Mix A vs. Placebo	Observer	0.793	1.4861	0.5941
Mix B vs. Placebo	Observer	1.4383	1.4818	0.3327
Mix A vs. Placebo	CPTCom	2.1591	1.6892	0.2027
Mix B vs. Placebo	CPTCom	2.208	1.6884	0.1925
Mix A vs. Placebo	CPThit	0.4611	0.7875	0.5589
Mix B vs. Placebo	CPThit	1.0339	0.7871	0.1905
Mix A vs. Placebo	CPTdpr	-0.01725	0.03306	0.6025
Mix B vs. Placebo	CPTdpr	-0.03695	0.03304	0.2648
Mix A vs. Placebo	CPTbet	0.01165	0.07102	0.8699
Mix B vs. Placebo	CPTbet	0.05985	0.07099	0.4002

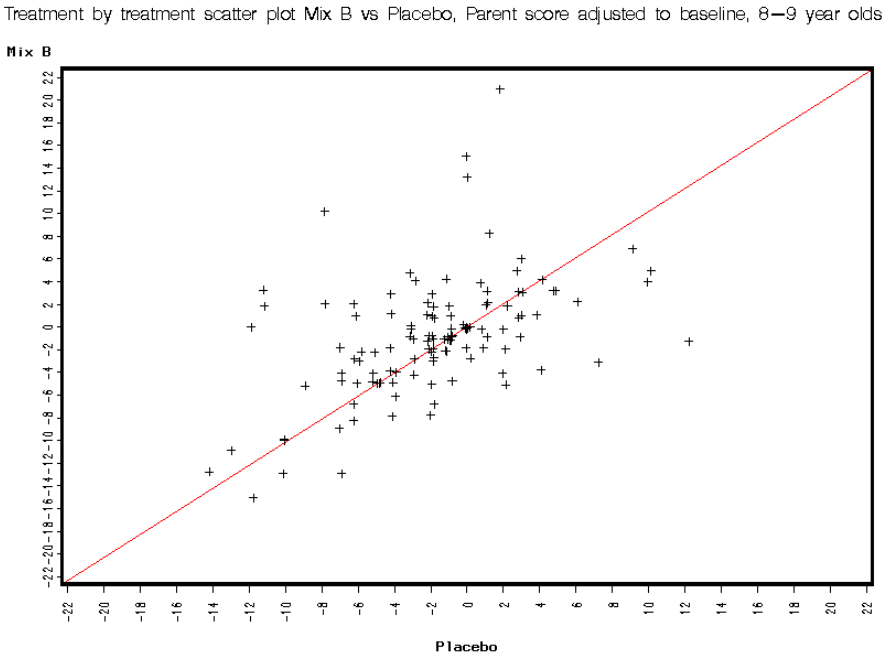


Figure 8-11 Treatment by treatment scatter plot Mix B vs. Placebo Parent score, 8/9Y year group, both sexes

Table 8-12 Entire sample, 8-9 year old, males:

Test	score	Estimate	StdErr	P-Value
Mix A vs. Placebo	Parent	0.1704	0.6079	0.7797
Mix B vs. Placebo	Parent	1.0299	0.6069	0.0921
Mix A vs. Placebo	Teacher	-0.06219	0.5951	0.9169
Mix B vs. Placebo	Teacher	0.2913	0.5863	0.6202
Mix A vs. Placebo	Observer	-0.3251	2.4365	0.8941
Mix B vs. Placebo	Observer	2.3671	2.4165	0.3291
Mix A vs. Placebo	**CPTCom**	**5.0607**	**2.1497**	**0.0206**
Mix B vs. Placebo	CPTCom	4.0152	2.1326	0.0627
Mix A vs. Placebo	CPThit	0.1799	1.2323	0.8842
Mix B vs. Placebo	CPThit	1.5217	1.2228	0.2163
Mix A vs. Placebo	CPTdpr	-0.01613	0.04248	0.7049
Mix B vs. Placebo	CPTdpr	-0.0266	0.04216	0.5296
Mix A vs. Placebo	CPTbet	0.00444	0.1022	0.9654
Mix B vs. Placebo	CPTbet	0.08054	0.1014	0.4291

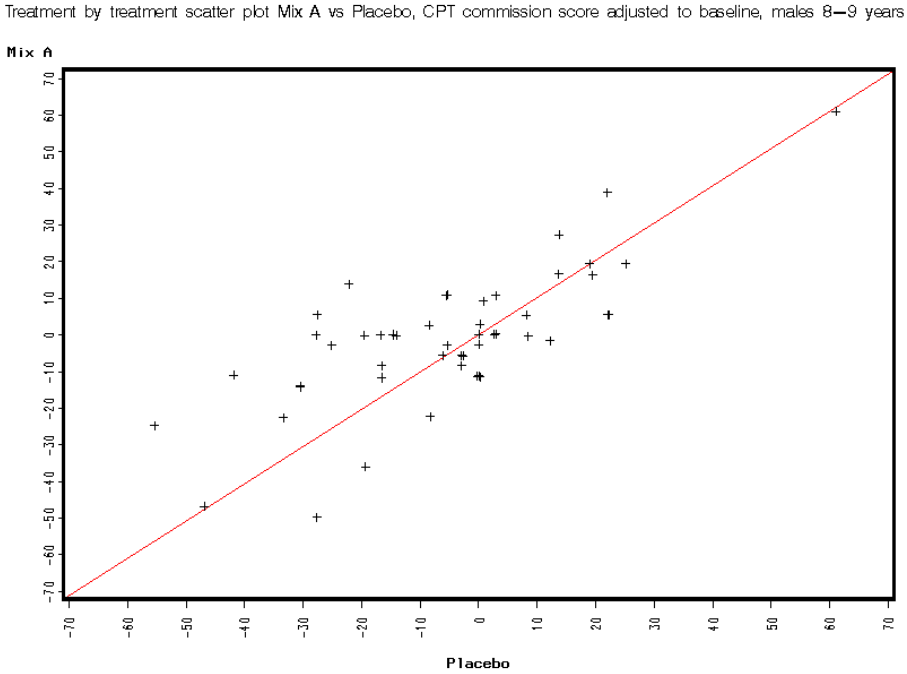


Figure 8-13 Treatment by treatment scatter plot Mix A vs. Placebo, CPTCom score, 8/9Y year group, males

Table 8-14 Entire sample, 8-9 year old, females:

Test	score	Estimate	StdErr	P-Value
Mix A vs. Placebo	Parent	0.1707	0.5993	0.7763
Mix B vs. Placebo	Parent	0.6376	0.5959	0.2869
Mix A vs. Placebo	Teacher	-0.4375	0.3766	0.2478
Mix B vs. Placebo	Teacher	-0.5627	0.3762	0.1374
Mix A vs. Placebo	Observer	2.6862	1.7116	0.1192
Mix B vs. Placebo	Observer	0.4576	1.7029	0.7886
Mix A vs. Placebo	CPTCom	-0.8642	2.6699	0.7469
Mix B vs. Placebo	CPTCom	-0.2705	2.6898	0.9201
Mix A vs. Placebo	CPThit	0.4012	1.0358	0.6994
Mix B vs. Placebo	CPThit	0.456	1.0431	0.663
Mix A vs. Placebo	CPTdpr	0.00188	0.05236	0.9714
Mix B vs. Placebo	CPTdpr	-0.03838	0.05275	0.4686
Mix A vs. Placebo	CPTbet	0.009068	0.1057	0.9318
Mix B vs. Placebo	CPTbet	0.02831	0.1062	0.7905

Table 8-15 ≥85% consumption, both year groups, both sexes:

Test	score	Estimate	StdErr	P-Value
Mix A vs. Placebo	Parent	0.4784	0.3303	0.1483
Mix B vs. Placebo	**Parent**	**0.788**	**0.3265**	**0.0163**
Mix A vs. Placebo	Teacher	0.04531	0.3008	0.8804
Mix B vs. Placebo	Teacher	0.3655	0.2968	0.219
Mix A vs. Placebo	Observer	1.2068	1.0666	0.2586
Mix B vs. Placebo	Observer	0.4436	1.0511	0.6732
Mix A vs. Placebo	CPTCom	2.0004	1.8664	0.2855
Mix B vs. Placebo	CPTCom	2.968	1.8371	0.1082
Mix A vs. Placebo	CPThit	0.2699	0.8749	0.7581
Mix B vs. Placebo	CPThit	0.8544	0.8612	0.3227
Mix A vs. Placebo	CPTdpr	0.01144	0.03635	0.7533
Mix B vs. Placebo	CPTdpr	-0.03139	0.03579	0.3818
Mix A vs. Placebo	CPTbet	-0.03885	0.07316	0.5962
Mix B vs. Placebo	CPTbet	0.08324	0.07202	0.2496

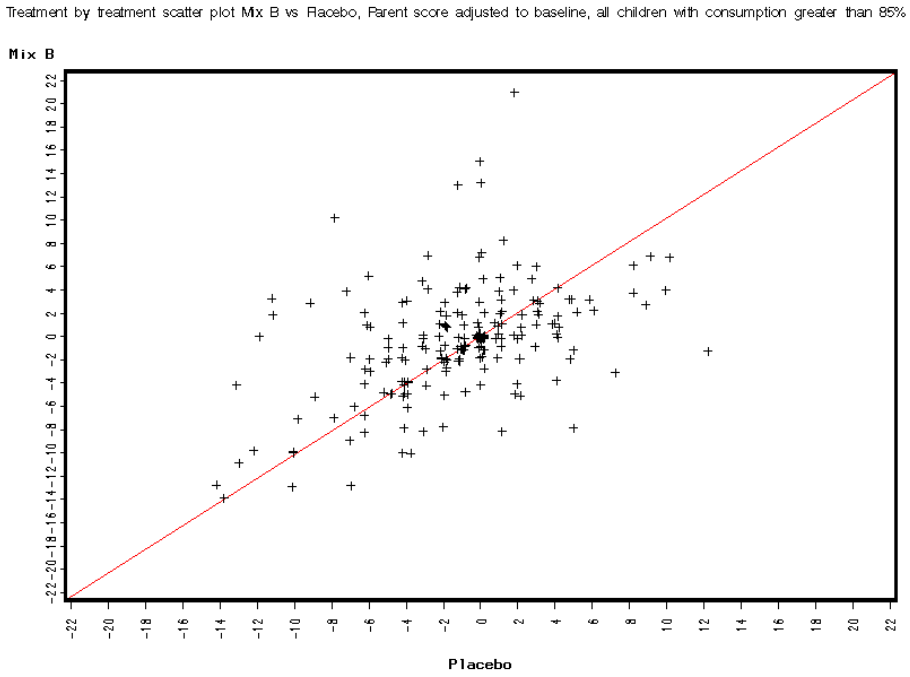


Figure 8-16 Treatment by treatment scatter plot Mix B vs. Placebo, Parent score, both year groups, both sexes

Table 8-17 ≥85% consumption, both year groups, males:

Test	score	Estimate	StdErr	P-Value
Mix A vs. Placebo	Parent	0.632	0.4527	0.1642
Mix B vs. Placebo	Parent	0.8443	0.4504	0.0622
Mix A vs. Placebo	Teacher	-0.1771	0.4612	0.7014
Mix B vs. Placebo	Teacher	0.4176	0.4535	0.3582
Mix A vs. Placebo	Observer	0.3901	1.6478	0.8131
Mix B vs. Placebo	Observer	0.2666	1.6205	0.8695
Mix A vs. Placebo	**CPTCom**	**5.0867**	**2.0922**	**0.0172**
Mix B vs. Placebo	**CPTCom**	**6.0375**	**2.0077**	**0.0035**
Mix A vs. Placebo	CPThit	-0.02723	1.2124	0.9821
Mix B vs. Placebo	CPThit	0.7441	1.1637	0.5243
Mix A vs. Placebo	CPTdpr	-0.02059	0.0458	0.6543
Mix B vs. Placebo	CPTdpr	-0.05848	0.04399	0.1874
Mix A vs. Placebo	CPTbet	-0.05105	0.1111	0.647
Mix B vs. Placebo	CPTbet	0.07259	0.1066	0.498

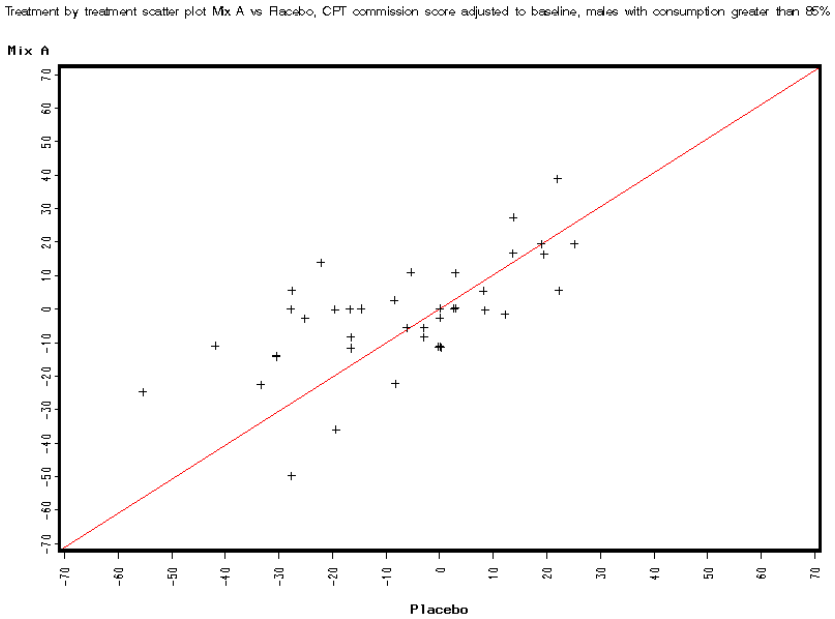


Figure 8-18 Treatment by treatment scatter plot Mix A vs. Placebo, CPTCom score, both year groups, male

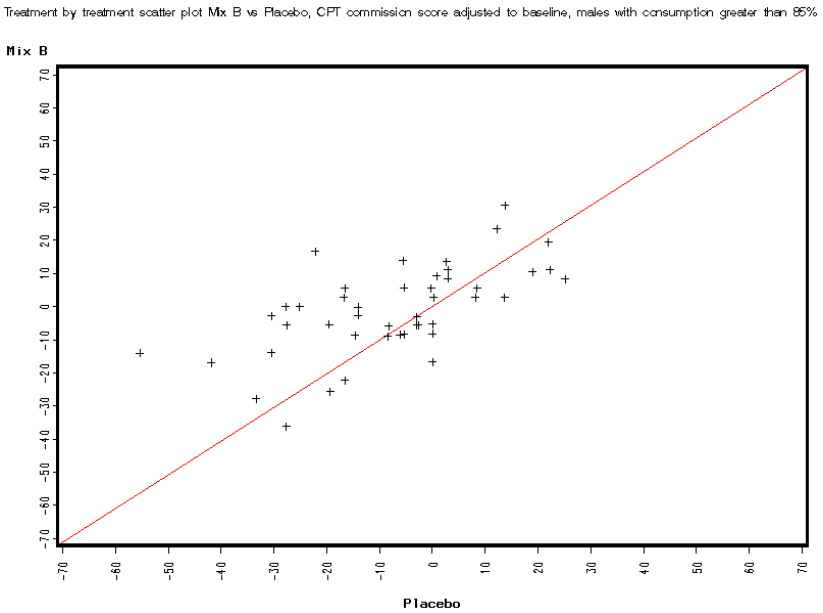


Figure 8-19 Treatment by treatment scatter plot Mix B vs. Placebo, CPTCom score, both year groups, male

Table 8-20 ≥85% consumption, both year groups, females:

Test	score	Estimate	StdErr	P-Value
Mix A vs. Placebo	Parent	0.2703	0.4772	0.5718
Mix B vs. Placebo	Parent	0.5716	0.4702	0.2257
Mix A vs. Placebo	Teacher	0.3327	0.3694	0.3689
Mix B vs. Placebo	Teacher	0.3338	0.3672	0.3645
Mix A vs. Placebo	Observer	2.135	1.293	0.1005
Mix B vs. Placebo	Observer	0.5166	1.2812	0.6873
Mix A vs. Placebo	CPTCom	-2.1759	3.2193	0.5014
Mix B vs. Placebo	CPTCom	-1.9687	3.277	0.55
Mix A vs. Placebo	CPThit	0.02704	1.3084	0.9836
Mix B vs. Placebo	CPThit	1.1053	1.3317	0.4095
Mix A vs. Placebo	CPTdpr	0.07202	0.05923	0.2283
Mix B vs. Placebo	CPTdpr	0.01188	0.06032	0.8444
Mix A vs. Placebo	CPTbet	-0.02995	0.09866	0.7624
Mix B vs. Placebo	CPTbet	0.08953	0.1002	0.375

Table 8-21 ≥85% consumption, 3 year old, both sexes:

