

Meta-Analysis of Attention-Deficit/Hyperactivity Disorder or Attention-Deficit/Hyperactivity Disorder Symptoms, Restriction Diet, and Synthetic Food Color Additives

Joel T. Nigg, Ph.D., Kara Lewis, Ph.D., Tracy Edinger, N.D., Michael Falk, Ph.D.

Objective: The role of diet and of food colors in attention-deficit/hyperactivity disorder (ADHD) or its symptoms warrants updated quantitative meta-analysis, in light of recent divergent policy in Europe and the United States. **Method:** Studies were identified through a literature search using the PubMed, Cochrane Library, and PsycNET databases through February 2011. Twenty-four publications met inclusion criteria for synthetic food colors; 10 additional studies informed analysis of dietary restriction. A random-effects meta-analytic model generated summary effect sizes. **Results:** Restriction diets reduced ADHD symptoms at an effect of $g = 0.29$ (95% CI, 0.07–0.53). For food colors, parent reports yielded an effect size of $g = 0.18$ (95% CI, 0.08–0.24; $p = .0007$), which decreased to 0.12 (95% CI, 0.01–0.23; $p < .05$) after adjustment for possible publication bias. The effect was reliable in studies restricted to food color additives ($g = 0.21$, 95% CI = 0.06–0.36) but did not survive correction for possible publication bias and was not reliable in studies confined to Food and Drug Administration–approved food colors. Teacher/observer reports yielded a nonsignificant effect of 0.07 (95% CI = –0.03 to 0.18; $p = .14$). However, high-quality studies confined to color additives yielded a reliable effect ($g = 0.22$, 95% CI = 0.10–0.41, $p = .030$) that survived correction. In psychometric tests of attention, the summary effect size was 0.27 (95% CI = 0.07–0.47; $p = .007$) and survived correction. An estimated 8% of children with ADHD may have symptoms related to synthetic food colors. **Conclusions:** A restriction diet benefits some children with ADHD. Effects of food colors were notable were but susceptible to publication bias or were derived from small, nongeneralizable samples. Renewed investigation of diet and ADHD is warranted. *J. Am. Acad. Child Adolesc. Psychiatry*, 2012;51(1):86–97. **Key words:** ADHD, meta-analysis, synthetic food color additives, restriction diet

Attention-deficit/hyperactivity disorder (ADHD) is a common developmental syndrome that confers elevated risk of school failure, peer rejection, family conflict, substance use disorders, delinquency, underemployment, depression, accidental death, suicide, and physical health problems.¹ Quantitative genetic data suggest that ADHD exists on a spectrum in the population, such that the disorder constitutes a clinical cut point on a dimension;² thus, studies

of symptom variation as well as of the disorder *per se* are of interest and were included. The causes of ADHD and the modifiers of its course are multifactorial. Twin and adoption studies have converged on a heritability of liability for ADHD symptoms of 65% to 75%, but some of the genetic effect is likely due to gene-by-environment interaction.^{3,4} It is suggested that, for some susceptible children, an environmental exposure may influence expression of ADHD.⁵

Numerous environmental factors are suspected to influence ADHD, including prenatal and postnatal toxicant exposures, teratogens, perinatal events, low birth weight, and postnatal



Supplemental material cited in this article is available online.

environmental conflict and stress.⁶ Various dietary effects have been of long-standing interest.

Among the dietary theories, the hypothesis that allergies or else hypersensitivity to certain foods or ingredients cause learning and behavior problems entered the literature as early as the 1920s.⁷ A specific hypothesis that food additives, which include synthetic food colorings and flavors, influence ADHD (at that time, hyperkinetic reaction), via either allergenic or pharmacologic mechanisms, was introduced in the 1970s by Feingold.⁸ He suggested initially that children who are allergic to aspirin are susceptible to synthetic food colors as well as naturally occurring salicylates, but he later focused on food color additives. To treat this reaction, Feingold proposed a diet free of foods with a natural salicylate radical and all synthetic colors and flavors.⁸

The topic of synthetic color additives and hyperactivity was heavily studied in the 1970s and 1980s. In 1982, the National Institutes of Health convened a consensus development conference on defined diets and childhood hyperactivity, which recommended further study. A 1983⁹ meta-analysis included 23 studies regarding the efficacy of the Feingold diet; the authors concluded that the composite effect size ($d = 0.11$) was too small to be important, setting the tone for two decades of professional skepticism as to the value of dietary intervention in ADHD. In a more recent meta-analysis, Schab and Trinh¹⁰ reviewed 15 double-blind, placebo-controlled studies, plus six others for their supplemental analysis. They concluded that there was a reliable effect ($d = 0.28$) linking synthetic colors to ADHD symptoms in parent ratings, but not in teacher or observer ratings, and that the effect was carried by individuals preselected to be diet responsive. The effects seemed to be similar whether or not children were initially selected to be hyperactive. That report helped revive scientific interest in the role of synthetic food colors.

Several considerations warrant an updated meta-analysis at the present time. Subsequent studies have appeared, and significant differences on the risks of food colors have emerged among authorities. In particular, the authors of a population-based study conducted in England¹¹ concluded that food additives contribute to hyperactivity, prompting the European Union Parliament recently to require warning labels on foods containing six colors, not all of which are approved for use in the United States. (The U.S.

Food and Drug Administration [FDA]¹² has approved nine synthetic colors for use in food subject to batch certification: FD&C Blue #1 (brilliant blue), FD&C Blue #2 (Indigotine), FD&C Green #3 (Green S; fast green), Orange B, Citrus Red #2 (Amaranth), FD&C Red #3 (Erythrosine), FD&C Red #40 (Allura Red), FD&C Yellow #5 (Tartrazine), and FD&C Yellow #6 (Sunset Yellow). All but Orange B are also approved for use in Europe, but in Europe warning labels are now required on FD&C Red #40 (Allura Red AC), FD&C Yellow #5 (Tartrazine), FD&C Yellow #6 (Sunset yellow), and three colors used in Europe but not the United States: Quinoline Yellow, Carmoisine, and Ponceau.

In 2008, the Center for Science in the Public Interest (CSPI), a consumer advocacy organization, petitioned the FDA to regulate food color additives. They provided an unpublished literature review arguing that colorings contributed to behavior problems, and contended that there was little justification for incurring any health risks because, in their view, food colors provide no health benefits.¹³ The FDA subsequently commissioned its own review, which concluded in 2011 that the evidence fell short of a causal association. Both the CSPI and FDA reviews were qualitative; neither included a quantitative meta-analysis. Another major qualitative review in 2011¹⁴ reached a somewhat different conclusion: A subgroup of children with ADHD are sensitive to synthetic color additives, flavors, or salicylates and could benefit from a restricted diet. However, the authors did not quantify the magnitude of the behavioral effect.

In all, there is a crucial lack of consensus and lack of recent quantification about these alleged effects. If dietary interventions would be beneficial in a substantial portion of ADHD cases, then treatment guidelines would require revision. If evidence were to support a role for color additives in ADHD, then further policy and regulatory review would be needed. In addition, if such an association exists, it could inform investigation into causal mechanisms to gain clues as to how ADHD might develop. An updated quantification of effect sizes is therefore timely and of considerable importance.