Test	score	Estimate	StdErr	P-Value
Mix A vs. Placebo	Parent	0.7301	0.4663	0.1191
Mix B vs. Placebo	Parent	0.5582	0.464	0.2305
Mix A vs. Placebo	Teacher	0.1096	0.4523	0.8088
Mix B vs. Placebo	Teacher	0.4874	0.4489	0.2791
Mix A vs. Placebo	Observer	1.2026	1.1839	0.3111
Mix B vs. Placebo	Observer	-0.399	1.1745	0.7345

Table 8-22 ≥85% consumption, 3 year old, males

Test	score	Estimate	StdErr	P-Value
Mix A vs. Placebo	Parent	0.9864	0.6581	0.1373
Mix B vs. Placebo	Parent	0.5061	0.6656	0.449
Mix A vs. Placebo	Teacher	-0.3435	0.6635	0.606
Mix B vs. Placebo	Teacher	0.237	0.6675	0.7234
Mix A vs. Placebo	Observer	1.2439	1.7786	0.4861
Mix B vs. Placebo	Observer	-1.3797	1.7896	0.4428

Table 8-23 ≥85% consumption, 3 year old, females:

Test	score	Estimate	StdErr	P-Value
Mix A vs. Placebo	Parent	0.1628	0.6637	0.8068
Mix B vs. Placebo	Parent	0.5362	0.6421	0.4058
Mix A vs. Placebo	Teacher	0.6647	0.6218	0.2881
Mix B vs. Placebo	Teacher	0.6884	0.6003	0.2547
Mix A vs. Placebo	Observer	1.0705	1.5947	0.5038
Mix B vs. Placebo	Observer	0.7843	1.5396	0.6117

Table 8-24 ≥85% consumption, 8-9 year old, both sexes:

Test	score	Estimate	StdErr	P-Value
Mix A vs. Placebo	Parent	0.2044	0.4665	0.6618
Mix B vs. Placebo	**Parent**	**0.9687**	**0.4584**	**0.0359**
Mix A vs. Placebo	Teacher	-0.0471	0.4008	0.9066
Mix B vs. Placebo	Teacher	0.234	0.3938	0.5531
Mix A vs. Placebo	Observer	1.0664	1.7279	0.5378
Mix B vs. Placebo	Observer	1.3445	1.6937	0.4283
Mix A vs. Placebo	CPTCom	2.0004	1.8664	0.2855
Mix B vs. Placebo	CPTCom	2.968	1.8371	0.1082
Mix A vs. Placebo	CPThit	0.2699	0.8749	0.7581
Mix B vs. Placebo	CPThit	0.8544	0.8612	0.3227
Mix A vs. Placebo	CPTdpr	0.01144	0.03635	0.7533
Mix B vs. Placebo	CPTdpr	-0.03139	0.03579	0.3818
Mix A vs. Placebo	CPTbet	-0.03885	0.07316	0.5962
Mix B vs. Placebo	CPTbet	0.08324	0.07202	0.2496

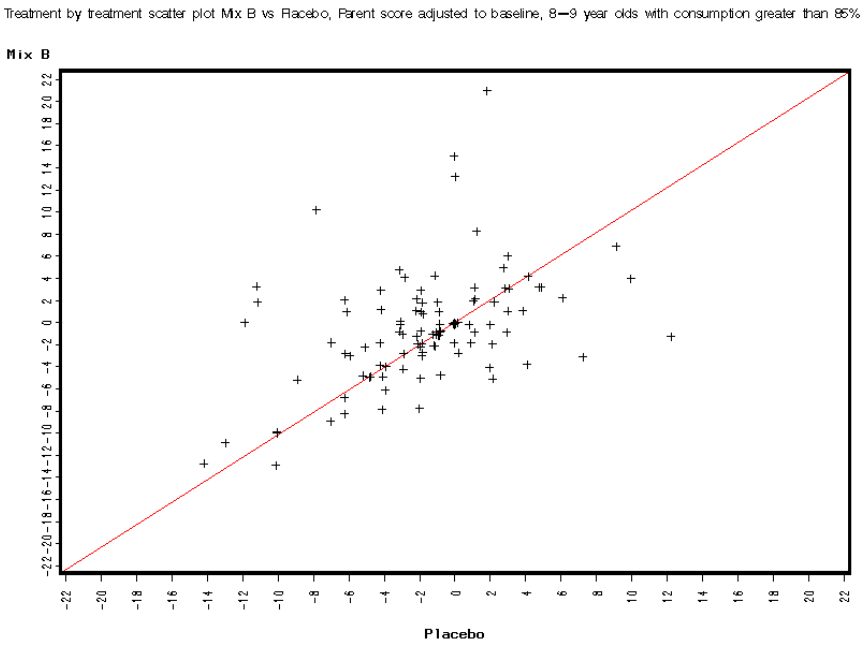


Figure 8-25 Treatment by treatment scatter plot Mix B vs. Placebo, Parent score, 8/9Y year group, both sexes

Table 8-26 ≥85% consumption, 8-9 year old, males:

Test	score	Estimate	StdErr	P-Value
Mix A vs. Placebo	Parent	0.1331	0.6362	0.8346
Mix B vs. Placebo	Parent	1.0254	0.6214	0.1018
Mix A vs. Placebo	Teacher	0.04486	0.6454	0.9447
Mix B vs. Placebo	Teacher	0.4307	0.6256	0.4926
Mix A vs. Placebo	Observer	0.531	2.6376	0.8408
Mix B vs. Placebo	Observer	2.4231	2.5576	0.3455
Mix A vs. Placebo	**CPTCom**	**5.0867**	**2.0922**	**0.0172**
Mix B vs. Placebo	**CPTCom**	**6.0375**	**2.0077**	**0.0035**
Mix A vs. Placebo	CPThit	-0.02723	1.2124	0.9821
Mix B vs. Placebo	CPThit	0.7441	1.1637	0.5243
Mix A vs. Placebo	CPTdpr	-0.02059	0.0458	0.6543
Mix B vs. Placebo	CPTdpr	-0.05848	0.04399	0.1874
Mix A vs. Placebo	CPTbet	-0.05105	0.1111	0.647
Mix B vs. Placebo	CPTbet	0.07259	0.1066	0.498

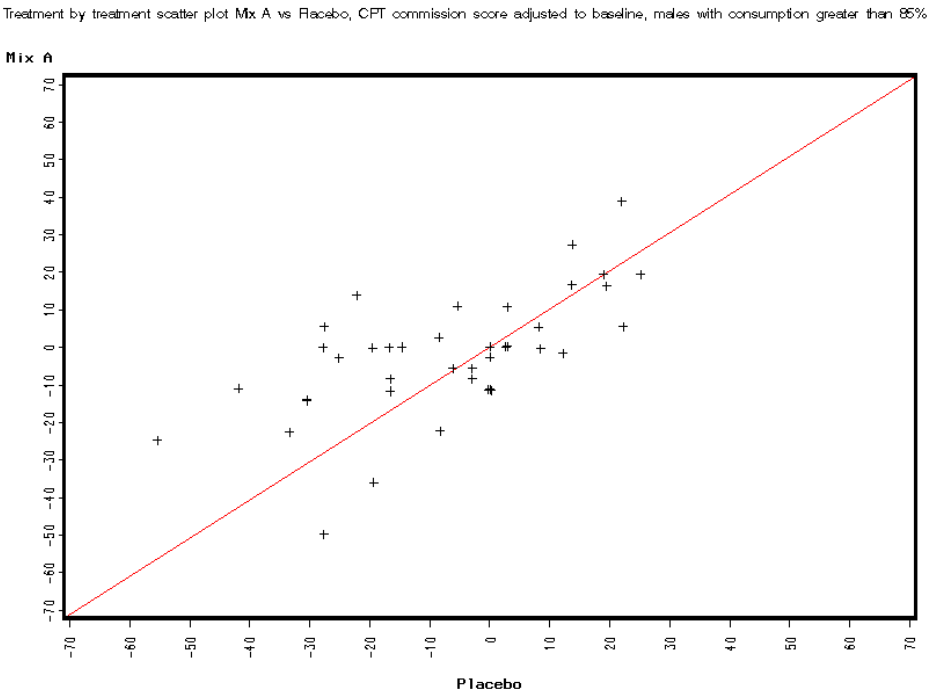


Figure 8-27 Treatment by treatment scatter plot Mix A vs. Placebo, CPTCom score, 8/9Y year group, males

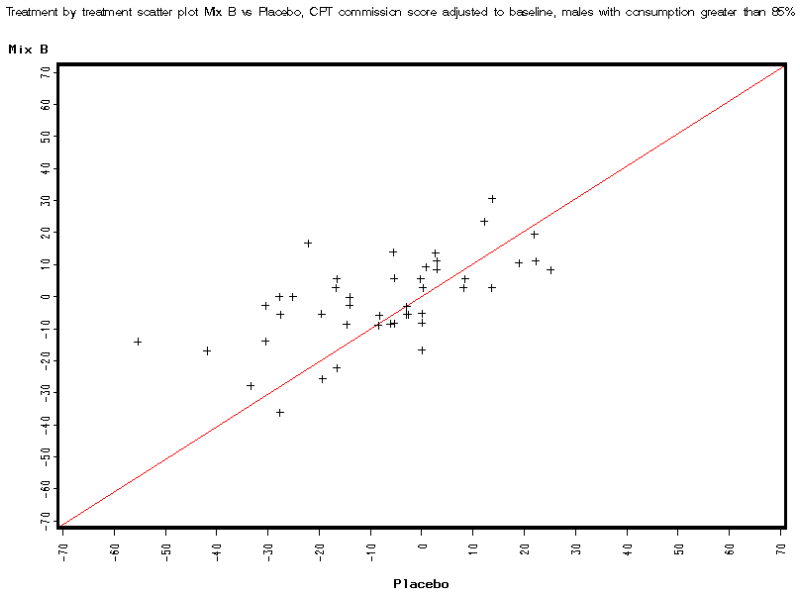


Figure 8-28 Treatment by treatment scatter plot Mix B vs. Placebo, CPTCom score, 8/9Y year group, males

Table 8-29 ≥85% consumption, 8-9 year old, females:

Test	score	Estimate	StdErr	P-Value
Mix A vs. Placebo	Parent	0.3337	0.7187	0.6436
Mix B vs. Placebo	Parent	0.7256	0.703	0.305
Mix A vs. Placebo	Teacher	-0.03036	0.4141	0.9417
Mix B vs. Placebo	Teacher	-0.1026	0.4108	0.8035
Mix A vs. Placebo	Observer	2.4603	2.1333	0.2522
Mix B vs. Placebo	Observer	0.2494	2.1019	0.9059
Mix A vs. Placebo	CPTCom	-2.1759	3.2193	0.5014
Mix B vs. Placebo	CPTCom	-1.9687	3.277	0.55
Mix A vs. Placebo	CPThit	0.02704	1.3084	0.9836
Mix B vs. Placebo	CPThit	1.1053	1.3317	0.4095
Mix A vs. Placebo	CPTdpr	0.07202	0.05923	0.2283
Mix B vs. Placebo	CPTdpr	0.01188	0.06032	0.8444
Mix A vs. Placebo	CPTbet	-0.02995	0.09866	0.7624
Mix B vs. Placebo	CPTbet	0.08953	0.1002	0.375

Table 8-30 Complete case, both year groups, both sexes:

Test	score	Estimate	StdErr	P-Value
Mix A vs. Placebo	Parent	0.6646	0.3603	0.066
Mix B vs. Placebo	**Parent**	**0.7726**	**0.3611**	**0.0332**
Mix A vs. Placebo	Teacher	0.3201	0.3173	0.3138
Mix B vs. Placebo	Teacher	0.455	0.3175	0.1528
Mix A vs. Placebo	Observer	0.7676	1.1998	0.5228
Mix B vs. Placebo	Observer	0.9208	1.2004	0.4436
Mix A vs. Placebo	CPTCom	2.5671	2.0325	0.2087
Mix B vs. Placebo	CPTCom	3.1417	2.0452	0.1268
Mix A vs. Placebo	CPThit	0.8879	0.9376	0.3453
Mix B vs. Placebo	CPThit	1.6799	0.9434	0.0772
Mix A vs. Placebo	CPTdpr	0.006658	0.03938	0.866
Mix B vs. Placebo	CPTdpr	-0.03562	0.03963	0.3703
Mix A vs. Placebo	CPTbet	-0.02996	0.07937	0.7064
Mix B vs. Placebo	CPTbet	0.08221	0.07986	0.3051

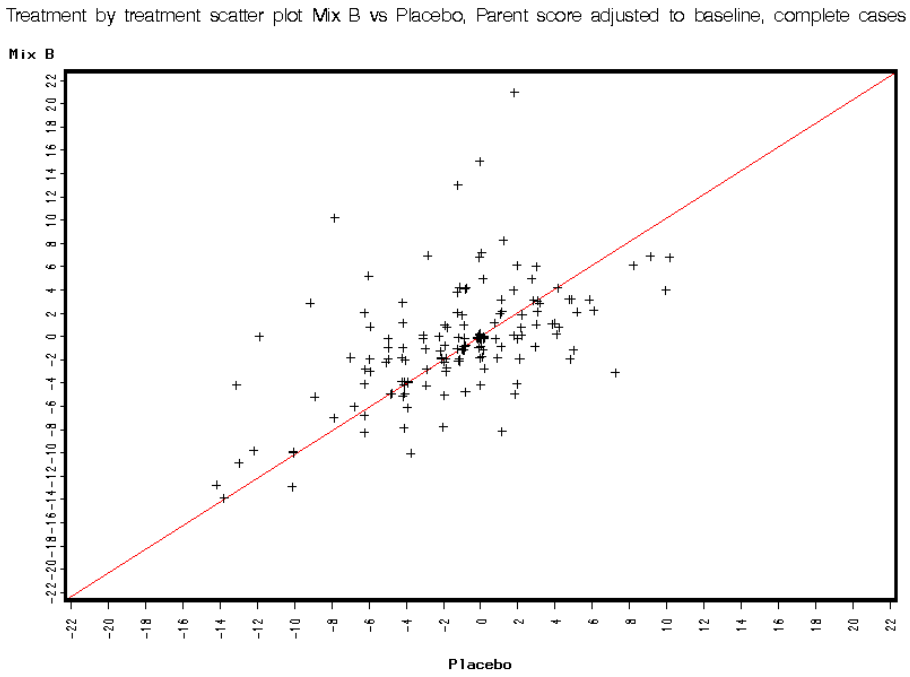


Figure 8-31 Treatment by treatment scatter plot Mix B vs. Placebo, Parent score, both year groups, both sexes

Table 8-32 Complete case, both year groups, males:

Test	score	Estimate	StdErr	P-Value
Mix A vs. Placebo	Parent	0.5854	0.4907	0.2345
Mix B vs. Placebo	Parent	0.6603	0.4911	0.1805
Mix A vs. Placebo	Teacher	0.1481	0.4706	0.7533
Mix B vs. Placebo	Teacher	0.4749	0.4692	0.3128
Mix A vs. Placebo	Observer	0.08688	1.8329	0.9622
Mix B vs. Placebo	Observer	1.3772	1.8274	0.4521
Mix A vs. Placebo	**CPTCom**	**5.5423**	**2.2612**	**0.0166**
Mix B vs. Placebo	**CPTCom**	**6.3442**	**2.2428**	**0.006**
Mix A vs. Placebo	CPThit	0.4741	1.2699	0.71
Mix B vs. Placebo	CPThit	1.5185	1.2596	0.2318
Mix A vs. Placebo	CPTdpr	-0.01048	0.04925	0.832
Mix B vs. Placebo	CPTdpr	-0.05424	0.04885	0.2705
Mix A vs. Placebo	CPTbet	-0.07036	0.1213	0.5636
Mix B vs. Placebo	CPTbet	0.07258	0.1203	0.5482

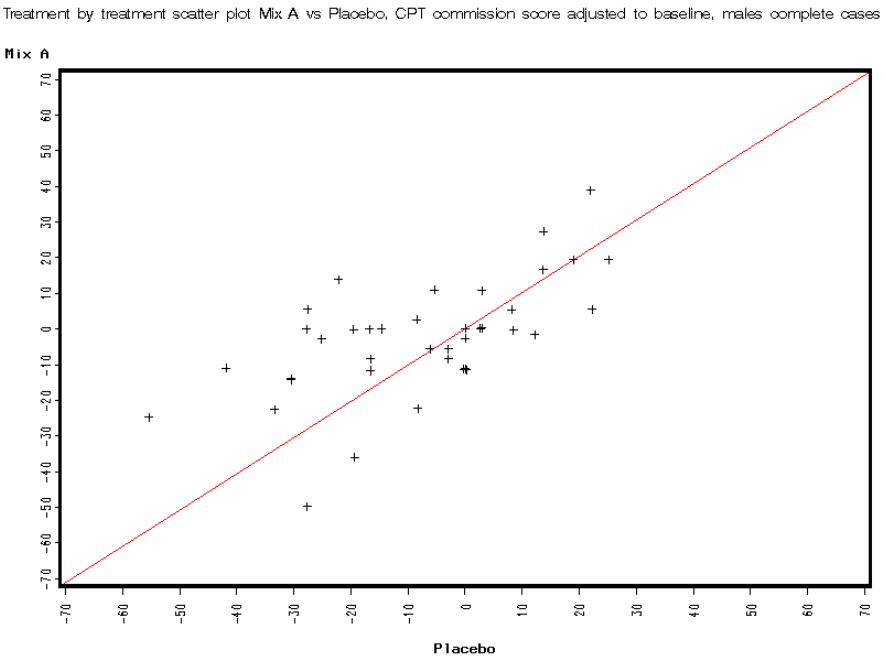


Figure 8-33 Treatment by treatment scatter plot Mix A vs. Placebo, CPTCom score, both year groups, males