METHOD

Studies were identified through a literature search using the PubMed, Cochrane Library, and PsycNET electronic databases. We initially used the following

combination of terms: behavior or ADHD or hyperactivity or impulsivity and food coloring; behavior or ADHD or hyperactivity or impulsivity and diet; behavior or ADHD or hyperactivity or impulsivity; and the name of each individual food coloring by either its formal or generic name (listed in parenthetical statement earlier), as well as the following additional terms: azorubine, Brilliant Blue, Brown FX, Fast Green FCF, Patent Blue V, Brown HT. A subsequent PubMed search using Medical Subject Heading (MeSH) terms was performed as follows to obtain additional human studies: limiting to clinical trials, randomized controlled trials, and/or comparative studies, and human: [((Food coloring agent or Food coloring agents) or (Food additive or Food additives)) and ((Attention Deficit Disorder with Hyperactivity or hyperactivity) or (hyperkinesis or hyperkinetic syndrome) or Child behavior disorders)]. Bibliographic searches of reviews and prior meta-analyses were necessary to identify earlier human studies. We included studies that evaluated the behavioral effects (relevant to inattention or hyperactivity) of the elimination of synthetic food colorings from the diet of subjects and/or challenged subjects with one or a combination of synthetic food colorings. We excluded studies that focused on the allergenic properties of food colorings and lacked behavioral end points, although in passing we note results of three human physiological studies.

This process identified 53 human studies, published from 1976 to February 2011, which were then graded for relevance. For Part 1 of our results (overview of the effects of a restriction diet), we noted open-label trials, but then required that studies meet the following criteria: be double blind and placebo controlled; have either a random assignment or crossover design; be conducted in a child or adolescent population; and evaluate restriction diets that include removal of food colors as part of a more general elimination diet. For Part 2 of our results (detailed examination of food color effects), we selected studies that met the following criteria: featured a double-blind, placebo-controlled, crossover design; used synthetic food colors as the manipulation of choice; and enabled computation of the effect size. In all, 35 publications were eligible for this review. Of these, 25 were studies of food colors, and six were studies of diet effects. The remaining four studies reported only open-label diet trials, and these studies were added to open-label trials reported in 10 of the prior studies for our open-label review. The remaining 18 articles examined nonbehavioral outcomes, conducted an uncontrolled food color study, had fewer than three participants, or allowed no effect size estimation.

Rational Grouping of Parent and Observer Ratings Data

Because a key issue in meta-analysis is appropriately combining studies based on sufficient similarity of

their measures, the following rational structure was adopted. First, for parent ratings, most studies had one type of parent data reported. When more than one type of parent data was reported, we chose the most psychometrically well-established measure (usually, early versions of the Conners rating scales). We then grouped the studies into the following two categories: "high-quality" outcome measures, which used ratings measures with published reliability and validity data; or "low-quality" outcomes, which either were custom measures, reported only undefined response rates, or had insufficient information to evaluate measure quality. We also coded studies for objective verification of blind, and we gave each study a global quality rating (see Table S1, available online, for details).

For nonparent reporters, we made the following decisions. First, if a teacher or observer provided more than one outcome rating, the most psychometrically established measure was used. When that information was not available, we pooled all ratings. Second, the resulting data were then coded as either high quality (psychometrically sound measure or formal observer coding with adequate inter-coder reliability) or low quality (unpublished rating scale, impressionistic observation rating, or observer ratings without psychometric data). Third, when a study reported both teacher and observer ratings, these were pooled into a summary effect size. However, we checked the effects by stratifying the ratings in three ways for secondary analysis: by data type (observations versus ratings), by reporter (teacher versus clinician/observer), and by setting (classroom versus clinic/research laboratory).

Rational Grouping of Psychometric Laboratory Measures

With regard to psychometric tests, the various tests reported clearly measured different cognitive abilities or functions. To address this, three clinical neuropsychologists (J.N. plus two neuropsychologists unfamiliar with the studies or the current review) sorted the tasks from all studies according to the cognitive abilities they assessed. The three raters agreed 100% on the assignment of tasks to those measuring attention (broadly defined) and those measuring other abilities (generally motor or language measures), with one exception noted. Tests that were classified as measuring attention broadly defined were the auditory memory test,¹⁵ visual memory test,¹⁵ matching familiar figures,¹⁶ zero-input tracking apparatus,¹⁷ and Paired Associate Learning Test (PALT).^{16,18,19} The motor measures were Beery visual motor integration,¹⁵ Handwriting,¹⁵ Draw a Child,¹⁵ balance on one leg, a test of coordination, and Ayres S.C. motor test.²⁰ Raters agreed unanimously on all test designations except the PALT, which was 2/3, with one rater voting that it measured neither attention nor motor functioning. We note in passing the results of three studies of discrete physiological functions (actigraphy,¹⁹ heart

rate response,²¹ and brain electrical activity)²²; we omitted one study of behavioral laboratory ratings.²³

Derivation of Effect Sizes

Because most studies did not report sufficient data to compute the correlation of scores across conditions, effect sizes were derived whenever possible from statistical parameters (t , F , exact p , or r). We computed effect sizes based on mean difference, standard deviation of differences, and across-condition correlation when those data were available. When we had to impute a correlation to compute the effect size based on means and standard deviations, we used the mean of all correlations available to us from six data sets for parent, teacher, and observer ratings, which was $r = 0.70$ (range, 0.40–0.91). For psychometric tests, which typically have lower test–retest reliability than rating scales, we set $r = 0.50$ and conducted sensitivity analyses of different correlations. When only a percent response was provided, odds ratios were calculated. When studies reported outcomes for multiple subgroups,^{11,24} the effect sizes were computed separately and pooled using a fixed-effects model within study. All effect sizes were converted to Hedges' g , a bias-free measure of standardized mean differences that can be interpreted in the same manner as the familiar Cohen's d .

Computations

All meta-analytic computations were performed using Comprehensive Meta-analysis software (Biostat Inc, Englewood, NJ). Studies with fewer than three participants were excluded, as they preclude calculation of a standard error for Hedges' g . A random-effects model was chosen for analysis due to the likelihood that different studies would reflect different effect sizes based on their methods and samples.²⁵ Effect-size heterogeneity was described using the Q statistic, τ (standard deviation of study effects), and I^2 (percentage of between-study variance that is not attributable to random sampling variation); however, examination of subgroups was driven by the *a priori* questions indicated below. When heterogeneity was high, we also considered sensitivity of effects via a leave-one-study-out procedure, in which effect sizes were recalculated with each study in turn removed, and we conducted sensitivity analysis by removing studies with significant ($z > 1.96$) residuals. The examination of publication bias was based on funnel plots and the trim-and-fill procedure. We also computed Orwin's fail-safe N of studies of various effects needed to reduce any observed effects to a trivial size (defined as $g = 0.05$).

Moderators and Subgroups

The *a priori* moderators that we reasoned were most important were as follows.

Quality or Type of Outcome Measure (ADHD or Its Symptoms). In the case of parents and observers, this was the quality rating of outcome, whether blind was validated, and global quality rating. In the case of tasks, we required that they assess attention, broadly defined.

Selection of the Sample to Be Diet Responsive. Preselected samples would be expected to yield larger effects than nonpreselected samples. We also considered whether samples were limited to children who were hyperactive or included typically developing children.

Content of the Challenge. We considered whether the challenge was restricted to food colors (some included other additives), as well as whether they were confined to synthetic colors allowed in the United States (FDA-approved colors). We conducted secondary meta-regression on the effects of dose and duration of challenge.

Overall, we highlight informative results in this report and provide additional details online, as noted later.