Table 8-34 Complete case, both year groups, females:

Test	score	Estimate	StdErr	P-Value
Mix A vs. Placebo	Parent	0.6451	0.5319	0.2272
Mix B vs. Placebo	Parent	0.764	0.5356	0.1559
Mix A vs. Placebo	Teacher	0.5735	0.4141	0.1682
Mix B vs. Placebo	Teacher	0.4363	0.4168	0.297
Mix A vs. Placebo	Observer	1.5459	1.4716	0.2952
Mix B vs. Placebo	Observer	0.1812	1.4818	0.9028
Mix A vs. Placebo	CPTCom	-1.7885	3.5952	0.6207
Mix B vs. Placebo	CPTCom	-2.0092	3.7482	0.5939
Mix A vs. Placebo	CPThit	0.7744	1.4192	0.5874
Mix B vs. Placebo	CPThit	2.2789	1.4791	0.1287
Mix A vs. Placebo	CPTdpr	0.05453	0.06497	0.4046
Mix B vs. Placebo	CPTdpr	-0.00642	0.06774	0.9249
Mix A vs. Placebo	CPTbet	0.02522	0.104	0.8093
Mix B vs. Placebo	CPTbet	0.1013	0.1084	0.3538

Table 8-35 Complete case, 3 year, both sexes:

Test	score	Estimate	StdErr	P-Value
Mix A vs. Placebo	**Parent**	**1.2448**	**0.5585**	**0.0274**
Mix A vs. Placebo	Teacher	0.594	0.491	0.2284
Mix A vs. Placebo	Observer	1.1165	1.3502	0.4096
Mix B vs. Placebo	Parent	0.8326	0.5578	0.1377
Mix B vs. Placebo	Teacher	0.6967	0.4907	0.1579
Mix B vs. Placebo	Observer	-0.2299	1.3486	0.8649

Table 8-36 Complete case, 3 year, males:

Test	score	Estimate	StdErr	P-Value
Mix A vs. Placebo	Parent	1.1678	0.8319	0.1648
Mix A vs. Placebo	Teacher	0.2437	0.6394	0.7043
Mix A vs. Placebo	Observer	1.5451	1.9817	0.4382
Mix B vs. Placebo	Parent	0.7659	0.8451	0.3679
Mix B vs. Placebo	Teacher	0.542	0.6495	0.4069
Mix B vs. Placebo	Observer	-0.4845	2.0129	0.8105

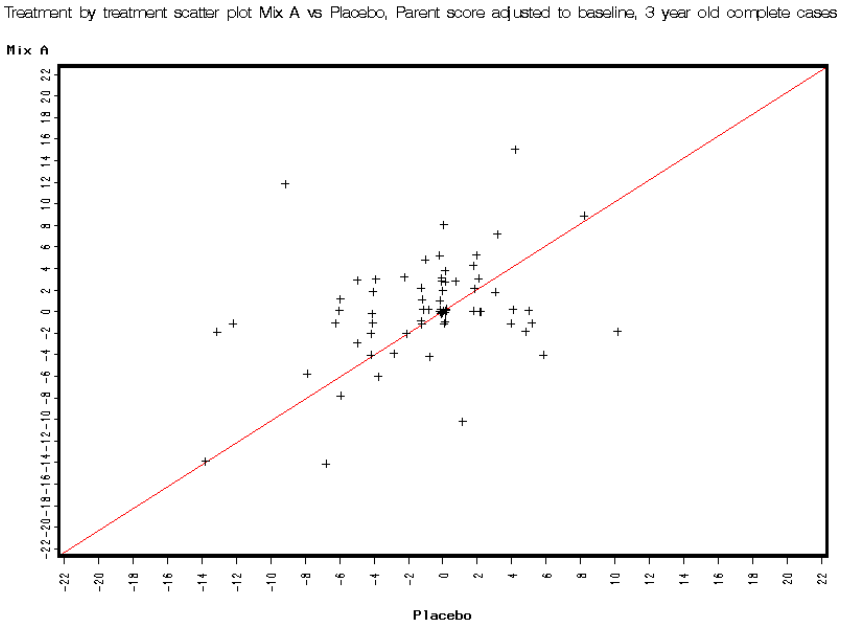


Figure 8-37 Treatment by treatment scatter plot Mix A vs. Placebo, Parent score, 3Y year group, both sexes

Table 8-38 Complete case, 3 year, females:

Test	score	Estimate	StdErr	P-Value
Mix A vs. Placebo	Parent	1.1726	0.762	0.1285
Mix A vs. Placebo	Teacher	1.0946	0.7745	0.1623
Mix A vs. Placebo	Observer	0.4053	1.9061	0.8323
Mix B vs. Placebo	Parent	1.0389	0.7386	0.1641
Mix B vs. Placebo	Teacher	0.7543	0.7506	0.3186
Mix B vs. Placebo	Observer	0.2391	1.8477	0.8974

Table 8-39 Complete case, 8-9 year, both sexes:

Test	score	Estimate	StdErr	P-Value
Mix A vs. Placebo	Parent	0.1776	0.4707	0.7064
Mix A vs. Placebo	Teacher	0.08908	0.4192	0.8319
Mix A vs. Placebo	Observer	0.4485	1.8512	0.8089
Mix A vs. Placebo	CPTCom	2.5671	2.0325	0.2087
Mix A vs. Placebo	CPThit	0.8879	0.9376	0.3453
Mix A vs. Placebo	CPTdpr	0.006658	0.03938	0.866
Mix A vs. Placebo	CPTbet	-0.02996	0.07937	0.7064
Mix B vs. Placebo	Parent	0.6808	0.4733	0.152
Mix B vs. Placebo	Teacher	0.2648	0.4199	0.529
Mix B vs. Placebo	Observer	1.9962	1.8542	0.2831
Mix B vs. Placebo	CPTCom	3.1417	2.0452	0.1268
Mix B vs. Placebo	CPThit	1.6799	0.9434	0.0772
Mix B vs. Placebo	CPTdpr	-0.03562	0.03963	0.3703
Mix B vs. Placebo	CPTbet	0.08221	0.07986	0.3051

Table 8-40 Complete case, 8-9 year, males:

Test	score	Estimate	StdErr	P-Value
Mix A vs. Placebo	Parent	0.1285	0.6145	0.8348
Mix A vs. Placebo	Teacher	0.09107	0.6764	0.8932
Mix A vs. Placebo	Observer	-0.08674	2.779	0.9752
Mix A vs. Placebo	**CPTCom**	**5.5423**	**2.2612**	**0.0166**
Mix A vs. Placebo	CPThit	0.4741	1.2699	0.71
Mix A vs. Placebo	CPTdpr	-0.01048	0.04925	0.832
Mix A vs. Placebo	CPTbet	-0.07036	0.1213	0.5636
Mix B vs. Placebo	Parent	0.5417	0.6156	0.381
Mix B vs. Placebo	Teacher	0.3069	0.6716	0.6486
Mix B vs. Placebo	Observer	3.5794	2.7594	0.1975
Mix B vs. Placebo	**CPTCom**	**6.3442**	**2.2428**	**0.006**
Mix B vs. Placebo	CPThit	1.5185	1.2596	0.2318
Mix B vs. Placebo	CPTdpr	-0.05424	0.04885	0.2705
Mix B vs. Placebo	CPTbet	0.07258	0.1203	0.5482

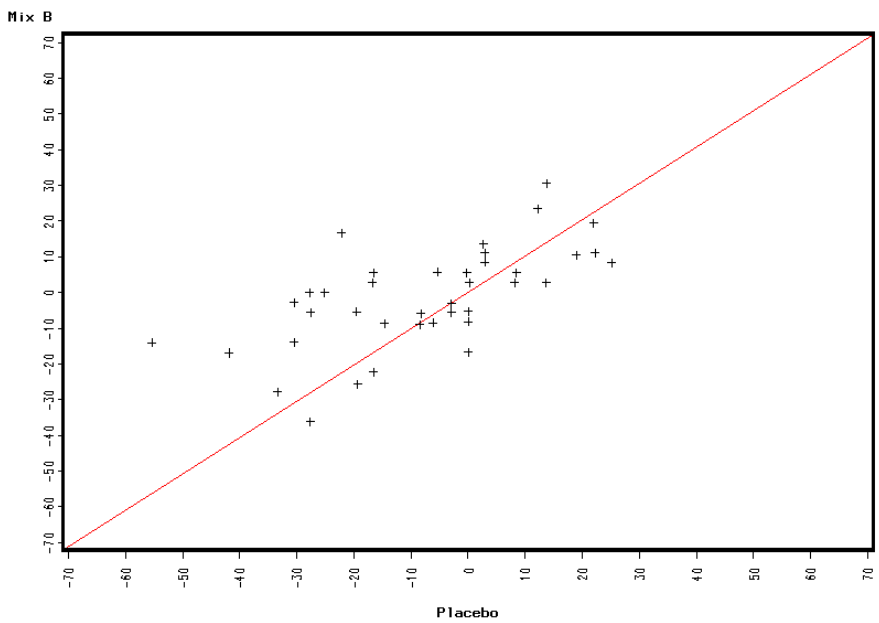


Figure 8-41 Treatment by treatment scatter plot Mix B vs. Placebo, CPTCom score, 8/9Y year group, males

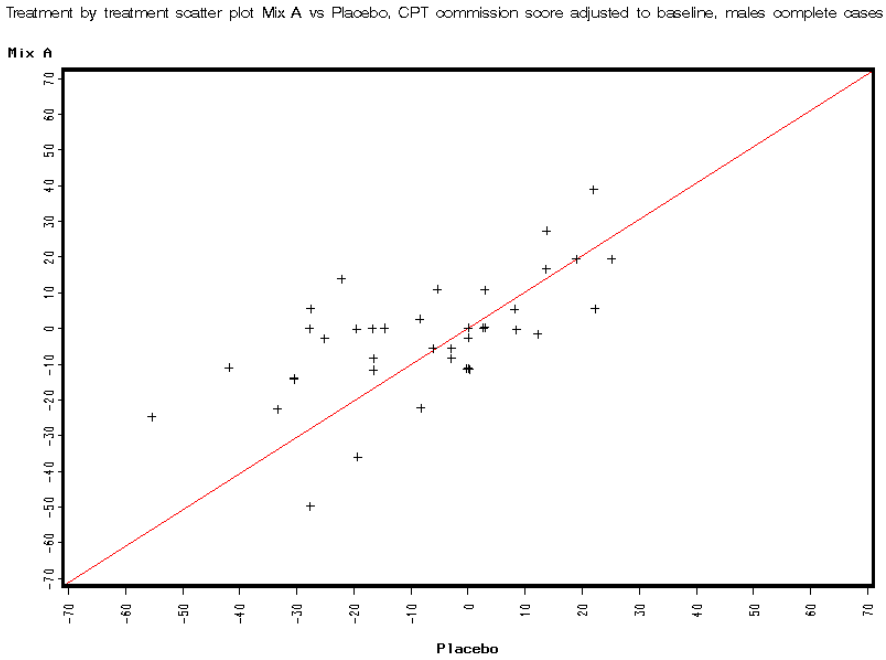


Figure 8-42 Treatment by treatment scatter plot Mix A vs. Placebo, CPTCom score, 8/9Y year group, males

Table 8-43 Complete case, 8-9 year, females:

Test	score	Estimate	StdErr	P-Value
Mix A vs. Placebo	Parent	0.2266	0.772	0.7699
Mix A vs. Placebo	Teacher	0.141	0.4014	0.7265
Mix A vs. Placebo	Observer	1.8074	2.3048	0.4355
Mix A vs. Placebo	CPTCom	-1.7885	3.5952	0.6207
Mix A vs. Placebo	CPThit	0.7744	1.4192	0.5874
Mix A vs. Placebo	CPTdpr	0.05453	0.06497	0.4046
Mix A vs. Placebo	CPTbet	0.02522	0.104	0.8093
Mix B vs. Placebo	Parent	0.6839	0.784	0.3859
Mix B vs. Placebo	Teacher	0.06091	0.4077	0.8816
Mix B vs. Placebo	Observer	-0.01938	2.3406	0.9934
Mix B vs. Placebo	CPTCom	-2.0092	3.7482	0.5939
Mix B vs. Placebo	CPThit	2.2789	1.4791	0.1287
Mix B vs. Placebo	CPTdpr	-0.00642	0.06774	0.9249
Mix B vs. Placebo	CPTbet	0.1013	0.1084	0.3538

Table 8-44 Summary of all significant cases found in treatment group comparisons, for each age group:

Test	score	Year Group	Sex	Estimate	StdErr	P-Value
Entire sample						
Mix A vs. Placebo	CPTCom	8/9Y	M	5.0607	2.1497	0.0206
Mix B vs. Placebo	Parent	8/9Y	both	0.9017	0.4186	0.0322
≥85% consumption						
Mix A vs. Placebo	CPTCom	8/9Y	M	5.0867	2.0922	0.0172
Mix B vs. Placebo	CPTCom	8/9Y	M	6.0375	2.0077	0.0035
Mix B vs. Placebo	Parent	8/9Y	both	0.9687	0.4584	0.0359
Complete case						
Mix A vs. Placebo	CPTCom	8/9Y	M	5.5414	2.2611	0.0166
Mix B vs. Placebo	CPTCom	8/9Y	M	6.3442	2.2427	0.006
Mix A vs. Placebo	Parent	3Y	both	1.2448	0.5585	0.0274

Table 8-45 Summary of all significant cases found in treatment group comparisons, when age groups are pooled:

Test	score	Year Group	Sex	Estimate	StdErr	P-Value
Entire sample						
Mix A vs. Placebo	Observer	both	F	2.1135	1.0503	0.0453
Mix B vs. Placebo	Parent	both	both	0.6385	0.2925	0.0295
≥85% consumption						
Mix B vs. Placebo	Parent	both	both	0.788	0.3265	0.0163
Complete case						
Mix B vs. Placebo	Parent	both	both	0.7726	0.3611	0.0332

8.2 Tests of the period (week of treatment) effect

Hereafter are reported all significant cases observed for "week of treatment" effect tests, for all the contrasts investigated.

Table 8-46 Summary of all significant tests (period and treatment effects), using the entire sample (P-Values <0.05)

Test	score	Year Group	sex	Estimate	StdErr	P-Value
Week 2 vs. Week 6	Observer	3Y	both	4.5922	1.341	0.0008
Week 4 vs. Week 6	Observer	3Y	both	4.0797	1.3521	0.003
Week 2 vs. Week 6	Observer	3Y	M	5.9341	2.0085	0.0043
Week 4 vs. Week 6	Observer	3Y	M	5.6344	2.0041	0.0064
Week 2 vs. Week 6	CPThit	both	both	-5.1107	0.9365	<.0001
Week 2 vs. Week 6	CPThit	both	F	-6.7067	1.4192	<.0001
Week 4 vs. Week 6	Observer	both	both	4.2824	1.2004	0.0004
Week 2 vs. Week 6	CPThit	both	M	-3.7591	1.2661	0.004
Week 4 vs. Week 6	Observer	both	M	4.9287	1.8274	0.0077
Week 2 vs. Week 6	Parent	both	M	1.3042	0.4921	0.0088
Week 4 vs. Week 6	Parent	both	M	1.2231	0.4911	0.0137
Week 2 vs. Week 6	Parent	both	both	0.8506	0.3599	0.0187
Week 4 vs. Week 6	Observer	both	F	3.4398	1.4818	0.0217
Week 4 vs. Week 6	CPThit	both	both	-1.9948	0.9407	0.0358
Week 2 vs. Week 6	CPTdpr	both	F	0.1339	0.06497	0.0437
Week 4 vs. Week 6	CPThit	both	M	-2.5152	1.2596	0.0495
Week 2 vs. Week 6	CPThit	8/9Y	F	-6.7067	1.4192	<.0001
Week 2 vs. Week 6	CPThit	8/9Y	both	-5.1107	0.9365	<.0001
Week 2 vs. Week 6	CPThit	8/9Y	M	-3.7591	1.2661	0.004
Week 2 vs. Week 6	Parent	8/9Y	both	1.2811	0.4723	0.0073
Week 4 vs. Week 6	Observer	8/9Y	both	4.4394	1.8512	0.0175
Week 2 vs. Week 6	Parent	8/9Y	M	1.4906	0.6176	0.0176
Week 4 vs. Week 6	CPThit	8/9Y	both	-1.9948	0.9407	0.0358
Week 2 vs. Week 6	CPTdpr	8/9Y	F	0.1339	0.06497	0.0437
Week 4 vs. Week 6	CPThit	8/9Y	M	-2.5152	1.2596	0.0495

8.3 Conclusions and comments on the single scores results

It is important to note that the conclusions drawn from these *post hoc* analyses are necessarily highly tentative. The main results are as follows.