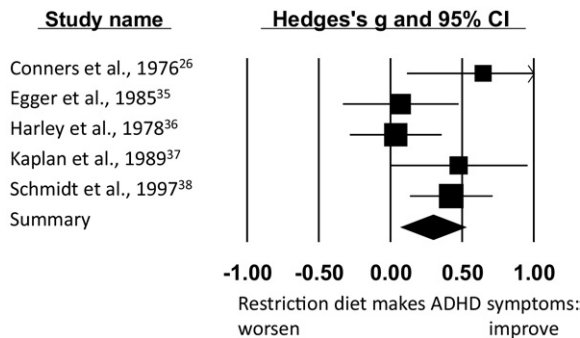
RESULTS

Dietary Restriction

Open-Label Trials. Before evaluating the effects of food colors, we examined the evidence regarding dietary restriction, in which synthetic food colors and/or other additives were eliminated as a component of the restrictive diet and ADHD symptoms were evaluated. Our literature review identified 14 open-label trials of hyperactive children that offered an operational definition of a "responder" (e.g., 25% improvement in symptoms), with aggregated $N = 2025$.^{20,24,26-34} (Note that several of these open-label trials of restriction diet were followed by a randomized controlled trial of food colors, so they fall into both groups of studies—open-label for diet, but randomized controlled trial for food color challenge.) In all, these studies yielded a random-effects-weighted response rate of 47.4% (95% confidence interval [CI] = 0.33–0.62), but with extreme heterogeneity ($Q = 332$; $p < .001$). This result, obviously vulnerable to placebo and experimenter effects, places an upper bound on restriction diet response rates.

Controlled Trials. We identified six restriction diet studies that used either a placebo-controlled diet challenge or a crossover design,^{26,33,35-38} which in aggregate examined 195 children for improvement in hyperactive symptoms. These studies yielded a summed response rate of 41.5% (95% CI = 22%–64%), again with substantial heteroge-

FIGURE 1 Restriction dietary crossover studies: homogenous results. The size of the square indicates the study weight, and the width of the diamond is the 95% confidence interval (CI).



neity ($Q = 29.7$; $p < .001$). A sensitivity analysis indicated that with the largest outlier (residual $z = 2.8$ ³³) removed, the response rate was 33% (95% CI = 19%–52%; $n = 164$). That outlier also had the weakest blinding by our rating. Therefore, a conservative response rate estimate was 33%.

In the six above-mentioned studies,^{26,33,35-38} we examined the average size of the change in symptoms, because the definition of the term “responder” varied arbitrarily across studies. Pooling all sources of outcome information (parents, teachers, and observers) within studies, the aggregate effect was $g = 0.58$ (SE = 0.25 [95% CI = 0.10–1.1]; $p = .019$). Again, heterogeneity was extremely high ($Q = 45.8$; $p < .001$). To resolve heterogeneity, we examined residuals. One outlier³³ was noted (the same as in the prior paragraph; standardized residual $z = 2.33$, $p = .020$; all other residuals $z < 1.03$, $p = \text{NS}$). It was a single-blind design with weak blinding that we had difficulty classifying. With that study removed, the effect was $g = 0.29$ (SE = 0.12 [95% CI = 0.16–0.52]; $p = .014$). Between-study variance was thereby substantially explained, with only modest variation remaining ($Q = 7.39$, $p = .12$). The final five studies for this effect are displayed in Figure 1. Within these studies, effects were similar in magnitude for parent versus teacher/observer ratings.

For the five homogeneous studies,^{26,35-38} there was no publication bias using the trim-and-fill estimator³⁹ with no missing studies imputed. The fail-safe analysis indicated that it would have required four studies (nearly as many as were published) with a negative effect of -0.25 to reduce the summary effect to a trivial $g = 0.05$.

One study of EEG effects of restriction diet showed reliable changes in frontal EEG power²² of $g = 0.71$, 95% CI = 0.11 to 1.31, $p = .021$, but no effects in central or posterior regions; a study of heart rate²¹ showed no consistent effects of food challenge.

Overall, although the handful of double-blind, crossover, diet-restriction studies used a relatively small total number of participants, their pooling indicated that dietary restriction produces a reliable and clinically meaningful benefit in children with ADHD. The effect size of 0.29 is approximately one-third the size of a summary medication effect of 0.9,⁴⁰ and is equivalent to a change from the 50th to 62nd percentile, thus potentially clinically meaningful. We therefore proceeded to our detailed analysis of food colors, the hypothesized component of these diets that is most discussed in terms of policy implications.

Synthetic Food Additives

Table 1 lists the studies used for our food colors meta-analysis (see Table S2, available online, for details on study inclusion and exclusion in each of the subsequent analyses). Typically, participants first followed a version of the Feingold diet or other restriction diet that eliminated synthetic color additives as well as other foods and additives (see previous section), and were then challenged with specially prepared cookies, juices, or capsules containing either one or a mixture of synthetic colors and, in some studies, other additives.

Parent Reports. For parent reports, we had 20 studies with 794 participants. Figure 2 portrays the results for the random-effects model across all available studies, with parent studies in the top portion of the figure. As shown, the effect is highly reliable, with a point estimate effect size of 0.18 (see Table S3, available online, for details on each study in Figure 2. Table 2 provides summary data on the model; Note that, as evident in Table 2, this result masked substantial heterogeneity across studies. Heterogeneity and effect size were not explained by any single study; a sensitivity analysis (removing one study at a time) yielded a range of effect sizes from 0.17 ($p = .001$) to 0.23. However, removal of the studies with significant residuals ($z > 1.96$)^{15,17,19} resulted in more acceptable homogeneity as indicated in the second line of Table 2.

We explored other moderation effects, as summarized in Table 2. As it shows, the effect held

TABLE 1 All Studies Considered in Meta-Analysis of Food Colors

First Author	Year	Nation	S and T ^a	N	Age	Selected ^b	Trial ^c	Dose ^d	Data ^d
Adams ¹⁵	1981	USA	Yes	18	7.5	Yes	FDA	26	P, O, A, M
Bateman et al ²³	2004	UK	No	277	3.7	No	Non	20	P, M
Conners ¹⁹	1980	USA	Yes	30	7.6	Yes	FDA	13	P
Conners et al ⁴¹	1980	USA	No	9	7.5	Yes	FDA	26	A
David ⁴²	1987	UK	Yes	24	5.2	Yes	FDA	250	P, O
Goyette et al a ¹⁷	1978	USA	Yes	16	8.3	Yes	FDA	26	A
Goyette et al b ¹⁷	1978	USA	Yes	13	6.2	Yes	FDA	26	P
Harley et al ³⁶	1978	USA	Yes	9	9.25	Yes	FDA	26	P, O
Levy et al ⁴³	1978	Australia	Yes	8	NR	Yes	Non	50	P
Mattes et al ⁴⁴	1981	USA	Yes	11	9.1	Yes	FDA	78	P, C, A
McCann et al a ^{11e}	2007	UK	No	130	3	No	Non	30	P, O
McCann et al b ^{11f}	2007	UK	No	119	8.5	No	Non	45	P, O, A
Pollock et al ⁴⁵	1990	UK	Yes	19	8.9	Yes	Non	125	P
Rapp ²⁰	1978	USA	Yes	24	10	No	FDA	NR	P, M
Rowe ⁴⁶	1988	Australia	Yes	8	9	Yes	Non	50	P
Rowe et al ²⁴	1994	Australia	Yes	54	7.1	Yes	FDA	50	P
Sarantinos et al ³⁴	1990	Australia	Yes	13	9	Yes	FDA	10	P
Spring et al ⁴⁷	1981	USA	Yes	6	10.2	Yes	FDA	26	P, T
Swanson et al ¹⁸	1980	USA	Yes	20	10	Hyp	FDA	150	A
Thorley ⁴⁸	1984	UK	Yes	10	11.7	MI	Non	91.8	T, A
Weiss et al ⁴⁹	1980	USA	Yes	22	4.5	Yes	FDA	35	P
Williams et al ⁵⁰	1978	CAN	Yes	26	8.5	Hyp	Non	26	P, T
Wilson et al ⁵¹	1989	UK	No	4 ^g	5.5	Yes	FDA	15	P

Note: NR = not reported

^aS and T, included in Schab and Trinh.¹⁰

^bThe term selected refers to how the sample was identified. "Yes," hyperactive children preselected as responsive to either a few foods diet or food colors by parent report or an open trial; "hyp," selected to be hyperactive but not selected to be diet or color responsive; "No," a community or population sample; "MI," a mentally impaired (as described in the publication), institutionalized sample.