- No statistically significant component effects were observed that did not coincide with effects seen already in the authors' overall GHA analysis, and no statistically significant effects were seen in the females alone.
- For the 3 year olds, only the Mix A versus Placebo effect with the parental score was statistically significant (complete case only, P=0.027, male and female combined). Neither of other components, teacher or observer, showed any evidence of an effect in any consumption group.
- For the 8-9 year olds, statistically significant Mix B versus Placebo effects were seen for the parental score when both sexes pooled (entire sample P=0.03, high consumers P=0.04) and Computer Commission scores for males, (high consumers P=0.004, complete case P=0.006, males only). For the Computer Commission score also there were statistically significant Mix A versus Placebo effects in the males (entire sample P=0.02, high consumers P=0.02, complete case P=0.02).
- Full consistency across consumption groups could not be observed except for the Computer Commission score for males. There was no obvious pattern over consumption groups.
- All statistically significant component effects have a positive sign, i.e. the component scores were greater on average under the active treatments.

In conclusion there is a suggestion from these analyses that the statistically significant effects seen in the 3 year olds (Mix A versus Placebo) and in the 8-9 year olds (Mix B versus Placebo) are largely driven in the data by the parental scores and, in the older males in both comparisons, by the Computer Commission score. It was why an additional analysis of GHA scores without parental score has been performed as reported in Section 9.

Section 8.2 shows fairly strong statistical evidence for a period or "week" effect, for all scores except the teacher and the Computer Beta and Commission scores which do not show significance in any case. There is no general pattern for all scores although it can be noticed that:

- This effect goes in the sense of decreasing of hyperactivity over time or in the sense of a "learning" trend for the Computer Hit Rate score which decreases with time;
- It shows a peak of hyperactivity at week 4 for the Observer and Parent scores.

The significant cases shown were in general consistent with those found in Model 2 of the McCann *et al.* paper.

Note that for the significant cases observed for period and treatment effects, the size of effect was of the same order of magnitude.

9 Results of the additional analyses

9.1 Histograms by treatment groups for Parent and Computer Commission scores

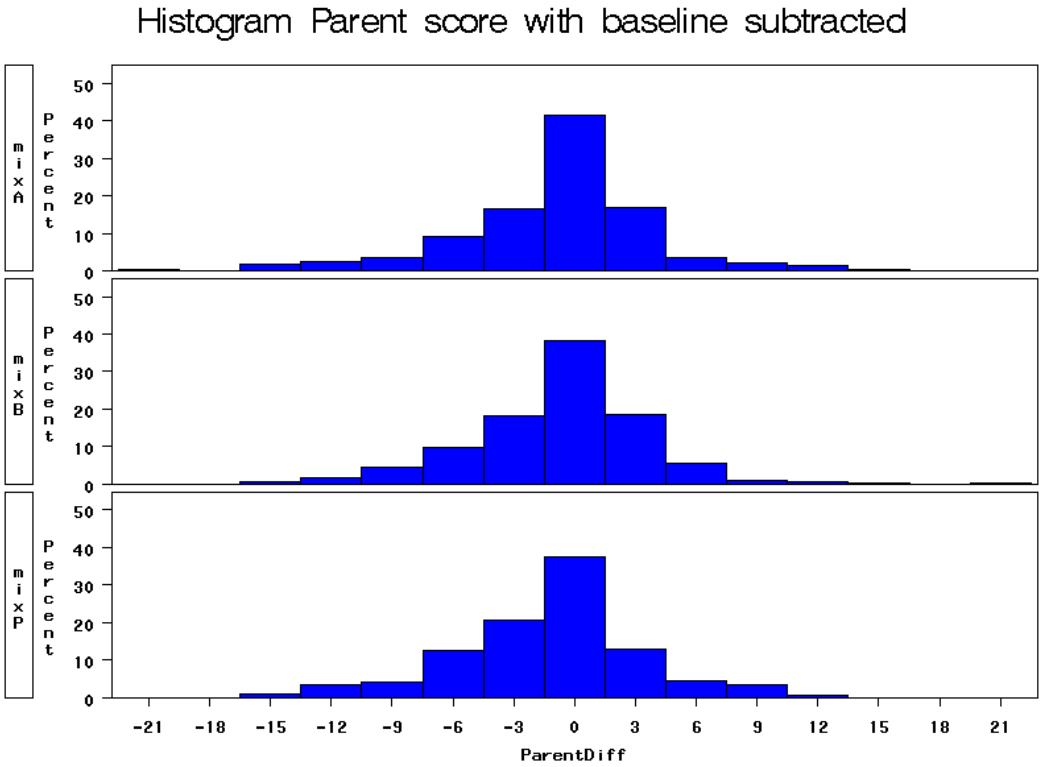


Figure 9-1 Histograms by treatment groups for Parent scores adjusted to individual baselines, Entire sample

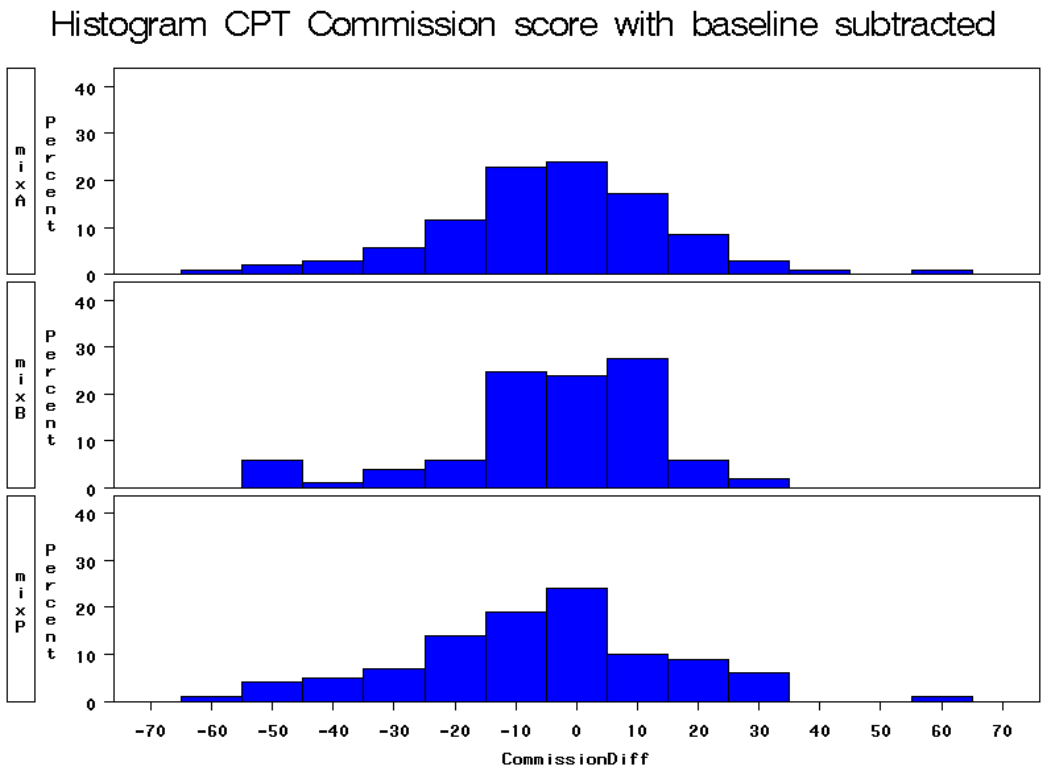


Figure 9-2 Histograms by treatment groups for Computer Commission scores adjusted to individual baselines, Entire sample

From the histograms in **Figure 9-1** and **Figure 9-2**, no obvious subgroup of responders could be identified.

9.2 Analysis of the GHA without parental score

The analysis was done for pooled and non-pooled sexes as well as for age groups pooled and non-pooled, for each consumption group. Among the 54 treatment contrasts investigated, only one appeared to be significant, namely Mix B vs. Placebo for 8-9 year old completers with sexes pooled (P=0.0421).

9.3 Analysis of the interaction Baseline*Treatment for the single scores

The output of this model-based analysis was a set of p-values from testing the effects of baseline values and their interaction with the treatment group, for each single variable. The analysis was done for each age group separately and for each consumption group. The results are reported in **Table 9-3** below. It showed that the Baseline mean effect was, as expected, significant in most cases. On the other hand, the interaction terms Baseline*Treatment appeared to be much less significant for high consumers and completers. However, an effect of the Baseline*Treatment is much

clearly evident for the Teacher score in the whole data set analysis. As it is not consistent across consumption groups, it would suggest that this may be due to external confounding variables such as School.

Table 9-3 Summary of all significant cases found for Baseline*Treatment interactions in single variables:

Test	score	Year Group	Sex	Estimate	StdErr	P-Value
Entire sample						
Baseline*MixA	Teacher	3Y	M	-0.4916	0.08161	<.0001
Baseline*MixB	Teacher	3Y	M	-0.4033	0.08074	<.0001
Baseline*MixB	Teacher	3Y	F	-0.4139	0.09357	<.0001
Baseline*MixB	Teacher	8/9Y	M	-0.356	0.06391	<.0001
Baseline*MixA	Teacher	8/9Y	M	-0.2426	0.06362	0.0002
Baseline*MixA	Teacher	8/9Y	F	-0.3043	0.07807	0.0002
Baseline*MixA	Teacher	3Y	F	-0.3219	0.08598	0.0003
Baseline*MixB	Teacher	8/9Y	F	-0.2455	0.07922	0.0024
Baseline*MixB	CPTCom	8/9Y	M	0.3516	0.1238	0.0055
Baseline*MixA	Observer	8/9Y	both	-0.1713	0.06658	0.0107
Baseline*MixA	Observer	8/9Y	M	-0.229	0.09213	0.0142
≥85% consumption						
Baseline*MixA	Observer	8/9Y	M	-0.2302	0.0944	0.0164
Baseline*MixA	Teacher	3Y	M	-0.2257	0.09798	0.0237
Baseline*MixA	Observer	8/9Y	both	-0.1604	0.07213	0.0273
Baseline*MixB	Parent	3Y	M	0.6299	0.3116	0.0461
Complete case						
Baseline*MixA	Observer	8/9Y	both	-0.1838	0.075	0.0152
Baseline*MixA	Observer	8/9Y	M	-0.2359	0.09614	0.0159
Baseline*MixA	Teacher	3Y	M	-0.2803	0.1167	0.019

10 Overall conclusions and discussion

The statistical approach proposed in this report for the analysis of the aggregated data from the Southampton Study provides some minor improvements over that originally used, although none of the modifications would be expected to make major differences to the conclusions. The supplementary analyses allow some tentative suggestions to be made concerning the impact of the individual component scores on the results seen from the aggregates.

As expected the results from the analysis of the aggregated scores were broadly consistent with the findings of the McCann *et al.* Study.

The analysis of single scores was in general consistent with the findings of the McCann *et al.* study. It shows that treatment groups show an increase in the Parent score for both age groups of children and in the Computer Commission scores for 8/9-year-old boys. The overall conclusion therefore supports that of the original paper, with slightly reduced statistical significance. Only two real effects are suggested, albeit somewhat weakly, from this analysis: the Mix A versus Placebo comparison in the three year olds and the Mix B versus Placebo comparison in the 8-9 year olds. It is noted that the effects do not appear to be supported consistent across age groups, neither across component measures, sexes or consumption groups.

A 'Week' effect was shown on both single and aggregated scores but only for the "Week 4 vs. Week 6" comparison in the latter. The general trend was that hyperactivity generally went up from Week 2 to Week 4 and then significantly decreased from Week 4 to Week 6. The size of this period effect was globally of the same order of magnitude as those observed for the treatment effects, in both single and aggregated scores. Finally, it appears that the Placebo group was consistently showing a decrease compared to Week 0 (baseline), such a decrease was significant in about half of the cases. It illustrates how large is the intra-individual variability over time and interpretation of all statistical results should be done in the light of this result.

11 References

Bonate P., Howard D. (2004). *Pharmacokinetics in drug development* (vol. 1 and 2), Springer

Bozdogan, H. (1987). Model selection and Akaike's information criteria (AIC): the general theory and its analytical extensions. *Psychometrika* 52, 345-370

Chatterjee S., Hadi A. S., Price B., (2000) *Regression analysis by example*, Ed Lavoisier

Gelman A., Meng X.-L., Stern H. (1996) Posterior predicitive assessment of model fitness via realized discrepancies, *Statistica Sinica* 6, 733-807

Hosmer D.W. and Lemeshow S. (1989). *Applied Logistic Regression*. New York : John Wiley and Sons

McCann D., Barrett A., Cooper A., Crumpler D., Dalen L., Grimshaw K., Kitchin E., Lok K., Porteous L., Prince E., Sonuga-Barke E., Warner J.O., Stevenson J. (2007). Food additives and hyperactive behaviour in 3 and 8/9 year old children in the community. *The Lancet* (on line 6 September 2007)

Milliken G.A. and Johnson D.E. (1992), *Analysis of Messy Data*, volI, Chapman &Hall

Molenberghs G. and Verbeke G. (2006). *Models for discrete longitudinal data*, Springer

Senn SJ. (1993). *Cross over trials in clinical research*, New-York, John Wiley and Sons.

Senn SJ. (1997) Are placebo run ins justified? *British Medical Journal*; 314(7088):1191-3.

Senn SJ. (2001) The Misunderstood Placebo. *Applied Clinical Trials*;10(5):40-46

Appendix 1: Variables tested in the mixed models

Model 1 (both age groups, entire sample and high consumers)

Subjects	Participant ID
Repeated	Week of study
Dependent	Valid Agg
Factors	Treatment
Covariates	none
Fixed effect	Treatment
Random effect	Participant . ID
Estimated means	Treatment .

Model 2 (both age groups, entire sample and high consumers)

Subjects	Participant ID
Repeated	Week of study
Dependent	Valid Agg
Factors	Treatment, Week of study, Sex, Rschooling, Msoclev
Covariates	Aggwk0, Foodadd
Fixed effect	Treatment , Week of study, Sex, Rschool. Msoclev, Aggwk0, Foodadd
Random effect	Participant . ID
Estimated means	Treatment

Data dictionary

Participant ID	ID nr of participant	Categorical	
Week of study			
Valid Agg	z score of Aggregate behaviour scores in challenge weeks	Continuous	
Treatment	Which mix received at a given week of study	Categorical 3 levels	1. Mix A vs. Placebo 2. Mix B 3. placebo
Sex	Sex of the participant	Categorical 2 levels	1. Male 2. Female
Rschool			
Msoclev	Mother's social Class level	Categorical 2 levels	1. Lower occupations and long term unemployed 2. Intermediate and higher occupation
Aggwk0			
Foodadd	Pre-study 24hr recall of number of additives consumed	Discrete	Count of the number of food items containing additive eaten in a 24 hour period before the trial. Range:0-6

Appendix 2: Mixed-model outputs

For Entire sample:

Test	score	Year Group	sex	Estimate	StdErr	P-Value
Week 2 vs. Week 6	Parent	3Y	M	0.3899	0.5967	0.5146
Week 2 vs. Week 6	Parent	3Y	F	0.1095	0.5468	0.8415
Week 2 vs. Week 6	Teacher	3Y	M	0.4735	0.5944	0.4273
Week 2 vs. Week 6	Teacher	3Y	F	-0.3376	0.5411	0.5338
Week 2 vs. Week 6	Observer	3Y	M	3.7604	1.4793	0.0122
Week 2 vs. Week 6	Observer	3Y	F	1.4446	1.3091	0.272
Week 2 vs. Week 6	Parent	3Y	both	0.2306	0.4076	0.572
Week 2 vs. Week 6	Teacher	3Y	both	0.0991	0.4009	0.805
Week 2 vs. Week 6	Observer	3Y	both	2.685	0.982	0.0067
Week 4 vs. Week 6	Parent	3Y	M	1.2354	0.6068	0.0437
Week 4 vs. Week 6	Parent	3Y	F	-0.7583	0.5532	0.1729
Week 4 vs. Week 6	Teacher	3Y	M	-0.05496	0.6081	0.9281
Week 4 vs. Week 6	Teacher	3Y	F	0.05864	0.5398	0.9137
Week 4 vs. Week 6	Observer	3Y	M	3.6006	1.5128	0.0188
Week 4 vs. Week 6	Observer	3Y	F	1.1654	1.3207	0.3793
Week 4 vs. Week 6	Parent	3Y	both	0.271	0.4124	0.5117
Week 4 vs. Week 6	Teacher	3Y	both	-0.02738	0.4045	0.9461
Week 4 vs. Week 6	Observer	3Y	both	2.4296	0.9964	0.0154
Mix A vs. Placebo	Parent	3Y	M	0.5895	0.6036	0.3305
Mix A vs. Placebo	Parent	3Y	F	0.2933	0.5538	0.5972
Mix A vs. Placebo	Teacher	3Y	M	-0.2682	0.6009	0.6562
Mix A vs. Placebo	Teacher	3Y	F	0.2744	0.5406	0.6127
Mix A vs. Placebo	Observer	3Y	M	0.8724	1.4954	0.5606
Mix A vs. Placebo	Observer	3Y	F	1.3643	1.3085	0.2992
Mix A vs. Placebo	Parent	3Y	both	0.5615	0.4097	0.1717
Mix A vs. Placebo	Teacher	3Y	both	-0.05585	0.4009	0.8893
Mix A vs. Placebo	Observer	3Y	both	1.0569	0.9818	0.2827
Mix B vs. Placebo	Parent	3Y	M	0.366	0.5988	0.5421
Mix B vs. Placebo	Parent	3Y	F	0.3998	0.5448	0.4644
Mix B vs. Placebo	Teacher	3Y	M	0.1665	0.5879	0.7775
Mix B vs. Placebo	Teacher	3Y	F	0.301	0.5403	0.5785
Mix B vs. Placebo	Observer	3Y	M	-1.0092	1.4719	0.4942
Mix B vs. Placebo	Observer	3Y	F	0.731	1.3107	0.578
Mix B vs. Placebo	Parent	3Y	both	0.3825	0.4081	0.3494
Mix B vs. Placebo	Teacher	3Y	both	0.2279	0.399	0.5684
Mix B vs. Placebo	Observer	3Y	both	-0.2815	0.982	0.7746
Week 2 vs. Week 6	Parent	both	both	0.6985	0.2925	0.0173
Week 2 vs. Week 6	Teacher	both	both	0.07775	0.2652	0.7695
Week 2 vs. Week 6	Observer	both	both	1.7698	0.8893	0.0471
Week 2 vs. Week 6	CPTCom	both	both	-1.8369	1.6816	0.276

Test	score	Year Group	sex	Estimate	StdErr	P-Value
Week 2 vs. Week 6	CPThit	both	both	-4.1397	0.7839	<.0001
Week 2 vs. Week 6	CPTdpr	both	both	0.05818	0.03291	0.0786
Week 2 vs. Week 6	CPTbet	both	both	0.003789	0.07069	0.9573
Week 2 vs. Week 6	Parent	both	M	0.8232	0.4236	0.053
Week 2 vs. Week 6	Parent	both	F	0.5585	0.3998	0.1637
Week 2 vs. Week 6	Teacher	both	M	0.2172	0.414	0.6003
Week 2 vs. Week 6	Teacher	both	F	-0.0891	0.3281	0.7862
Week 2 vs. Week 6	Observer	both	M	1.8734	1.4133	0.1861
Week 2 vs. Week 6	Observer	both	F	1.7016	1.0548	0.108
Week 2 vs. Week 6	CPTCom	both	M	-0.9258	2.1346	0.6654
Week 2 vs. Week 6	CPTCom	both	F	-3.8157	2.6608	0.1549
Week 2 vs. Week 6	CPThit	both	M	-3.5874	1.2237	0.0042
Week 2 vs. Week 6	CPThit	both	F	-4.6967	1.032	<.0001
Week 2 vs. Week 6	CPTdpr	both	M	0.008453	0.04219	0.8416
Week 2 vs. Week 6	CPTdpr	both	F	0.1081	0.05218	0.0409
Week 2 vs. Week 6	CPTbet	both	M	0.0213	0.1015	0.8342
Week 2 vs. Week 6	CPTbet	both	F	-0.02696	0.1052	0.7983
Week 4 vs. Week 6	Parent	both	both	0.6441	0.2941	0.029
Week 4 vs. Week 6	Teacher	both	both	0.1656	0.266	0.534
Week 4 vs. Week 6	Observer	both	both	3.6252	0.8961	<.0001
Week 4 vs. Week 6	CPTCom	both	both	-1.6531	1.6797	0.3262
Week 4 vs. Week 6	CPThit	both	both	-1.5533	0.7831	0.0487
Week 4 vs. Week 6	CPTdpr	both	both	0.02005	0.03288	0.5427
Week 4 vs. Week 6	CPTbet	both	both	0.0639	0.07065	0.3668
Week 4 vs. Week 6	Parent	both	M	1.39	0.4254	0.0012
Week 4 vs. Week 6	Parent	both	F	-0.168	0.4026	0.6767
Week 4 vs. Week 6	Teacher	both	M	0.2117	0.4168	0.6119
Week 4 vs. Week 6	Teacher	both	F	0.09633	0.3275	0.7689
Week 4 vs. Week 6	Observer	both	M	4.6505	1.4251	0.0012
Week 4 vs. Week 6	Observer	both	F	2.4881	1.0617	0.0199
Week 4 vs. Week 6	CPTCom	both	M	0.332	2.1281	0.8764
Week 4 vs. Week 6	CPTCom	both	F	-4.0766	2.6766	0.1311
Week 4 vs. Week 6	CPThit	both	M	-1.8359	1.2202	0.1357
Week 4 vs. Week 6	CPThit	both	F	-1.4915	1.0385	0.1542
Week 4 vs. Week 6	CPTdpr	both	M	0.009444	0.04207	0.8228
Week 4 vs. Week 6	CPTdpr	both	F	0.02564	0.0525	0.6263
Week 4 vs. Week 6	CPTbet	both	M	0.05676	0.1012	0.5762
Week 4 vs. Week 6	CPTbet	both	F	0.06108	0.106	0.5657
Mix A vs. Placebo	Parent	both	both	0.3552	0.2933	0.2265
Mix A vs. Placebo	Teacher	both	both	-0.1729	0.2655	0.5152
Mix A vs. Placebo	Observer	both	both	0.947	0.8911	0.2884
Mix A vs. Placebo	CPTCom	both	both	2.1591	1.6892	0.2027
Mix A vs. Placebo	CPThit	both	both	0.4611	0.7875	0.5589
Mix A vs. Placebo	CPTdpr	both	both	-0.01725	0.03306	0.6025
Mix A vs. Placebo	CPTbet	both	both	0.01165	0.07102	0.8699
Mix A vs. Placebo	Parent	both	M	0.452	0.4237	0.287

Test	score	Year Group	sex	Estimate	StdErr	P-Value
Mix A vs. Placebo	Parent	both	F	0.2334	0.401	0.5612
Mix A vs. Placebo	Teacher	both	M	-0.2474	0.4163	0.5529
Mix A vs. Placebo	Teacher	both	F	-0.1104	0.3259	0.7352
Mix A vs. Placebo	Observer	both	M	-0.1714	1.4212	0.9041
Mix A vs. Placebo	Observer	both	F	2.1135	1.0503	0.0453
Mix A vs. Placebo	CPTCom	both	M	5.0607	2.1497	0.0206
Mix A vs. Placebo	CPTCom	both	F	-0.8642	2.6699	0.7469
Mix A vs. Placebo	CPThit	both	M	0.1799	1.2323	0.8842
Mix A vs. Placebo	CPThit	both	F	0.4012	1.0358	0.6994
Mix A vs. Placebo	CPTdpr	both	M	-0.01613	0.04248	0.7049
Mix A vs. Placebo	CPTdpr	both	F	0.00188	0.05236	0.9714
Mix A vs. Placebo	CPTbet	both	M	0.00444	0.1022	0.9654
Mix A vs. Placebo	CPTbet	both	F	0.009068	0.1057	0.9318
Mix B vs. Placebo	Parent	both	both	0.6385	0.2925	0.0295
Mix B vs. Placebo	Teacher	both	both	0.06112	0.2643	0.8172
Mix B vs. Placebo	Observer	both	both	0.601	0.8897	0.4997
Mix B vs. Placebo	CPTCom	both	both	2.208	1.6884	0.1925
Mix B vs. Placebo	CPThit	both	both	1.0339	0.7871	0.1905
Mix B vs. Placebo	CPTdpr	both	both	-0.03695	0.03304	0.2648
Mix B vs. Placebo	CPTbet	both	both	0.05985	0.07099	0.4002
Mix B vs. Placebo	Parent	both	M	0.7304	0.4237	0.0859
Mix B vs. Placebo	Parent	both	F	0.4571	0.4001	0.2543
Mix B vs. Placebo	Teacher	both	M	0.22	0.4105	0.5926
Mix B vs. Placebo	Teacher	both	F	-0.1352	0.3289	0.6814
Mix B vs. Placebo	Observer	both	M	0.5354	1.4107	0.7046
Mix B vs. Placebo	Observer	both	F	0.5275	1.0592	0.6189
Mix B vs. Placebo	CPTCom	both	M	4.0152	2.1326	0.0627
Mix B vs. Placebo	CPTCom	both	F	-0.2705	2.6898	0.9201
Mix B vs. Placebo	CPThit	both	M	1.5217	1.2228	0.2163
Mix B vs. Placebo	CPThit	both	F	0.456	1.0431	0.663
Mix B vs. Placebo	CPTdpr	both	M	-0.0266	0.04216	0.5296
Mix B vs. Placebo	CPTdpr	both	F	-0.03838	0.05275	0.4686
Mix B vs. Placebo	CPTbet	both	M	0.08054	0.1014	0.4291
Mix B vs. Placebo	CPTbet	both	F	0.02831	0.1062	0.7905
Week 2 vs. Week 6	Parent	8/9Y	M	1.2334	0.6094	0.045
Week 2 vs. Week 6	Parent	8/9Y	F	1.0549	0.5956	0.0792
Week 2 vs. Week 6	Teacher	8/9Y	M	-0.06317	0.5914	0.9151
Week 2 vs. Week 6	Teacher	8/9Y	F	0.07635	0.376	0.8394
Week 2 vs. Week 6	Observer	8/9Y	M	-0.05094	2.4211	0.9832
Week 2 vs. Week 6	Observer	8/9Y	F	2.0222	1.7019	0.2372
Week 2 vs. Week 6	CPTCom	8/9Y	M	-0.9258	2.1346	0.6654
Week 2 vs. Week 6	CPTCom	8/9Y	F	-3.8157	2.6608	0.1549
Week 2 vs. Week 6	CPThit	8/9Y	M	-3.5874	1.2237	0.0042
Week 2 vs. Week 6	CPThit	8/9Y	F	-4.6967	1.032	<.0001
Week 2 vs. Week 6	CPTdpr	8/9Y	M	0.008453	0.04219	0.8416
Week 2 vs. Week 6	CPTdpr	8/9Y	F	0.1081	0.05218	0.0409

Test	score	Year Group	sex	Estimate	StdErr	P-Value
Week 2 vs. Week 6	CPTbet	8/9Y	M	0.0213	0.1015	0.8342
Week 2 vs. Week 6	CPTbet	8/9Y	F	-0.02696	0.1052	0.7983
Week 2 vs. Week 6	Parent	8/9Y	both	1.1669	0.4194	0.0058
Week 2 vs. Week 6	Teacher	8/9Y	both	0.06689	0.3497	0.8485
Week 2 vs. Week 6	Observer	8/9Y	both	0.8733	1.4809	0.5559
Week 2 vs. Week 6	CPTCom	8/9Y	both	-1.8369	1.6816	0.276
Week 2 vs. Week 6	CPThit	8/9Y	both	-4.1397	0.7839	<.0001
Week 2 vs. Week 6	CPTdpr	8/9Y	both	0.05818	0.03291	0.0786
Week 2 vs. Week 6	CPTbet	8/9Y	both	0.003789	0.07069	0.9573
Week 4 vs. Week 6	Parent	8/9Y	M	1.4149	0.6084	0.0216
Week 4 vs. Week 6	Parent	8/9Y	F	0.5296	0.5981	0.3777
Week 4 vs. Week 6	Teacher	8/9Y	M	0.4613	0.5868	0.4332
Week 4 vs. Week 6	Teacher	8/9Y	F	0.1064	0.3757	0.7775
Week 4 vs. Week 6	Observer	8/9Y	M	5.0573	2.4081	0.0376
Week 4 vs. Week 6	Observer	8/9Y	F	3.6833	1.707	0.033
Week 4 vs. Week 6	CPTCom	8/9Y	M	0.332	2.1281	0.8764
Week 4 vs. Week 6	CPTCom	8/9Y	F	-4.0766	2.6766	0.1311
Week 4 vs. Week 6	CPThit	8/9Y	M	-1.8359	1.2202	0.1357
Week 4 vs. Week 6	CPThit	8/9Y	F	-1.4915	1.0385	0.1542
Week 4 vs. Week 6	CPTdpr	8/9Y	M	0.009444	0.04207	0.8228
Week 4 vs. Week 6	CPTdpr	8/9Y	F	0.02564	0.0525	0.6263
Week 4 vs. Week 6	CPTbet	8/9Y	M	0.05676	0.1012	0.5762
Week 4 vs. Week 6	CPTbet	8/9Y	F	0.06108	0.106	0.5657
Week 4 vs. Week 6	Parent	8/9Y	both	1.0296	0.4189	0.0147
Week 4 vs. Week 6	Teacher	8/9Y	both	0.3581	0.3488	0.3056
Week 4 vs. Week 6	Observer	8/9Y	both	4.803	1.4801	0.0013
Week 4 vs. Week 6	CPTCom	8/9Y	both	-1.6531	1.6797	0.3262
Week 4 vs. Week 6	CPThit	8/9Y	both	-1.5533	0.7831	0.0487
Week 4 vs. Week 6	CPTdpr	8/9Y	both	0.02005	0.03288	0.5427
Week 4 vs. Week 6	CPTbet	8/9Y	both	0.0639	0.07065	0.3668
Mix A vs. Placebo	Parent	8/9Y	M	0.1704	0.6079	0.7797
Mix A vs. Placebo	Parent	8/9Y	F	0.1707	0.5993	0.7763
Mix A vs. Placebo	Teacher	8/9Y	M	-0.06219	0.5951	0.9169
Mix A vs. Placebo	Teacher	8/9Y	F	-0.4375	0.3766	0.2478
Mix A vs. Placebo	Observer	8/9Y	M	-0.3251	2.4365	0.8941
Mix A vs. Placebo	Observer	8/9Y	F	2.6862	1.7116	0.1192
Mix A vs. Placebo	CPTCom	8/9Y	M	5.0607	2.1497	0.0206
Mix A vs. Placebo	CPTCom	8/9Y	F	-0.8642	2.6699	0.7469
Mix A vs. Placebo	CPThit	8/9Y	M	0.1799	1.2323	0.8842
Mix A vs. Placebo	CPThit	8/9Y	F	0.4012	1.0358	0.6994
Mix A vs. Placebo	CPTdpr	8/9Y	M	-0.01613	0.04248	0.7049
Mix A vs. Placebo	CPTdpr	8/9Y	F	0.00188	0.05236	0.9714
Mix A vs. Placebo	CPTbet	8/9Y	M	0.00444	0.1022	0.9654
Mix A vs. Placebo	CPTbet	8/9Y	F	0.009068	0.1057	0.9318
Mix A vs. Placebo	Parent	8/9Y	both	0.1391	0.4193	0.7404
Mix A vs. Placebo	Teacher	8/9Y	both	-0.2868	0.3503	0.4136

Test	score	Year Group	sex	Estimate	StdErr	P-Value
Mix A vs. Placebo	Observer	8/9Y	both	0.793	1.4861	0.5941
Mix A vs. Placebo	CPTCom	8/9Y	both	2.1591	1.6892	0.2027
Mix A vs. Placebo	CPThit	8/9Y	both	0.4611	0.7875	0.5589
Mix A vs. Placebo	CPTdpr	8/9Y	both	-0.01725	0.03306	0.6025
Mix A vs. Placebo	CPTbet	8/9Y	both	0.01165	0.07102	0.8699
Mix B vs. Placebo	Parent	8/9Y	M	1.0299	0.6069	0.0921
Mix B vs. Placebo	Parent	8/9Y	F	0.6376	0.5959	0.2869
Mix B vs. Placebo	Teacher	8/9Y	M	0.2913	0.5863	0.6202
Mix B vs. Placebo	Teacher	8/9Y	F	-0.5627	0.3762	0.1374
Mix B vs. Placebo	Observer	8/9Y	M	2.3671	2.4165	0.3291
Mix B vs. Placebo	Observer	8/9Y	F	0.4576	1.7029	0.7886
Mix B vs. Placebo	CPTCom	8/9Y	M	4.0152	2.1326	0.0627
Mix B vs. Placebo	CPTCom	8/9Y	F	-0.2705	2.6898	0.9201
Mix B vs. Placebo	CPThit	8/9Y	M	1.5217	1.2228	0.2163
Mix B vs. Placebo	CPThit	8/9Y	F	0.456	1.0431	0.663
Mix B vs. Placebo	CPTdpr	8/9Y	M	-0.0266	0.04216	0.5296
Mix B vs. Placebo	CPTdpr	8/9Y	F	-0.03838	0.05275	0.4686
Mix B vs. Placebo	CPTbet	8/9Y	M	0.08054	0.1014	0.4291
Mix B vs. Placebo	CPTbet	8/9Y	F	0.02831	0.1062	0.7905
Mix B vs. Placebo	Parent	8/9Y	both	0.9017	0.4186	0.0322
Mix B vs. Placebo	Teacher	8/9Y	both	-0.1082	0.3489	0.7566
Mix B vs. Placebo	Observer	8/9Y	both	1.4383	1.4818	0.3327
Mix B vs. Placebo	CPTCom	8/9Y	both	2.208	1.6884	0.1925
Mix B vs. Placebo	CPThit	8/9Y	both	1.0339	0.7871	0.1905
Mix B vs. Placebo	CPTdpr	8/9Y	both	-0.03695	0.03304	0.2648
Mix B vs. Placebo	CPTbet	8/9Y	both	0.05985	0.07099	0.4002

For ≥85% consumption:

Test	score	Year Group	sex	Estimate	StdErr	P-Value
Week 2 vs. Week 6	Parent	3Y	M	0.3661	0.6722	0.5873
Week 2 vs. Week 6	Parent	3Y	F	0.08004	0.6548	0.903
Week 2 vs. Week 6	Teacher	3Y	M	-0.4231	0.6852	0.5385
Week 2 vs. Week 6	Teacher	3Y	F	0.0812	0.6147	0.8952
Week 2 vs. Week 6	Observer	3Y	M	5.646	1.8367	0.0028
Week 2 vs. Week 6	Observer	3Y	F	2.7469	1.561	0.082
Week 2 vs. Week 6	Parent	3Y	both	0.1477	0.4699	0.7536
Week 2 vs. Week 6	Teacher	3Y	both	-0.09265	0.4584	0.84
Week 2 vs. Week 6	Observer	3Y	both	4.2095	1.1949	0.0005
Week 4 vs. Week 6	Parent	3Y	M	1.3952	0.6756	0.0417
Week 4 vs. Week 6	Parent	3Y	F	-0.9971	0.6415	0.1234
Week 4 vs. Week 6	Teacher	3Y	M	-0.8016	0.6961	0.2527
Week 4 vs. Week 6	Teacher	3Y	F	0.4668	0.5947	0.4347
Week 4 vs. Week 6	Observer	3Y	M	5.9359	1.8662	0.002

Test	score	Year Group	sex	Estimate	StdErr	P-Value
Week 4 vs. Week 6	Observer	3Y	F	1.8789	1.5355	0.2244
Week 4 vs. Week 6	Parent	3Y	both	0.2158	0.4671	0.6446
Week 4 vs. Week 6	Teacher	3Y	both	-0.1665	0.4547	0.7147
Week 4 vs. Week 6	Observer	3Y	both	3.8874	1.194	0.0014
Mix A vs. Placebo	Parent	3Y	M	0.9864	0.6581	0.1373
Mix A vs. Placebo	Parent	3Y	F	0.1628	0.6637	0.8068
Mix A vs. Placebo	Teacher	3Y	M	-0.3435	0.6635	0.606
Mix A vs. Placebo	Teacher	3Y	F	0.6647	0.6218	0.2881
Mix A vs. Placebo	Observer	3Y	M	1.2439	1.7786	0.4861
Mix A vs. Placebo	Observer	3Y	F	1.0705	1.5947	0.5038
Mix A vs. Placebo	Parent	3Y	both	0.7301	0.4663	0.1191
Mix A vs. Placebo	Teacher	3Y	both	0.1096	0.4523	0.8088
Mix A vs. Placebo	Observer	3Y	both	1.2026	1.1839	0.3111
Mix B vs. Placebo	Parent	3Y	M	0.5061	0.6656	0.449
Mix B vs. Placebo	Parent	3Y	F	0.5362	0.6421	0.4058
Mix B vs. Placebo	Teacher	3Y	M	0.237	0.6675	0.7234
Mix B vs. Placebo	Teacher	3Y	F	0.6884	0.6003	0.2547
Mix B vs. Placebo	Observer	3Y	M	-1.3797	1.7896	0.4428
Mix B vs. Placebo	Observer	3Y	F	0.7843	1.5396	0.6117
Mix B vs. Placebo	Parent	3Y	both	0.5582	0.464	0.2305
Mix B vs. Placebo	Teacher	3Y	both	0.4874	0.4489	0.2791
Mix B vs. Placebo	Observer	3Y	both	-0.399	1.1745	0.7345
Week 2 vs. Week 6	Parent	both	both	0.69	0.3304	0.0374
Week 2 vs. Week 6	Teacher	both	both	-0.146	0.3013	0.6283
Week 2 vs. Week 6	Observer	both	both	1.6457	1.0648	0.1231
Week 2 vs. Week 6	CPTCom	both	both	-1.3115	1.8528	0.4801
Week 2 vs. Week 6	CPThit	both	both	-4.7883	0.8685	<.0001
Week 2 vs. Week 6	CPTdpr	both	both	0.04128	0.0361	0.2547
Week 2 vs. Week 6	CPTbet	both	both	-0.0141	0.07265	0.8464
Week 2 vs. Week 6	Parent	both	M	0.9468	0.4584	0.0401
Week 2 vs. Week 6	Parent	both	F	0.5036	0.472	0.2874
Week 2 vs. Week 6	Teacher	both	M	-0.2649	0.4656	0.57
Week 2 vs. Week 6	Teacher	both	F	0.02094	0.3676	0.9546
Week 2 vs. Week 6	Observer	both	M	1.5319	1.6634	0.3581
Week 2 vs. Week 6	Observer	both	F	1.901	1.2773	0.1385
Week 2 vs. Week 6	CPTCom	both	M	0.2983	2.0851	0.8866
Week 2 vs. Week 6	CPTCom	both	F	-4.3194	3.1731	0.178
Week 2 vs. Week 6	CPThit	both	M	-3.7977	1.2083	0.0023
Week 2 vs. Week 6	CPThit	both	F	-6.0042	1.2897	<.0001
Week 2 vs. Week 6	CPTdpr	both	M	-0.00976	0.04565	0.8312
Week 2 vs. Week 6	CPTdpr	both	F	0.1167	0.05837	0.0497
Week 2 vs. Week 6	CPTbet	both	M	-0.0091	0.1107	0.9346
Week 2 vs. Week 6	CPTbet	both	F	-0.01188	0.09733	0.9032
Week 4 vs. Week 6	Parent	both	both	0.5969	0.3285	0.07
Week 4 vs. Week 6	Teacher	both	both	0.1941	0.2992	0.5168
Week 4 vs. Week 6	Observer	both	both	4.2055	1.0626	<.0001

Test	score	Year Group	sex	Estimate	StdErr	P-Value
Week 4 vs. Week 6	CPTCom	both	both	-0.6818	1.8375	0.7111
Week 4 vs. Week 6	CPThit	both	both	-1.8608	0.8614	0.0323
Week 4 vs. Week 6	CPTdpr	both	both	0.01308	0.03579	0.7152
Week 4 vs. Week 6	CPTbet	both	both	0.03677	0.07203	0.6105
Week 4 vs. Week 6	Parent	both	M	1.4569	0.4583	0.0017
Week 4 vs. Week 6	Parent	both	F	-0.3358	0.4677	0.4737
Week 4 vs. Week 6	Teacher	both	M	0.09621	0.4671	0.837
Week 4 vs. Week 6	Teacher	both	F	0.2947	0.3615	0.4161
Week 4 vs. Week 6	Observer	both	M	5.0776	1.669	0.0027
Week 4 vs. Week 6	Observer	both	F	3.1691	1.27	0.0135
Week 4 vs. Week 6	CPTCom	both	M	1.5729	2.0585	0.447
Week 4 vs. Week 6	CPTCom	both	F	-4.4719	3.2145	0.1688
Week 4 vs. Week 6	CPThit	both	M	-2.258	1.1929	0.0619
Week 4 vs. Week 6	CPThit	both	F	-1.2636	1.3063	0.3369
Week 4 vs. Week 6	CPTdpr	both	M	-0.00439	0.04507	0.9226
Week 4 vs. Week 6	CPTdpr	both	F	0.03805	0.05916	0.5223
Week 4 vs. Week 6	CPTbet	both	M	0.0524	0.1093	0.6328
Week 4 vs. Week 6	CPTbet	both	F	0.02528	0.09835	0.7979
Mix A vs. Placebo	Parent	both	both	0.4784	0.3303	0.1483
Mix A vs. Placebo	Teacher	both	both	0.04531	0.3008	0.8804
Mix A vs. Placebo	Observer	both	both	1.2068	1.0666	0.2586
Mix A vs. Placebo	CPTCom	both	both	2.0004	1.8664	0.2855
Mix A vs. Placebo	CPThit	both	both	0.2699	0.8749	0.7581
Mix A vs. Placebo	CPTdpr	both	both	0.01144	0.03635	0.7533
Mix A vs. Placebo	CPTbet	both	both	-0.03885	0.07316	0.5962
Mix A vs. Placebo	Parent	both	M	0.632	0.4527	0.1642
Mix A vs. Placebo	Parent	both	F	0.2703	0.4772	0.5718
Mix A vs. Placebo	Teacher	both	M	-0.1771	0.4612	0.7014
Mix A vs. Placebo	Teacher	both	F	0.3327	0.3694	0.3689
Mix A vs. Placebo	Observer	both	M	0.3901	1.6478	0.8131
Mix A vs. Placebo	Observer	both	F	2.135	1.293	0.1005
Mix A vs. Placebo	CPTCom	both	M	5.0867	2.0922	0.0172
Mix A vs. Placebo	CPTCom	both	F	-2.1759	3.2193	0.5014
Mix A vs. Placebo	CPThit	both	M	-0.02723	1.2124	0.9821
Mix A vs. Placebo	CPThit	both	F	0.02704	1.3084	0.9836
Mix A vs. Placebo	CPTdpr	both	M	-0.02059	0.0458	0.6543
Mix A vs. Placebo	CPTdpr	both	F	0.07202	0.05923	0.2283
Mix A vs. Placebo	CPTbet	both	M	-0.05105	0.1111	0.647
Mix A vs. Placebo	CPTbet	both	F	-0.02995	0.09866	0.7624
Mix B vs. Placebo	Parent	both	both	0.788	0.3265	0.0163
Mix B vs. Placebo	Teacher	both	both	0.3655	0.2968	0.219
Mix B vs. Placebo	Observer	both	both	0.4436	1.0511	0.6732
Mix B vs. Placebo	CPTCom	both	both	2.968	1.8371	0.1082
Mix B vs. Placebo	CPThit	both	both	0.8544	0.8612	0.3227
Mix B vs. Placebo	CPTdpr	both	both	-0.03139	0.03579	0.3818
Mix B vs. Placebo	CPTbet	both	both	0.08324	0.07202	0.2496

Test	score	Year Group	sex	Estimate	StdErr	P-Value
Mix B vs. Placebo	Parent	both	M	0.8443	0.4504	0.0622
Mix B vs. Placebo	Parent	both	F	0.5716	0.4702	0.2257
Mix B vs. Placebo	Teacher	both	M	0.4176	0.4535	0.3582
Mix B vs. Placebo	Teacher	both	F	0.3338	0.3672	0.3645
Mix B vs. Placebo	Observer	both	M	0.2666	1.6205	0.8695
Mix B vs. Placebo	Observer	both	F	0.5166	1.2812	0.6873
Mix B vs. Placebo	CPTCom	both	M	6.0375	2.0077	0.0035
Mix B vs. Placebo	CPTCom	both	F	-1.9687	3.277	0.55
Mix B vs. Placebo	CPThit	both	M	0.7441	1.1637	0.5243
Mix B vs. Placebo	CPThit	both	F	1.1053	1.3317	0.4095
Mix B vs. Placebo	CPTdpr	both	M	-0.05848	0.04399	0.1874
Mix B vs. Placebo	CPTdpr	both	F	0.01188	0.06032	0.8444
Mix B vs. Placebo	CPTbet	both	M	0.07259	0.1066	0.498
Mix B vs. Placebo	CPTbet	both	F	0.08953	0.1002	0.375
Week 2 vs. Week 6	Parent	8/9Y	M	1.431	0.6387	0.0271
Week 2 vs. Week 6	Parent	8/9Y	F	1.0198	0.6974	0.1475
Week 2 vs. Week 6	Teacher	8/9Y	M	-0.2402	0.643	0.7095
Week 2 vs. Week 6	Teacher	8/9Y	F	-0.1291	0.4051	0.7508
Week 2 vs. Week 6	Observer	8/9Y	M	-2.1742	2.6281	0.4099
Week 2 vs. Week 6	Observer	8/9Y	F	1.1499	2.0763	0.5812
Week 2 vs. Week 6	CPTCom	8/9Y	M	0.2983	2.0851	0.8866
Week 2 vs. Week 6	CPTCom	8/9Y	F	-4.3194	3.1731	0.178
Week 2 vs. Week 6	CPThit	8/9Y	M	-3.7977	1.2083	0.0023
Week 2 vs. Week 6	CPThit	8/9Y	F	-6.0042	1.2897	<.0001
Week 2 vs. Week 6	CPTdpr	8/9Y	M	-0.00976	0.04565	0.8312
Week 2 vs. Week 6	CPTdpr	8/9Y	F	0.1167	0.05837	0.0497
Week 2 vs. Week 6	CPTbet	8/9Y	M	-0.0091	0.1107	0.9346
Week 2 vs. Week 6	CPTbet	8/9Y	F	-0.01188	0.09733	0.9032
Week 2 vs. Week 6	Parent	8/9Y	both	1.2224	0.4631	0.009
Week 2 vs. Week 6	Teacher	8/9Y	both	-0.1919	0.3977	0.63
Week 2 vs. Week 6	Observer	8/9Y	both	-0.7486	1.7106	0.6621
Week 2 vs. Week 6	CPTCom	8/9Y	both	-1.3115	1.8528	0.4801
Week 2 vs. Week 6	CPThit	8/9Y	both	-4.7883	0.8685	<.0001
Week 2 vs. Week 6	CPTdpr	8/9Y	both	0.04128	0.0361	0.2547
Week 2 vs. Week 6	CPTbet	8/9Y	both	-0.0141	0.07265	0.8464
Week 4 vs. Week 6	Parent	8/9Y	M	1.4099	0.6325	0.0278
Week 4 vs. Week 6	Parent	8/9Y	F	0.4541	0.7042	0.5208
Week 4 vs. Week 6	Teacher	8/9Y	M	0.77	0.6363	0.2288
Week 4 vs. Week 6	Teacher	8/9Y	F	0.08739	0.408	0.8309
Week 4 vs. Week 6	Observer	8/9Y	M	3.9307	2.6009	0.1336
Week 4 vs. Week 6	Observer	8/9Y	F	4.3787	2.0994	0.0401
Week 4 vs. Week 6	CPTCom	8/9Y	M	1.5729	2.0585	0.447
Week 4 vs. Week 6	CPTCom	8/9Y	F	-4.4719	3.2145	0.1688
Week 4 vs. Week 6	CPThit	8/9Y	M	-2.258	1.1929	0.0619
Week 4 vs. Week 6	CPThit	8/9Y	F	-1.2636	1.3063	0.3369
Week 4 vs. Week 6	CPTdpr	8/9Y	M	-0.00439	0.04507	0.9226