^cTrial type, Food and Drug Administration (FDA)-approved colors only; non, other colors/additives included. All of the non-FDA studies examined food color additives alone, except Bateman et al.²³ and McCann et al., 2007.¹¹ Those studies included preservatives as well as food colors.

^dDose=mg/day maximum during study.

^eData types available are: parent ratings or observations (P); observer coding behavior frequency (O); clinician or child care worker impression rating (C); teacher rating (T); attention measure by computer or other psychometric measure (A); motor or other non-attention psychometric task (M).

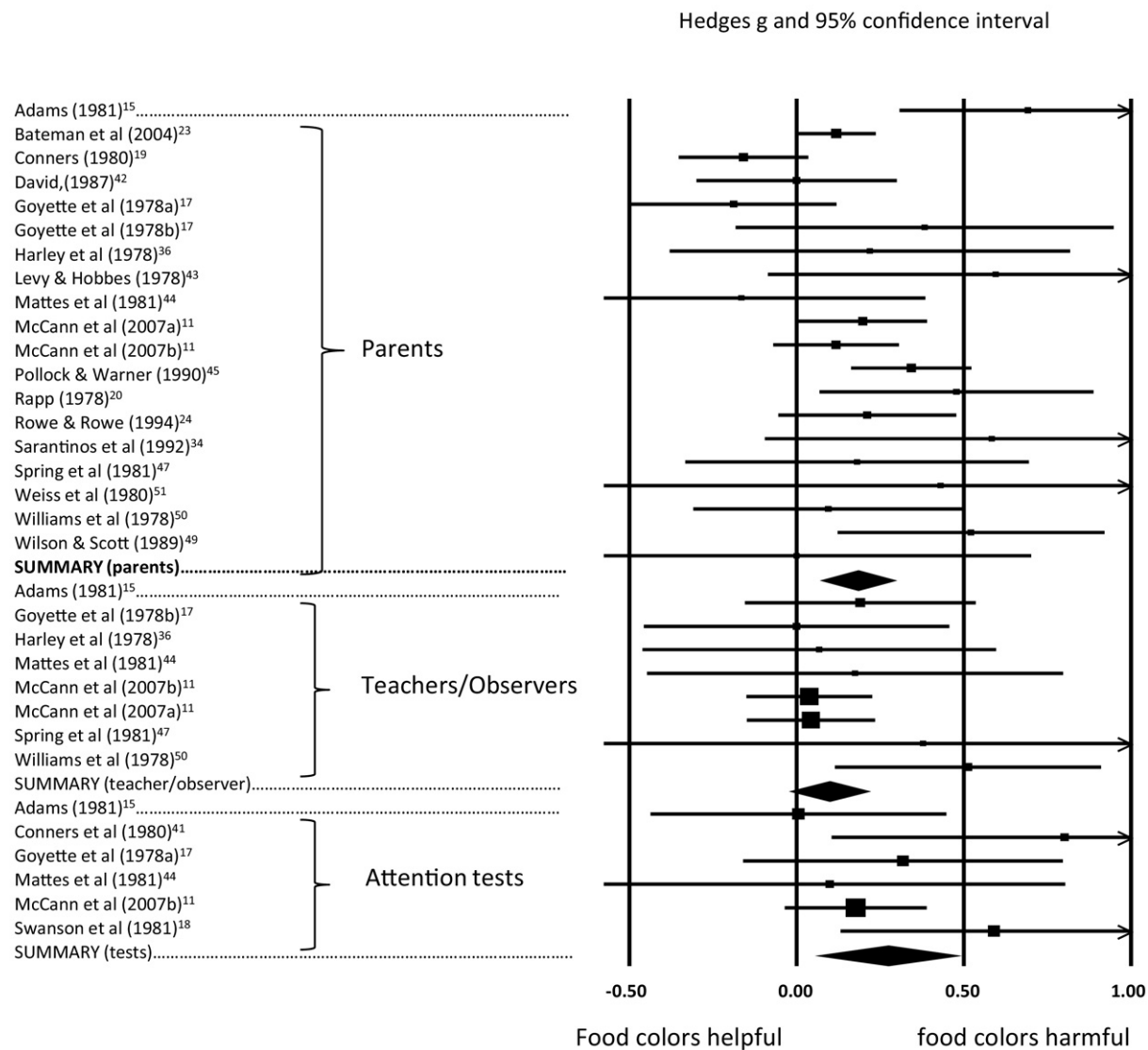
^fMcCann et al.¹¹ a is 3-year-olds and b is 8- to 9-year-olds. Their "mix A" and "mix B" conditions were pooled for estimated effect sizes.

^gWilson and Scott⁵¹ studied 19 children, but only 4 of them received a challenge with food colors based on our reading of the methods.

when blinding was carefully validated, but fell to just below significance when studies were restricted to high-quality outcome measures. However, effects were significant when restricted to those studies that challenged with food colors, but not when selected for high quality measures or for FDA-approved colors only. The significant effect seemed to be carried by studies of unselected samples, which overlapped with studies outside the United States that included additional colors and/or preservatives. Hyperactivity level of the sample did not affect results. Meta-regressions (see Table S1, available online) showed that parents reported larger effects with higher dose, but no effect of dose duration or of global quality.

Teacher/Observer Reports. For teacher and observer reports, we had 10 reports with a combined sample size of 323 participants. Effects by reporter (teacher vs observers) were nearly identical. Effects by setting suggested slightly more effect in classroom than laboratory or clinic settings, but were difficult to evaluate because of the small number of nonclassroom observations. Pooling across these effects, the random-effects model yielded a nonsignificant effect of food colors of 0.07 ($p = .14$; Table 2). As shown in Table 2, heterogeneity was minimal, lending confidence to the summary statistic. The effect grew slightly larger with higher quality measures ($g = 0.08$) but still shy of significance. It was reliable for studies confined to food colors ($p =$

FIGURE 2 Food colors results for parent report: all studies, teacher/observer reports of high quality studies, and psychometric tests of attention. The size of the square indicates the study weight, the line for individual studies indicates the 95% confidence interval (CI). The summary effects are shown for parent, teacher/observer, and test results by the diamond; the center of the diamond is the effect size and the width of the diamond is the 95% confidence interval. Comparison across studies using the graph metrics can only be done within domain.



.03) but not when restricted to FDA-approved colors (Table 2). However, there was only 1 small study of FDA-approved food colors using valid and reliable outcome measures and non-preselected samples. When the entire pool of studies was divided into preselected and nonpreselected samples, power was reduced and neither group separately yielded a significant g value. The forest plot for eight studies with reliable outcome measures is included in Figure 2, in the middle portion, with details on each study available in Table S3, available online).