Test	score	Year Group	sex	Estimate	StdErr	P-Value
Week 4 vs. Week 6	CPTdpr	8/9Y	F	0.03805	0.05916	0.5223
Week 4 vs. Week 6	CPTbet	8/9Y	M	0.0524	0.1093	0.6328
Week 4 vs. Week 6	CPTbet	8/9Y	F	0.02528	0.09835	0.7979
Week 4 vs. Week 6	Parent	8/9Y	both	1.0068	0.4609	0.0301
Week 4 vs. Week 6	Teacher	8/9Y	both	0.5274	0.3948	0.1832
Week 4 vs. Week 6	Observer	8/9Y	both	4.5115	1.7014	0.0087
Week 4 vs. Week 6	CPTCom	8/9Y	both	-0.6818	1.8375	0.7111
Week 4 vs. Week 6	CPThit	8/9Y	both	-1.8608	0.8614	0.0323
Week 4 vs. Week 6	CPTdpr	8/9Y	both	0.01308	0.03579	0.7152
Week 4 vs. Week 6	CPTbet	8/9Y	both	0.03677	0.07203	0.6105
Mix A vs. Placebo	Parent	8/9Y	M	0.1331	0.6362	0.8346
Mix A vs. Placebo	Parent	8/9Y	F	0.3337	0.7187	0.6436
Mix A vs. Placebo	Teacher	8/9Y	M	0.04486	0.6454	0.9447
Mix A vs. Placebo	Teacher	8/9Y	F	-0.03036	0.4141	0.9417
Mix A vs. Placebo	Observer	8/9Y	M	0.531	2.6376	0.8408
Mix A vs. Placebo	Observer	8/9Y	F	2.4603	2.1333	0.2522
Mix A vs. Placebo	CPTCom	8/9Y	M	5.0867	2.0922	0.0172
Mix A vs. Placebo	CPTCom	8/9Y	F	-2.1759	3.2193	0.5014
Mix A vs. Placebo	CPThit	8/9Y	M	-0.02723	1.2124	0.9821
Mix A vs. Placebo	CPThit	8/9Y	F	0.02704	1.3084	0.9836
Mix A vs. Placebo	CPTdpr	8/9Y	M	-0.02059	0.0458	0.6543
Mix A vs. Placebo	CPTdpr	8/9Y	F	0.07202	0.05923	0.2283
Mix A vs. Placebo	CPTbet	8/9Y	M	-0.05105	0.1111	0.647
Mix A vs. Placebo	CPTbet	8/9Y	F	-0.02995	0.09866	0.7624
Mix A vs. Placebo	Parent	8/9Y	both	0.2044	0.4665	0.6618
Mix A vs. Placebo	Teacher	8/9Y	both	-0.0471	0.4008	0.9066
Mix A vs. Placebo	Observer	8/9Y	both	1.0664	1.7279	0.5378
Mix A vs. Placebo	CPTCom	8/9Y	both	2.0004	1.8664	0.2855
Mix A vs. Placebo	CPThit	8/9Y	both	0.2699	0.8749	0.7581
Mix A vs. Placebo	CPTdpr	8/9Y	both	0.01144	0.03635	0.7533
Mix A vs. Placebo	CPTbet	8/9Y	both	-0.03885	0.07316	0.5962
Mix B vs. Placebo	Parent	8/9Y	M	1.0254	0.6214	0.1018
Mix B vs. Placebo	Parent	8/9Y	F	0.7256	0.703	0.305
Mix B vs. Placebo	Teacher	8/9Y	M	0.4307	0.6256	0.4926
Mix B vs. Placebo	Teacher	8/9Y	F	-0.1026	0.4108	0.8035
Mix B vs. Placebo	Observer	8/9Y	M	2.4231	2.5576	0.3455
Mix B vs. Placebo	Observer	8/9Y	F	0.2494	2.1019	0.9059
Mix B vs. Placebo	CPTCom	8/9Y	M	6.0375	2.0077	0.0035
Mix B vs. Placebo	CPTCom	8/9Y	F	-1.9687	3.277	0.55
Mix B vs. Placebo	CPThit	8/9Y	M	0.7441	1.1637	0.5243
Mix B vs. Placebo	CPThit	8/9Y	F	1.1053	1.3317	0.4095
Mix B vs. Placebo	CPTdpr	8/9Y	M	-0.05848	0.04399	0.1874
Mix B vs. Placebo	CPTdpr	8/9Y	F	0.01188	0.06032	0.8444
Mix B vs. Placebo	CPTbet	8/9Y	M	0.07259	0.1066	0.498
Mix B vs. Placebo	CPTbet	8/9Y	F	0.08953	0.1002	0.375
Mix B vs. Placebo	Parent	8/9Y	both	0.9687	0.4584	0.0359

Test	score	Year Group	sex	Estimate	StdErr	P-Value
Mix B vs. Placebo	Teacher	8/9Y	both	0.234	0.3938	0.5531
Mix B vs. Placebo	Observer	8/9Y	both	1.3445	1.6937	0.4283
Mix B vs. Placebo	CPTCom	8/9Y	both	2.968	1.8371	0.1082
Mix B vs. Placebo	CPThit	8/9Y	both	0.8544	0.8612	0.3227
Mix B vs. Placebo	CPTdpr	8/9Y	both	-0.03139	0.03579	0.3818
Mix B vs. Placebo	CPTbet	8/9Y	both	0.08324	0.07202	0.2496

For Completers:

Test	score	Year Group	sex	Estimate	StdErr	P-Value
Week 2 vs. Week 6	Parent	3Y	M	1.1404	0.8432	0.1806
Week 2 vs. Week 6	Parent	3Y	F	-0.5299	0.7304	0.4706
Week 2 vs. Week 6	Teacher	3Y	M	-0.85	0.6481	0.194
Week 2 vs. Week 6	Teacher	3Y	F	0.4584	0.7511	0.5438
Week 2 vs. Week 6	Observer	3Y	M	5.9341	2.0085	0.0043
Week 2 vs. Week 6	Observer	3Y	F	3.1256	1.827	0.0917
Week 2 vs. Week 6	Parent	3Y	both	0.3277	0.5546	0.5556
Week 2 vs. Week 6	Teacher	3Y	both	-0.2005	0.4906	0.6834
Week 2 vs. Week 6	Observer	3Y	both	4.5922	1.341	0.0008
Week 4 vs. Week 6	Parent	3Y	M	1.2124	0.8413	0.154
Week 4 vs. Week 6	Parent	3Y	F	-0.8746	0.7513	0.2484
Week 4 vs. Week 6	Teacher	3Y	M	-0.5659	0.6467	0.3845
Week 4 vs. Week 6	Teacher	3Y	F	1.0027	0.7584	0.1907
Week 4 vs. Week 6	Observer	3Y	M	5.6344	2.0041	0.0064
Week 4 vs. Week 6	Observer	3Y	F	2.2058	1.8793	0.2446
Week 4 vs. Week 6	Parent	3Y	both	0.1981	0.5592	0.7236
Week 4 vs. Week 6	Teacher	3Y	both	0.1472	0.4896	0.7642
Week 4 vs. Week 6	Observer	3Y	both	4.0797	1.3521	0.003
Mix A vs. Placebo	Parent	3Y	M	1.1678	0.8319	0.1648
Mix A vs. Placebo	Parent	3Y	F	1.1726	0.762	0.1285
Mix A vs. Placebo	Teacher	3Y	M	0.2437	0.6394	0.7043
Mix A vs. Placebo	Teacher	3Y	F	1.0946	0.7745	0.1623
Mix A vs. Placebo	Observer	3Y	M	1.5451	1.9817	0.4382
Mix A vs. Placebo	Observer	3Y	F	0.4053	1.9061	0.8323
Mix A vs. Placebo	Parent	3Y	both	1.2448	0.5585	0.0274
Mix A vs. Placebo	Teacher	3Y	both	0.594	0.491	0.2284
Mix A vs. Placebo	Observer	3Y	both	1.1165	1.3502	0.4096
Mix B vs. Placebo	Parent	3Y	M	0.7659	0.8451	0.3679
Mix B vs. Placebo	Parent	3Y	F	1.0389	0.7386	0.1641
Mix B vs. Placebo	Teacher	3Y	M	0.542	0.6495	0.4069
Mix B vs. Placebo	Teacher	3Y	F	0.7543	0.7506	0.3186
Mix B vs. Placebo	Observer	3Y	M	-0.4845	2.0129	0.8105

Test	score	Year Group	sex	Estimate	StdErr	P-Value
Mix B vs. Placebo	Observer	3Y	F	0.2391	1.8477	0.8974
Mix B vs. Placebo	Parent	3Y	both	0.8326	0.5578	0.1377
Mix B vs. Placebo	Teacher	3Y	both	0.6967	0.4907	0.1579
Mix B vs. Placebo	Observer	3Y	both	-0.2299	1.3486	0.8649
Week 2 vs. Week 6	Parent	both	both	0.8506	0.3599	0.0187
Week 2 vs. Week 6	Teacher	both	both	-0.2888	0.3171	0.363
Week 2 vs. Week 6	Observer	both	both	1.6691	1.1961	0.1639
Week 2 vs. Week 6	CPTCom	both	both	-1.6701	2.03	0.4121
Week 2 vs. Week 6	CPThit	both	both	-5.1107	0.9365	<.0001
Week 2 vs. Week 6	CPTdpr	both	both	0.04813	0.03933	0.2232
Week 2 vs. Week 6	CPTbet	both	both	0.001545	0.07927	0.9845
Week 2 vs. Week 6	Parent	both	M	1.3042	0.4921	0.0088
Week 2 vs. Week 6	Parent	both	F	0.3139	0.5282	0.5533
Week 2 vs. Week 6	Teacher	both	M	-0.5401	0.4702	0.2523
Week 2 vs. Week 6	Teacher	both	F	0.05855	0.4134	0.8876
Week 2 vs. Week 6	Observer	both	M	1.478	1.8315	0.4208
Week 2 vs. Week 6	Observer	both	F	2.0369	1.4613	0.1655
Week 2 vs. Week 6	CPTCom	both	M	-0.1524	2.2543	0.9463
Week 2 vs. Week 6	CPTCom	both	F	-4.7975	3.5952	0.1872
Week 2 vs. Week 6	CPThit	both	M	-3.7591	1.2661	0.004
Week 2 vs. Week 6	CPThit	both	F	-6.7067	1.4192	<.0001
Week 2 vs. Week 6	CPTdpr	both	M	-0.01329	0.0491	0.7874
Week 2 vs. Week 6	CPTdpr	both	F	0.1339	0.06497	0.0437
Week 2 vs. Week 6	CPTbet	both	M	-0.00987	0.1209	0.9352
Week 2 vs. Week 6	CPTbet	both	F	0.03165	0.104	0.762
Week 4 vs. Week 6	Parent	both	both	0.5197	0.3611	0.1511
Week 4 vs. Week 6	Teacher	both	both	0.2956	0.3169	0.3517
Week 4 vs. Week 6	Observer	both	both	4.2824	1.2004	0.0004
Week 4 vs. Week 6	CPTCom	both	both	-0.2803	2.0391	0.8909
Week 4 vs. Week 6	CPThit	both	both	-1.9948	0.9407	0.0358
Week 4 vs. Week 6	CPTdpr	both	both	-0.00136	0.03951	0.9727
Week 4 vs. Week 6	CPTbet	both	both	0.08092	0.07962	0.3113
Week 4 vs. Week 6	Parent	both	M	1.2231	0.4911	0.0137
Week 4 vs. Week 6	Parent	both	F	-0.3261	0.5356	0.5435
Week 4 vs. Week 6	Teacher	both	M	0.2016	0.4692	0.6679
Week 4 vs. Week 6	Teacher	both	F	0.422	0.415	0.311
Week 4 vs. Week 6	Observer	both	M	4.9287	1.8274	0.0077
Week 4 vs. Week 6	Observer	both	F	3.4398	1.4818	0.0217
Week 4 vs. Week 6	CPTCom	both	M	1.3531	2.2428	0.5481
Week 4 vs. Week 6	CPTCom	both	F	-3.9943	3.7334	0.289
Week 4 vs. Week 6	CPThit	both	M	-2.5152	1.2596	0.0495
Week 4 vs. Week 6	CPThit	both	F	-1.1375	1.4733	0.4431
Week 4 vs. Week 6	CPTdpr	both	M	-0.0159	0.04885	0.7457
Week 4 vs. Week 6	CPTdpr	both	F	0.02438	0.06747	0.7192
Week 4 vs. Week 6	CPTbet	both	M	0.06789	0.1203	0.5743
Week 4 vs. Week 6	CPTbet	both	F	0.1002	0.108	0.357

Test	score	Year Group	sex	Estimate	StdErr	P-Value
Mix A vs. Placebo	Parent	both	both	0.6646	0.3603	0.066
Mix A vs. Placebo	Teacher	both	both	0.3201	0.3173	0.3138
Mix A vs. Placebo	Observer	both	both	0.7676	1.1998	0.5228
Mix A vs. Placebo	CPTCom	both	both	2.5671	2.0325	0.2087
Mix A vs. Placebo	CPThit	both	both	0.8879	0.9376	0.3453
Mix A vs. Placebo	CPTdpr	both	both	0.006658	0.03938	0.866
Mix A vs. Placebo	CPTbet	both	both	-0.02996	0.07937	0.7064
Mix A vs. Placebo	Parent	both	M	0.5854	0.4907	0.2345
Mix A vs. Placebo	Parent	both	F	0.6451	0.5319	0.2272
Mix A vs. Placebo	Teacher	both	M	0.1481	0.4706	0.7533
Mix A vs. Placebo	Teacher	both	F	0.5735	0.4141	0.1682
Mix A vs. Placebo	Observer	both	M	0.08688	1.8329	0.9622
Mix A vs. Placebo	Observer	both	F	1.5459	1.4716	0.2952
Mix A vs. Placebo	CPTCom	both	M	5.5423	2.2612	0.0166
Mix A vs. Placebo	CPTCom	both	F	-1.7885	3.5952	0.6207
Mix A vs. Placebo	CPThit	both	M	0.4741	1.2699	0.71
Mix A vs. Placebo	CPThit	both	F	0.7744	1.4192	0.5874
Mix A vs. Placebo	CPTdpr	both	M	-0.01048	0.04925	0.832
Mix A vs. Placebo	CPTdpr	both	F	0.05453	0.06497	0.4046
Mix A vs. Placebo	CPTbet	both	M	-0.07036	0.1213	0.5636
Mix A vs. Placebo	CPTbet	both	F	0.02522	0.104	0.8093
Mix B vs. Placebo	Parent	both	both	0.7726	0.3611	0.0332
Mix B vs. Placebo	Teacher	both	both	0.455	0.3175	0.1528
Mix B vs. Placebo	Observer	both	both	0.9208	1.2004	0.4436
Mix B vs. Placebo	CPTCom	both	both	3.1417	2.0452	0.1268
Mix B vs. Placebo	CPThit	both	both	1.6799	0.9434	0.0772
Mix B vs. Placebo	CPTdpr	both	both	-0.03562	0.03963	0.3703
Mix B vs. Placebo	CPTbet	both	both	0.08221	0.07986	0.3051
Mix B vs. Placebo	Parent	both	M	0.6603	0.4911	0.1805
Mix B vs. Placebo	Parent	both	F	0.764	0.5356	0.1559
Mix B vs. Placebo	Teacher	both	M	0.4749	0.4692	0.3128
Mix B vs. Placebo	Teacher	both	F	0.4363	0.4168	0.297
Mix B vs. Placebo	Observer	both	M	1.3772	1.8274	0.4521
Mix B vs. Placebo	Observer	both	F	0.1812	1.4818	0.9028
Mix B vs. Placebo	CPTCom	both	M	6.3442	2.2428	0.006
Mix B vs. Placebo	CPTCom	both	F	-2.0092	3.7482	0.5939
Mix B vs. Placebo	CPThit	both	M	1.5185	1.2596	0.2318
Mix B vs. Placebo	CPThit	both	F	2.2789	1.4791	0.1287
Mix B vs. Placebo	CPTdpr	both	M	-0.05424	0.04885	0.2705
Mix B vs. Placebo	CPTdpr	both	F	-0.00642	0.06774	0.9249
Mix B vs. Placebo	CPTbet	both	M	0.07258	0.1203	0.5482
Mix B vs. Placebo	CPTbet	both	F	0.1013	0.1084	0.3538
Week 2 vs. Week 6	Parent	8/9Y	M	1.4906	0.6176	0.0176
Week 2 vs. Week 6	Parent	8/9Y	F	1.0395	0.772	0.1824
Week 2 vs. Week 6	Teacher	8/9Y	M	-0.2976	0.6746	0.66
Week 2 vs. Week 6	Teacher	8/9Y	F	-0.3993	0.4014	0.3232

Test	score	Year Group	sex	Estimate	StdErr	P-Value
Week 2 vs. Week 6	Observer	8/9Y	M	-2.0528	2.7715	0.4606
Week 2 vs. Week 6	Observer	8/9Y	F	1.0263	2.3048	0.6575
Week 2 vs. Week 6	CPTCom	8/9Y	M	-0.1524	2.2543	0.9463
Week 2 vs. Week 6	CPTCom	8/9Y	F	-4.7975	3.5952	0.1872
Week 2 vs. Week 6	CPThit	8/9Y	M	-3.7591	1.2661	0.004
Week 2 vs. Week 6	CPThit	8/9Y	F	-6.7067	1.4192	<.0001
Week 2 vs. Week 6	CPTdpr	8/9Y	M	-0.01329	0.0491	0.7874
Week 2 vs. Week 6	CPTdpr	8/9Y	F	0.1339	0.06497	0.0437
Week 2 vs. Week 6	CPTbet	8/9Y	M	-0.00987	0.1209	0.9352
Week 2 vs. Week 6	CPTbet	8/9Y	F	0.03165	0.104	0.762
Week 2 vs. Week 6	Parent	8/9Y	both	1.2811	0.4723	0.0073
Week 2 vs. Week 6	Teacher	8/9Y	both	-0.3365	0.419	0.423
Week 2 vs. Week 6	Observer	8/9Y	both	-0.7631	1.8502	0.6805
Week 2 vs. Week 6	CPTCom	8/9Y	both	-1.6701	2.03	0.4121
Week 2 vs. Week 6	CPThit	8/9Y	both	-5.1107	0.9365	<.0001
Week 2 vs. Week 6	CPTdpr	8/9Y	both	0.04813	0.03933	0.2232
Week 2 vs. Week 6	CPTbet	8/9Y	both	0.001545	0.07927	0.9845
Week 4 vs. Week 6	Parent	8/9Y	M	1.2012	0.6138	0.0531
Week 4 vs. Week 6	Parent	8/9Y	F	0.332	0.7874	0.6745
Week 4 vs. Week 6	Teacher	8/9Y	M	0.7588	0.6698	0.2599
Week 4 vs. Week 6	Teacher	8/9Y	F	-0.07616	0.4094	0.853
Week 4 vs. Week 6	Observer	8/9Y	M	3.8037	2.7518	0.1699
Week 4 vs. Week 6	Observer	8/9Y	F	4.2515	2.3508	0.0747
Week 4 vs. Week 6	CPTCom	8/9Y	M	1.3531	2.2428	0.5481
Week 4 vs. Week 6	CPTCom	8/9Y	F	-3.9943	3.7334	0.289
Week 4 vs. Week 6	CPThit	8/9Y	M	-2.5152	1.2596	0.0495
Week 4 vs. Week 6	CPThit	8/9Y	F	-1.1375	1.4733	0.4431
Week 4 vs. Week 6	CPTdpr	8/9Y	M	-0.0159	0.04885	0.7457
Week 4 vs. Week 6	CPTdpr	8/9Y	F	0.02438	0.06747	0.7192
Week 4 vs. Week 6	CPTbet	8/9Y	M	0.06789	0.1203	0.5743
Week 4 vs. Week 6	CPTbet	8/9Y	F	0.1002	0.108	0.357
Week 4 vs. Week 6	Parent	8/9Y	both	0.8224	0.4724	0.0834
Week 4 vs. Week 6	Teacher	8/9Y	both	0.4402	0.4192	0.2951
Week 4 vs. Week 6	Observer	8/9Y	both	4.4394	1.8512	0.0175
Week 4 vs. Week 6	CPTCom	8/9Y	both	-0.2803	2.0391	0.8909
Week 4 vs. Week 6	CPThit	8/9Y	both	-1.9948	0.9407	0.0358
Week 4 vs. Week 6	CPTdpr	8/9Y	both	-0.00136	0.03951	0.9727
Week 4 vs. Week 6	CPTbet	8/9Y	both	0.08092	0.07962	0.3113
Mix A vs. Placebo	Parent	8/9Y	M	0.1285	0.6145	0.8348
Mix A vs. Placebo	Parent	8/9Y	F	0.2266	0.772	0.7699
Mix A vs. Placebo	Teacher	8/9Y	M	0.09107	0.6764	0.8932
Mix A vs. Placebo	Teacher	8/9Y	F	0.141	0.4014	0.7265
Mix A vs. Placebo	Observer	8/9Y	M	-0.08674	2.779	0.9752
Mix A vs. Placebo	Observer	8/9Y	F	1.8074	2.3048	0.4355
Mix A vs. Placebo	CPTCom	8/9Y	M	5.5423	2.2612	0.0166
Mix A vs. Placebo	CPTCom	8/9Y	F	-1.7885	3.5952	0.6207

Test	score	Year Group	sex	Estimate	StdErr	P-Value
Mix A vs. Placebo	CPThit	8/9Y	M	0.4741	1.2699	0.71
Mix A vs. Placebo	CPThit	8/9Y	F	0.7744	1.4192	0.5874
Mix A vs. Placebo	CPTdpr	8/9Y	M	-0.01048	0.04925	0.832
Mix A vs. Placebo	CPTdpr	8/9Y	F	0.05453	0.06497	0.4046
Mix A vs. Placebo	CPTbet	8/9Y	M	-0.07036	0.1213	0.5636
Mix A vs. Placebo	CPTbet	8/9Y	F	0.02522	0.104	0.8093
Mix A vs. Placebo	Parent	8/9Y	both	0.1776	0.4707	0.7064
Mix A vs. Placebo	Teacher	8/9Y	both	0.08908	0.4192	0.8319
Mix A vs. Placebo	Observer	8/9Y	both	0.4485	1.8512	0.8089
Mix A vs. Placebo	CPTCom	8/9Y	both	2.5671	2.0325	0.2087
Mix A vs. Placebo	CPThit	8/9Y	both	0.8879	0.9376	0.3453
Mix A vs. Placebo	CPTdpr	8/9Y	both	0.006658	0.03938	0.866
Mix A vs. Placebo	CPTbet	8/9Y	both	-0.02996	0.07937	0.7064
Mix B vs. Placebo	Parent	8/9Y	M	0.5417	0.6156	0.381
Mix B vs. Placebo	Parent	8/9Y	F	0.6839	0.784	0.3859
Mix B vs. Placebo	Teacher	8/9Y	M	0.3069	0.6716	0.6486
Mix B vs. Placebo	Teacher	8/9Y	F	0.06091	0.4077	0.8816
Mix B vs. Placebo	Observer	8/9Y	M	3.5794	2.7594	0.1975
Mix B vs. Placebo	Observer	8/9Y	F	-0.01938	2.3406	0.9934
Mix B vs. Placebo	CPTCom	8/9Y	M	6.3442	2.2428	0.006
Mix B vs. Placebo	CPTCom	8/9Y	F	-2.0092	3.7482	0.5939
Mix B vs. Placebo	CPThit	8/9Y	M	1.5185	1.2596	0.2318
Mix B vs. Placebo	CPThit	8/9Y	F	2.2789	1.4791	0.1287
Mix B vs. Placebo	CPTdpr	8/9Y	M	-0.05424	0.04885	0.2705
Mix B vs. Placebo	CPTdpr	8/9Y	F	-0.00642	0.06774	0.9249
Mix B vs. Placebo	CPTbet	8/9Y	M	0.07258	0.1203	0.5482
Mix B vs. Placebo	CPTbet	8/9Y	F	0.1013	0.1084	0.3538
Mix B vs. Placebo	Parent	8/9Y	both	0.6808	0.4733	0.152
Mix B vs. Placebo	Teacher	8/9Y	both	0.2648	0.4199	0.529
Mix B vs. Placebo	Observer	8/9Y	both	1.9962	1.8542	0.2831
Mix B vs. Placebo	CPTCom	8/9Y	both	3.1417	2.0452	0.1268
Mix B vs. Placebo	CPThit	8/9Y	both	1.6799	0.9434	0.0772
Mix B vs. Placebo	CPTdpr	8/9Y	both	-0.03562	0.03963	0.3703
Mix B vs. Placebo	CPTbet	8/9Y	both	0.08221	0.07986	0.3051

For the newly calculated aggregate score:

Test	Consumption Group	Year Group	sex	Estimate	StdErr	P-Value
Week 2 vs. Week 6	≥85% consumption	3Y	F	0.01474	0.1057	0.8894
Week 2 vs. Week 6	≥85% consumption	3Y	M	0.1908	0.1254	0.1317
Week 2 vs. Week 6	≥85% consumption	3Y	both	0.09483	0.08146	0.2459
Week 2 vs. Week 6	Entire sample	3Y	F	-0.01245	0.0895	0.8896
Week 2 vs. Week 6	Entire sample	3Y	M	0.1787	0.1031	0.0855
Week 2 vs. Week 6	Entire sample	3Y	both	0.08466	0.06833	0.2165
Week 2 vs. Week 6	Complete case	3Y	F	0.004043	0.1302	0.9753
Week 2 vs. Week 6	Complete case	3Y	M	0.2689	0.1461	0.07
Week 2 vs. Week 6	Complete case	3Y	both	0.1411	0.09697	0.1478
Week 4 vs. Week 6	≥85% consumption	3Y	F	0.001271	0.1045	0.9903
Week 4 vs. Week 6	≥85% consumption	3Y	M	0.3267	0.1274	0.012
Week 4 vs. Week 6	≥85% consumption	3Y	both	0.1621	0.08156	0.0483
Week 4 vs. Week 6	Entire sample	3Y	F	-0.01976	0.09072	0.8279
Week 4 vs. Week 6	Entire sample	3Y	M	0.2467	0.1057	0.0212
Week 4 vs. Week 6	Entire sample	3Y	both	0.1119	0.06949	0.1085
Week 4 vs. Week 6	Complete case	3Y	F	-0.02721	0.1339	0.8395
Week 4 vs. Week 6	Complete case	3Y	M	0.2863	0.1458	0.0536
Week 4 vs. Week 6	Complete case	3Y	both	0.1342	0.09777	0.1722
Mix A vs. Placebo	≥85% consumption	3Y	F	0.1909	0.1081	0.0807
Mix A vs. Placebo	≥85% consumption	3Y	M	0.1735	0.1215	0.1567
Mix A vs. Placebo	≥85% consumption	3Y	both	0.1962	0.08074	0.0161
Mix A vs. Placebo	Entire sample	3Y	F	0.162	0.08979	0.0737
Mix A vs. Placebo	Entire sample	3Y	M	0.06805	0.1043	0.5154
Mix A vs. Placebo	Entire sample	3Y	both	0.1196	0.06849	0.0819
Mix A vs. Placebo	Complete case	3Y	F	0.2306	0.1358	0.0941
Mix A vs. Placebo	Complete case	3Y	M	0.221	0.1442	0.1298
Mix A vs. Placebo	Complete case	3Y	both	0.2359	0.09764	0.0169
Mix B vs. Placebo	≥85% consumption	3Y	F	0.1923	0.1045	0.0691
Mix B vs. Placebo	≥85%	3Y	M	0.07238	0.1222	0.5551

Test	Consumption Group	Year Group	sex	Estimate	StdErr	P-Value
Mix B vs. Placebo	≥85% consumption	3Y	both	0.1302	0.08017	0.106
Mix B vs. Placebo	Entire sample	3Y	F	0.1422	0.08984	0.116
Mix B vs. Placebo	Entire sample	3Y	M	0.04744	0.1021	0.6431
Mix B vs. Placebo	Entire sample	3Y	both	0.08825	0.06821	0.1969
Mix B vs. Placebo	Complete case	3Y	F	0.1896	0.1316	0.1542
Mix B vs. Placebo	Complete case	3Y	M	0.1276	0.1464	0.3864
Mix B vs. Placebo	Complete case	3Y	both	0.1471	0.09752	0.1337
Week 2 vs. Week 6	≥85% consumption	both	F	0.02043	0.05923	0.7305
Week 2 vs. Week 6	≥85% consumption	both	M	0.06742	0.06179	0.2765
Week 2 vs. Week 6	≥85% consumption	both	both	0.03971	0.04304	0.3568
Week 2 vs. Week 6	Entire sample	both	F	0.02007	0.05024	0.6898
Week 2 vs. Week 6	Entire sample	both	M	0.0882	0.05591	0.1159
Week 2 vs. Week 6	Entire sample	both	both	0.05364	0.03771	0.1555
Week 2 vs. Week 6	Complete case	both	F	0.00913	0.06789	0.8932
Week 2 vs. Week 6	Complete case	both	M	0.09283	0.06652	0.1646
Week 2 vs. Week 6	Complete case	both	both	0.05338	0.04754	0.2624
Week 4 vs. Week 6	≥85% consumption	both	F	0.02397	0.05872	0.6837
Week 4 vs. Week 6	≥85% consumption	both	M	0.2121	0.062	0.0008
Week 4 vs. Week 6	≥85% consumption	both	both	0.1225	0.04289	0.0045
Week 4 vs. Week 6	Entire sample	both	F	0.01648	0.05043	0.7441
Week 4 vs. Week 6	Entire sample	both	M	0.1824	0.05636	0.0014
Week 4 vs. Week 6	Entire sample	both	both	0.1016	0.03795	0.0077
Week 4 vs. Week 6	Complete case	both	F	0.00464	0.06884	0.9464
Week 4 vs. Week 6	Complete case	both	M	0.1812	0.06638	0.007
Week 4 vs. Week 6	Complete case	both	both	0.101	0.04771	0.0351
Mix A vs. Placebo	≥85% consumption	both	F	0.1312	0.05972	0.0294
Mix A vs. Placebo	≥85% consumption	both	M	0.08458	0.06122	0.1686
Mix A vs. Placebo	≥85% consumption	both	both	0.1096	0.04304	0.0113
Mix A vs. Placebo	Entire sample	both	F	0.09706	0.04992	0.053
Mix A vs. Placebo	Entire sample	both	M	0.04395	0.05622	0.435

Test	Consumption Group	Year Group	sex	Estimate	StdErr	P-Value
Mix A vs. Placebo	Entire sample	both	both	0.07192	0.03775	0.0573
Mix A vs. Placebo	Complete case	both	F	0.1443	0.06836	0.0365
Mix A vs. Placebo	Complete case	both	M	0.09237	0.06657	0.1671
Mix A vs. Placebo	Complete case	both	both	0.1211	0.04769	0.0115
Mix B vs. Placebo	≥85% consumption	both	F	0.1224	0.05935	0.0407
Mix B vs. Placebo	≥85% consumption	both	M	0.07842	0.0602	0.1941
Mix B vs. Placebo	≥85% consumption	both	both	0.1059	0.04249	0.0131
Mix B vs. Placebo	Entire sample	both	F	0.07269	0.05049	0.1513
Mix B vs. Placebo	Entire sample	both	M	0.06843	0.05552	0.2189
Mix B vs. Placebo	Entire sample	both	both	0.07429	0.03765	0.0491
Mix B vs. Placebo	Complete case	both	F	0.1101	0.06884	0.1119
Mix B vs. Placebo	Complete case	both	M	0.09554	0.06638	0.1518
Mix B vs. Placebo	Complete case	both	both	0.1093	0.04771	0.0226
Week 2 vs. Week 6	≥85% consumption	8/9Y	F	0.01242	0.04554	0.7858
Week 2 vs. Week 6	≥85% consumption	8/9Y	M	-0.02196	0.04772	0.6462
Week 2 vs. Week 6	≥85% consumption	8/9Y	both	-0.00818	0.0331	0.805
Week 2 vs. Week 6	Entire sample	8/9Y	F	0.03855	0.04044	0.3425
Week 2 vs. Week 6	Entire sample	8/9Y	M	0.000805	0.04425	0.9855
Week 2 vs. Week 6	Entire sample	8/9Y	both	0.02446	0.02985	0.4134
Week 2 vs. Week 6	Complete case	8/9Y	F	-0.00656	0.04989	0.8958
Week 2 vs. Week 6	Complete case	8/9Y	M	-0.02186	0.05015	0.6638
Week 2 vs. Week 6	Complete case	8/9Y	both	-0.01467	0.03528	0.6781
Week 4 vs. Week 6	≥85% consumption	8/9Y	F	0.04984	0.04573	0.2791
Week 4 vs. Week 6	≥85% consumption	8/9Y	M	0.1108	0.04722	0.0207
Week 4 vs. Week 6	≥85% consumption	8/9Y	both	0.0938	0.03282	0.0047
Week 4 vs. Week 6	Entire sample	8/9Y	F	0.05203	0.04005	0.1966
Week 4 vs. Week 6	Entire sample	8/9Y	M	0.108	0.04398	0.0154
Week 4 vs. Week 6	Entire sample	8/9Y	both	0.09192	0.02972	0.0022
Week 4 vs. Week 6	Complete case	8/9Y	F	0.04279	0.05088	0.4032
Week 4 vs. Week 6	Complete case	8/9Y	M	0.09762	0.04979	0.0527

Test	Consumption Group	Year Group	sex	Estimate	StdErr	P-Value
Week 4 vs. Week 6	Complete case	8/9Y	both	0.08335	0.0353	0.0193
Mix A vs. Placebo	≥85% consumption	8/9Y	F	0.04658	0.04641	0.3186
Mix A vs. Placebo	≥85% consumption	8/9Y	M	0.02353	0.04789	0.6241
Mix A vs. Placebo	≥85% consumption	8/9Y	both	0.02733	0.03333	0.4132
Mix A vs. Placebo	Entire sample	8/9Y	F	0.02548	0.04016	0.527
Mix A vs. Placebo	Entire sample	8/9Y	M	0.03375	0.04454	0.4499
Mix A vs. Placebo	Entire sample	8/9Y	both	0.02194	0.02982	0.4627
Mix A vs. Placebo	Complete case	8/9Y	F	0.04852	0.04989	0.334
Mix A vs. Placebo	Complete case	8/9Y	M	0.02491	0.05029	0.6214
Mix A vs. Placebo	Complete case	8/9Y	both	0.03144	0.0353	0.3743
Mix B vs. Placebo	≥85% consumption	8/9Y	F	0.03855	0.04615	0.406
Mix B vs. Placebo	≥85% consumption	8/9Y	M	0.1116	0.04644	0.018
Mix B vs. Placebo	≥85% consumption	8/9Y	both	0.08348	0.03276	0.0116
Mix B vs. Placebo	Entire sample	8/9Y	F	0.000137	0.04046	0.9973
Mix B vs. Placebo	Entire sample	8/9Y	M	0.1115	0.04394	0.0124
Mix B vs. Placebo	Entire sample	8/9Y	both	0.05963	0.02981	0.0466
Mix B vs. Placebo	Complete case	8/9Y	F	0.03946	0.05066	0.4386
Mix B vs. Placebo	Complete case	8/9Y	M	0.1118	0.04993	0.0273
Mix B vs. Placebo	Complete case	8/9Y	both	0.08546	0.03536	0.0167