Psychometric Tests of Attention. For psychometric test measures of attention, we had six studies examining 154 children, with the results portrayed in Figure 2 (bottom portion), and test-by-test results provided in detail in Table S3 (available online). The attention studies yielded a reliable pooled effect of 0.27 ($p = .007$) (Table 2 provides model details). Again, heterogeneity was modest, lending confidence to the estimate. Because few of these effect-size estimates relied on imputation of the repeated-measures correlation, the effect was robust to sensitivity analyses

TABLE 2 Summary of All Food Color Results Grouped by *A Priori* Moderators^a

Selection	Studies	N	Effect Size			Variability Across Studies			
			G	95% CI	P	Q	P	τ	ρ
Parent Report									
All	20	794	0.18	0.08–0.29	.0007	39.5	<.01	.15	52%
Two outliers removed ^b	18	746	0.18	0.10–0.27	.00003	21.7	.19	.08	22%
High-quality outcome	11	619	0.13	–0.00–0.25	.053	24.3	<.1	.15	58%
Low-quality outcome	9	175	.29	0.11–0.46	.002	12.0	.15	.15	33%
Blind validated	5	504	.14	0.05–0.22	.002	0.7	.95	.00	0%
Blind not validated	15	290	.21	0.05–0.37	.009	38.6	.001	.23	64%
Hyperactive only	11	185	.21	–0.02–0.43	.069	31.5	.01	.29	68%
Not purely hyperactive	9	609	.18	0.11–0.25	.0001	6.4	.59	.00	0%
Diet responder only	15	255	.16	0.01–0.33	.038	33.0	<.01	.21	57%
Not diet resp only	5	539	.20	0.07–0.32	.001	6.8	.18	.08	36%
Colors only	17	305	.21	0.06–0.36	.006	38.7	<.01	.22	59%
High quality	8	130	.13	–0.12–0.37	.307	23.6	<.01	.26	70%
FDA colors only	13	243	.13	–0.04–0.30	.128	25.6	.01	.21	53%
High quality	6	85	.00	–0.25–0.04	.167	4.9	.43	.00	0%
Teacher/Observer									
All	10	323	.07	–0.03–0.18	.146	6.0	.73	.00	0%
Teachers	7	490	.06	–0.03–0.14	.206	5.7	.46	.00	0%
Observers	4	475	.06	–0.03–0.14	.211	0.7	.86	.00	0%
Classroom	8	280	.08	–0.03–0.19	.160	5.3	.62	.00	0%
Clinic/laboratory	3	53	.03	–0.17–0.24	.750	1.9	.39	.00	0%
High quality	8	288	.10	–0.01–0.21	.086	5.5	.60	.00	0%
Colors only	6	76	.22	0.10–0.41	.030	3.4	.65	.00	0%
FDA only	5	50	.12	–0.01–0.35	.290	0.5	.96	.00	0%
Attention Tests									
All	6	154	.27	0.07–0.47	.007	6.4	.27	.12	22%
Colors only ^c	5	68	.34	0.06–0.62	.017	5.4	.24	.16	30%
FDA only	5	68	.34	0.06–0.62	.017	5.4	.24	.16	30%

Note: CI = confidence interval.

^aPreselected, sample prescreened to be diet responsive prior to food color challenge; Food and Drug Administration (FDA) only, challenge exclusively includes US FDA-approved colors; other studies included FDA-approved colors in addition to other colors. For parent and teacher/observer analyses, time 1 to time 2 correlation is set as in methods; for psychometric tests of attention, it is set to 0.50. Sensitivity analyses indicated no change in results with correlations varied from 0.2 to 0.8.

^bLine 2 under parent studies: Two outliers removed are Adams¹⁵ and Goyette et al.,¹⁷ both with residual $z > 1.96$.

^cSensitivity analysis with one study at a time removed showed that this effect survived removal of some studies but not others; details provided in Table S4, available online.

for time 1 to time 2 correlations (if $r = 0.75$, $g = 0.244$, $p = .006$; if $r = 0.25$, $g = 0.243$, $p = .002$). As shown in Table 2, the attention task effect held for studies of colors only and for studies of FDA-approved colors only, albeit with a pooled sample size of only 68 participants. Although heterogeneity was minimal and there were no statistical outliers (largest residual $z = 1.47$), the color-only effect was vulnerable to single-study effects (see Table S4, available online, for details).

Motor and physiological results were too variable for meaningful pooling; however, significant effects were seen for individual measures in single studies (see Table S4, available online, for details).

When studies were grouped according to whether children had been preselected to be sensitive to food additives, results did not vary significantly across these groups of studies. Most results held when restricted to children not pre-selected (see Table S5, available online, for detailed results).

Publication Bias. Publication bias was evaluated with a funnel plot, and the effects of bias, if it might be present, were evaluated using a trim-and-fill procedure. This procedure imputes the point values for studies that are “missed” in the estimated bias and recomputes the effect size. We also considered Orwin’s fail-safe N. In the latter analysis, a “trivial” size g was set at

TABLE 3 Publication Bias Analysis for Food Color Studies^a

Selection	Trim and Fill		95% CI	P	Orwin's Fail-safe N		
	Imputed	Adjusted g			g = 0	g = -0.10	g = -0.25
Parent Report							
All	4	.119	0.01–0.23	<.05 ^b	42	14	7
Two outliers out	3	.153	0.06–0.24	<.05 ^b	45	15	8
Selected	4	.077	–0.08–0.24	>.05	27	9	5
Unselected	2	.140	–0.01–0.28	>.05	12	4	2
Blind verified	2	.135	–0.05–0.22	<.05 ^b	9	3	2
Colors only	6	.077	–0.08–0.23	>.05	41	14	7
Teacher/Observer							
Colors	0	No change	–	<.05 ^b	21	7	4
Psychometric Tests							
Attention only	1	.232	0.01–0.45	<.05 ^b	24	8	4
Colors/FDA	0	No change	–	<.05 ^b	29	10	4

Note: CI = confidence interval; FDA = Food and Drug Administration.
^aImputed studies indicate the number of studies imputed to eliminate publication bias; if the bias is due to publication bias, the adjusted g and 95% CI would be as shown. Irwin's fail-safe N shows how many studies with effects of 0, –0.10, and –0.25 (negative studies) would be needed to reduce the effect to g = 0.05.
^bIndicates effects that remain significant after adjusting for putative publication bias.

0.05, and we computed the number of studies needed to bring the point estimate below this level if they all had effect sizes of 0, –0.10, or –0.25. Table 3 shows the results for key analyses and reveals that the parent global result and the result for validated studies survive correction for publication bias, but the results for colors only (no other additives) do not. It should also be noted that the trim-and-fill procedure is less reliable in conditions of very high heterogeneity, as seen in the parent data; thus there may be true differences in effect sizes for the small versus large studies.

As shown in Table 3, teacher/observer results for studies confined to food colors were unaffected by potential publication bias. Results for psychometric tests of attention showed no evidence of publication bias, either. In all, the results that were insensitive to possible bias (the most dependable results) were the following: the parent-rated overall result (including studies to examine additives in addition to food colors); observer-rated data on food colors studies; and psychometric tests of attention.

Percentage of Children Reacting. Fourteen studies,^{15,19,20,24,34,42,45,47,49,51,52} including three studies reported by Connors,¹⁹ attempted to gauge the percentage of children reacting to food colors. All were based on children who had been preselected to be diet responsive by either an open-label trial or parent report. The criteria for response varied from a general impression to a

formal within-subject statistical test. Across all 14 reports (n = 241, with 51 “responders”), the event rate was 18% (95% CI = 0.09–0.32; p < .0001). Among the eight studies with what we coded as using well-defined criteria for responders,^{15,17,19,20,24,46,47,49} in (n = 176), including two studies in Goyette et al.,¹⁷ the event rate was 24% (95% CI = 0.13–0.40; p = .002). Variation among the studies was high (Q = 17.9; p < .01). Because the 24% figure is the response rate to food colors among diet responders (studies of general population samples are excluded here), we combined these findings with the estimation of how many children with ADHD are diet responders from Part 1 of our results. We calculated that 8% (24% × 33%) of children with ADHD may have symptoms related to food colors. A confidence interval around that figure could not be computed, but would be large.

However, these effects were not specific to food colors. Studies that considered the question of specificity of effects concluded that children who react to food colors also react to other food ingredients.^{20,35}

DISCUSSION

The relation of food color additives to ADHD has provoked periodic controversy for nearly 40 years, but has gained renewed currency with review by European and American authorities recently taking

different policy actions. The present meta-analysis provides new clarification and nuance, and overturns some previous conclusions. It is important to note that we emphasized randomized controlled trials, because these provide *prima facie* evaluation of causality. We began by examining whether a restriction diet improves ADHD symptoms, concluding that it does so, with perhaps 30% of children with ADHD being responsive.

We then proceeded to examine food color additives, one component of a restriction diet. First, across all studies, ratings from parents produced a highly reliable effect of synthetic food colors at a standardized effect of 0.18, which was reduced to 0.12 when restricted to studies with well-described measures of ADHD. Because of findings from more recent studies, it is no longer the case that this effect is carried by preselected samples. However, although the effect was reliable in studies of food colors without preservatives, it was not reliable for studies restricted to FDA-approved food colors only. Thus, the parent rating effect seemed to be carried by studies outside the United States, which either utilized additional coloring agents or added preservatives to their challenge, or both. It can be concluded that some food additives provoke increased symptoms as rated by parents, but it is not clear whether this effect can be attributed to FDA-approved food colors. Our effect size estimate is slightly smaller than that estimated by Schab and Trinh,¹⁰ who did not separate the effect size estimate by studies restricted to FDA-approved colors or quality of ADHD measure.

Results for observer/teacher ratings still fail to show reliable effects in aggregate despite the addition of new studies since the meta-analysis by Schab and Trinh.¹⁰ This observer-rated effect was reliable for studies of food colors only, but again not when restricted to FDA-approved colors. Although true variation in study effect sizes cannot be ruled out, and although controversy remains regarding the validity of correcting for publication bias⁵³ the result was a lack of a consistent finding across parent and teacher/observer ratings, after quality of measure was taken into account.

Notable in this light was that nearly all studies examined combinations of colors, with too little consistency in their mixtures for us to test comparative effect sizes of different mixtures or individual compounds. Likewise, dose and exposure length varied. Blinding quality also varied; only a

small number conducted blind verification and even those attempts were often incomplete by contemporary standards. These issues may account for some of the high variability in the parent-report data as well as some of the inconsistency across models that included different studies. Clearly needed are studies that quantify comparative effects of individual colors and additives or competing specific mixtures; those studies might benefit from updated consideration of the range of intakes by children today.

In view of these mixed results, it is important to note that no prior meta-analysis evaluated objective, computerized measures of attention. Thus, an important contribution here is that we observed reliable effects in this analysis, which were not explainable by publication bias, at an effect size of 0.27. This effect also held when the finding was restricted to studies of only FDA-approved colors. This finding suggests that the pursuit of studies of the effects on attention and mental control may be fruitful. Despite the reliability of the psychometric test finding, with 154 participants across all studies and only 68 participants in studies in the United States, the generalizability of these findings for clinical or policy recommendations is questionable at best and individual studies can still be criticized for methodological limitations. The result for studies restricted only to food colors was, due to the small number of studies and small aggregate sample, vulnerable to single study effects. Even so, the overall pooled result is sufficiently robust that it would not be easily overturned even if several unpublished, negative results exist.

Indeed, it is striking that despite several recent studies in Europe, in the past 20 years there have been no studies restricted only to FDA-approved food colors and almost no studies of unselected samples. These gaps reflect a complete absence of modern studies of this topic in the United States since the early 1990s. The literature remains limited by lack of validation of blinding in many studies, and wide variation in methodology which would be best addressed by a pooled analysis of individual data across studies—not possible with the old literature. In short, despite 35 years of research and evidence of an effect of food colors on objective measures of attention, the database that would confirm this possibility and generalize it for contemporary use is woefully out of date with regard to policy or clinical decisions in the United States.

Overall, a mixed conclusion must be drawn. Although the evidence is too weak to justify action recommendations absent a strong precautionary stance, it is too substantial to dismiss. As for substantiality, it was striking that restricted food diets could be helpful in managing ADHD in a clinically meaningful subset of cases, and should continue to be investigated. If predictors of dietary response could be identified, a treatment avenue might be opened for some children. It also seems that food additives can influence ADHD symptoms across different outcome measures, with effects that are in the range of one-sixth to one-third the size of a medication effect, depending on the outcome measure and challenge type. Although these average effect sizes are small in clinical terms, they could be quite substantial from the perspective of population-wide prevention efforts. Intriguing was that reliable effects were observed in psychometric laboratory tests of attention. However, in counterpoint, the limitations of these findings are important. The confidence in or generalizability of the food color findings is limited by the lack of consistency in the findings across information sources, small pooled samples for studies restricted only to FDA-approved colors and for the psychometric test data, outdated studies in the United States, and vulnerability to publication bias of findings from parent studies.

In conclusion, approximately 33% of children with ADHD may respond to a dietary interven-

tion. Although as many as 8% may have symptoms related to food colors, the source of most of this dietary response remains unclear. We thus conclude that dietary effects on and treatments of ADHD, including food additives and colors, deserve renewed investigation. &

Accepted October 24, 2011.

Drs. Nigg and Edinger are with Oregon Health and Science University, Portland, OR. Drs. Lewis and Falk are with the Life Sciences Research Organization, Bethesda, MD.

This study was supported by the North American Branch of the International Life Sciences Institute (ILSI) Technical Committee on Food and Chemical Safety. ILSI North America is a public, nonprofit foundation that provides a forum to advance understanding of scientific issues related to the nutritional quality and safety of the food supply by sponsoring research programs, educational seminars and workshops, and publications. ILSI North America receives support primarily from its industry membership. The Life Sciences Research Organization Inc and Dr. Nigg received a grant from the organization for their work reviewing, analyzing, and summarizing the information contained in this article. Partial funding for this project was also received from the National Confectioners Association.

The opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of either funding organization.

Disclosure: Drs. Falk and Lewis are employees of the Life Sciences Research Organization, Inc. (LSRO). LSRO has received research support from numerous federal government agencies (including the U.S. Food and Drug Administration), trade associations, and corporations (including food and dietary supplement manufacturers and distributors). Drs. Nigg, Lewis, Edinger, and Falk report no biomedical financial interests or potential conflicts of interest.

Correspondence to Joel T. Nigg, Ph.D., Department of Psychiatry, Oregon Health and Science University, 3181 Sam Jackson Park Road, Portland, OR 97239; e-mail: niggj@ohsu.edu

0890-8567/\$36.00/©2012 American Academy of Child and Adolescent Psychiatry

DOI: 10.1016/j.jaac.2011.10.015

REFERENCES

1. Chronis-Tuscano A, Molina BS, Pelham WE, *et al.* Very early predictors of adolescent depression and suicide attempts in children with attention-deficit/hyperactivity disorder. *Arch Gen Psychiatry.* 2010;67:1044-1051.
2. Marcus DK, Barry TD. Does attention-deficit/hyperactivity disorder have a dimensional latent structure? A taxometric analysis. *J Abnorm Psychol.* 2011;120:427-442.
3. Purcell S. Variance components models for gene-environment interaction in twin analysis. *Twin Res.* 2002;5:554-571.
4. Nigg J, Nikolas M, Burt SA. Measured gene-by-environment interaction in relation to attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry.* 2010;49:863-873.
5. Stevenson J, Sonuga-Barke E, McCann D, *et al.* The role of histamine degradation gene polymorphisms in moderating the effects of food additives on children's ADHD symptoms. *Am J Psychiatry.* 2010;167:1108-1115.
6. Nigg JT. Temperament and developmental psychopathology. *J Child Psychol Psychiatry.* 2006;47:395-422.
7. Shannon WR. Neuropathic manifestations in infants and children as a result of anaphylactic reactions to foods contained in their diet. *Am J Dis Child.* 1922;24:89-94.
8. Feingold BF. Hyperkinesis and learning disabilities linked to artificial food flavors and colors. *Am J Nurs.* 1975;75:797-803.
9. Kavale KA, Forness SR. Hyperactivity and diet treatment: a meta-analysis of the Feingold hypothesis. *J Learn Disabil.* 1983; 16:324-330.
10. Schab DW, Trinh NH. Do artificial food colors promote hyperactivity in children with hyperactive syndromes? A meta-analysis of double-blind placebo-controlled trials. *J Dev Behav Pediatr.* 2004;25:423-434.
11. McCann D, Barrett A, Cooper A, *et al.* Food additives and hyperactive behaviour in 3-year-old and 8/9-year-old children in the community: a randomised, double-blinded, placebo-controlled trial. *Lancet.* 2007;370:1560-1567.
12. FDA. US Food and Drug Administration. Summary of color additives for use in United States in foods, drugs, cosmetics, and medical devices. 2011. Available at: <http://www.fda.gov/forindustry/coloradditives/coloradditiveinventories/ucm115641.htm>. Accessed April 2, 2011.
13. Jacobson MF. Petition to ban the use of Yellow 5 and other food dyes, in the interim to docket require a warning on foods containing these dyes, to correct the information the Food and Drug Administration gives to consumers on the impact of these dyes on the behavior of some children, and to require neurotoxicity testing of new food additives and food colors. 2008. Available at: <http://www.cspinet.org/new/pdf/petition-food-dyes.pdf>. Accessed May 27, 2011.
14. Stevens LJ, Kuczek T, Burgess JR, Hurt E, Arnold LE. Dietary sensitivities and ADHD symptoms: thirty-five years of research. *Clin Pediatr.* 2011;50:279-293.
15. Adams W. Lack of behavioral effects from Feingold diet violations. *Percept Mot Skills.* 1981;52:307-313.

16. Carter CM, Urbanowicz M, Hemsley R, *et al.* Effects of a few food diet in attention deficit disorder. *Arch Dis Child.* 1993;69:564-568.
17. Goyette GH, Connors CK, Petti TA, Curtis LE. Effects of artificial colors on hyperkinetic children: a double-blind challenge study [proceedings]. *Psychopharmacol Bull.* 1978;14:39-40.
18. Swanson JM, Kinsbourne M. Food dyes impair performance of hyperactive children on a laboratory learning test. *Science.* 28 1980;207:1485-1487.
19. Connors CK. Foods, food dyes, and allergies. In: *Food Additives and Hyperactive Children.* New York: Plenum; 1980:77-85.
20. Rapp DJ. Does diet affect hyperactivity? *J Learn Disabil.* 1978;11: 383-389.
21. Salamy J, Shucard D, Alexander H, Peterson D, Braud L. Physiological changes in hyperactive children following the ingestion of food additives. *Int J Neurosci.* 1982;16:241-246.
22. Uhlig T, Merckenschlager A, Brandmaier R, Egger J. Topographic mapping of brain electrical activity in children with food-induced attention deficit hyperkinetic disorder. *Eur J Pediatr.* 1997;156: 557-561.
23. Bateman B, Warner JO, Hutchinson E, *et al.* The effects of a double blind, placebo controlled, artificial food colourings and benzoate preservative challenge on hyperactivity in a general population sample of preschool children. *Arch Dis Child.* 2004;89:506-511.
24. Rowe KS, Rowe KJ. Synthetic food coloring and behavior: a dose response effect in a double-blind, placebo-controlled, repeated-measures study. *J Pediatr.* 1994;125:691-698.
25. Borenstein M. *Introduction to Meta-Analysis.* Hoboken, NJ: Wiley; 2009.
26. Connors CK, Goyette CH, Southwick DA. Food additives and hyperkinesis: preliminary report of a double-blind crossover experiment. *Psychopharmacol Bull.* 1976;12:10-11.
27. Boris M, Mandel FS. Foods and additives are common causes of the attention deficit hyperactive disorder in children. *Ann Allergy.* 1994;72:462-468.
28. Breakey J, Hill M, Reilly C. A report on a trial of the low additive, low salicylate diet in the treatment of behaviour and learning problems in children. *Aust J Nutr Diet* 1991;48:89-94.
29. Cook PS, Woodhill JM. The Feingold dietary treatment of the hyperkinetic syndrome. *Med J Aust.* 1976;2:85-88, 90.
30. Holborow P, Elkins J, Berry P. The effect of the Feingold diet on 'normal' school children. *J Learn Disabil.* 1981;14:143-147.
31. Loblay R. Adverse reactions to tartrazine. *Food Technol Aust* 1985;37:508-510.
32. Pelsser LM, Frankena K, Toorman J, Savelkoul HF, Pereira RR, Buitelaar JK. A randomised controlled trial into the effects of food on ADHD. *Eur Child Adolesc Psychiatry.* 2009;18:12-19.
33. Pelsser LM, Frankena K, Toorman J, *et al.* Effects of a restricted elimination diet on the behaviour of children with attention-deficit hyperactivity disorder (INCA study): a randomised controlled trial. *Lancet.* 2011;377:494-503.
34. Sarantinos J, Rowe KS, Briggs DR. Synthetic food colouring and behavioural change in children with attention deficit disorder: a double-blind, placebo-controlled, repeated measures study. *Proc Nutr Soc Australia.* 1990;15:233.
35. Egger J, Carter CM, Graham PJ, Gumley D, Soothill JF. Controlled trial of oligoantigenic treatment in the hyperkinetic syndrome. *Lancet.* 1985;1:540-545.
36. Harley JP, Ray RS, Tomasi L, *et al.* Hyperkinesis and food additives: testing the Feingold hypothesis. *Pediatrics.* 1978;61:818-828.
37. Kaplan BJ, McNicol J, Conte RA, Moghadam HK. Dietary replacement in preschool-aged hyperactive boys. *Pediatrics.* 1989;83:7-17.
38. Schmidt MH, Mocks P, Lay B, *et al.* Does oligoantigenic diet influence hyperactive/conduct-disordered children—a controlled trial. *Eur Child Adolesc Psychiatry.* 1997;6:88-95.
39. Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics.* 2000;56:455-463.
40. Faraone SV, Buitelaar J. Comparing the efficacy of stimulants for ADHD in children and adolescents using meta-analysis. *Eur Child Adolesc Psychiatry.* 2010;19:353-364.
41. Connors CK, Goyette CH, Newman EB. Dose-time effect of artificial colors in hyperactive children. *J Learn Disabil.* 1980;13: 512-516.
42. David TJ. Reactions to dietary tartrazine. *Arch Dis Child.* 1987; 62:119-122.
43. Levy F, Hobbes G. Hyperkinesis and diet: a replication study. *Am J Psychiatry.* 1978;135:1559-1560.
44. Mattes JA, Gittelman R. Effects of artificial food colorings in children with hyperactive symptoms. A critical review and results of a controlled study. *Arch Gen Psychiatry.* 1981;38:714-718.
45. Pollock I, Warner JO. Effect of artificial food colours on childhood behaviour. *Arch Dis Child.* 1990;65:74-77.
46. Rowe KS. Synthetic food colourings and 'hyperactivity': a double-blind crossover study. *Aust Paediatr J.* 1988;24:143-147.
47. Spring C, Vermeersch J, Blunden D, Sterling H. Case studies of effects of artificial food colors on hyperactivity. *J Spec Educ* 1981;15:361-372.
48. Thorley G. Pilot study to assess behavioural and cognitive effects of artificial food colours in a group of retarded children. *Dev Med Child Neurol.* 1984;26:56-61.
49. Weiss B, Williams JH, Margen S, *et al.* Behavioral responses to artificial food colors. *Science.* 1980;207:1487-1489.
50. Williams JL, Cram DM, Tausig FT, Webster E. Relative effects of drugs and diet on hyperactive behaviors: an experimental study. *Pediatrics.* 1978;61:811-817.
51. Wilson N, Scott A. A double-blind assessment of additive intolerance in children using a 12 day challenge period at home. *Clin Exp Allergy.* 1989;19:267-272.
52. Harley JP, Matthews CG, Eichman P. Synthetic food colors and hyperactivity in children: a double-blind challenge experiment. *Pediatrics.* 1978;62:975-983.
53. Ioannidis JP. Interpretation of tests of heterogeneity and bias in meta-analysis. *J Eval Clin Pract.* 2008;14:951-957.

TABLE S1 Meta-regression Results for Age, Dose, and Study Quality on Parent and Teacher Reports

Meta-regression exposure duration, parent report, all studies							
	Estimate	SE	Lower	Upper	z	p	
Slope	0.000	0.000	0.000	0.001	0.750	.453	
Intercept	0.120	0.054	0.014	0.226	2.223	.026	
τ^2	0.027						
Q	df	p					
Model	0.563	1	.453				
Residual	38.629	17	.002				
Total	39.193	18	.003				
Meta-regression child age, parent report, all studies							
	Estimate	SE	Lower	Upper	z	p	
Slope	0.023	0.016	-0.008	0.054	1.461	0.144	
Intercept	0.005	0.106	-0.204	0.213	0.042	0.966	
τ^2	0.029						
Q	df	p					
Model	2.134	1	.144				
Residual	37.059	17	.003				
Total	39.193	18	.003				
Meta-regression dose, parent report, all studies ^a							
	Estimate	SE	Lower	Upper	z	p	
Slope	0.002	0.001	0.000	0.004	2.120	.034	
Intercept	0.064	0.053	-0.039	0.167	1.218	.223	
τ^2	0.025						
Q	df	p					
Model	4.496	1	.034				
Residual	31.278	15	.008				
Total	35.774	16	.003				
Meta-regression dose, parent report, all studies ^b							
	Estimate	SE	Lower	Upper	z	p	
Slope	0.000	0.003	-0.005	0.006	0.154	.878	
Intercept	0.113	0.089	-0.061	0.286	1.271	.204	
Q	df	p					
Model	0.024	1	.878				
Residual	30.816	14	.006				
Total	30.839	15	.009				

TABLE S1 Continued

Meta-regression 6-point quality score and parent rating ^c						
	Estimate	SE	Lower	Upper	z	p
Slope	-0.020	0.019	-0.058	0.018	-1.024	.306
Intercept	0.249	0.098	0.056	0.441	2.527	.012
r ²	0.027					
	Q	df	p			
Model	1.048	1	.306			
Residual	38.145	17	.002			
Total	39.193	18	.003			
Teacher data age meta-regression						
	Estimate	SE	Lower	Upper	z	p
Slope	0.008	0.019	-0.030	0.046	0.415	.678
Intercept	0.021	0.139	-0.252	0.294	0.151	.880
r ²	0					
	Q	Df	p			
Model	0.172	1	0.678			
Residual	5.854	8	0.664			
Total	6.026	9	0.737			
Teacher data dosage meta-regression						
	Estimate	SE	Lower	Upper	z	p
Slope	0.008	0.019	-0.030	0.046	0.415	.678
Intercept	0.021	0.139	-0.252	0.294	0.151	.880
r ²	0					
	Q	df	p			
Model	0.172	1	.678			
Residual	5.854	8	.664			
Total	6.026	9	.737			
Teacher data duration of trial meta-regression						
	Estimate	SE	Lower	Upper	Z	p
Slope	0.008	0.019	-0.030	0.046	0.415	.678
Intercept	0.021	0.139	-0.252	0.294	0.151	.880
r ²	0					
	Q	df	p			
Model	0.172	1	.678			
Residual	5.854	8	.664			
Total	6.026	9	.737			

TABLE S1 Continued

Teacher global quality ^d						
	Estimate	SE	Lower	Upper	z	p
Slope	-0.004	0.027	-0.057	0.048	-0.154	.878
Intercept	0.107	0.135	-0.156	0.371	0.798	.425
r^2	0					
	Q	df	p			
Model	0.024	1	.878			
Residual	6.218	8	.623			
Total	6.241	9	.716			

Note: SE = standard error.

^aExcludes Rapp¹ (dose no applicable/available, sublingual delivery), David² (dose outlier); the significant and slightly positive slope effect indicates a tendency for larger effects with larger doses.

^bAlso excludes Pollock *et al.*,³ as the next highest outlier, in an effort to ascertain any effect.

^cThe quality score gave one point for each of the following: standardized and reliable outcome measure, successful effort to validate blind; well explained and convincing blinding attempt; data presented clearly and completely; analytic approach consistent with a priori expectations of analysis (minimal capitalization on chance); and absence of any other obvious or major flaw in the study. The negative slope indicates a slight tendency for higher quality studies to yield smaller effect sizes, but this association was not statistically reliable.

^dThe quality score gave one point for each of the following: standardized and reliable outcome measure, successful effort to validate blind; well-explained and convincing blinding attempt; data presented clearly and completely; analytic approach consistent with a priori expectations of analysis (minimal capitalization on chance); and absence of any other obvious or major flaw in the study. The negative slope indicates a slight tendency for higher quality studies to yield smaller effect sizes, but this association was trivial in size and not statistically reliable.

TABLE S2 Details of Reasons for Exclusion and Inclusion of Each Study from Each Analysis

1 st Author, Year	Parent	Teacher	Test
Adams, 1981 ⁴	2	1	1,2,3,
Bateman et al, 2004 ⁵	3,4	NA	4
Conners, 1980, Ch 5 ⁶	1	NA	NA
Conners et al, 1980 ^{7 a}	NA	NA	1,4
David, 1987 ²	2	2	NA
Goyette et al, 1978a ^{8 b}	1	NA	1
Goyette et al, 1978b ^{8 c}	1	NA	NA
Gross et al, 1987 ⁹	NA	NA	NA
Harley et al, 1978 ¹⁰	1	1	NA
Levy and Hobbes, 1978 ¹¹	2, 4	NA	NA
Mattes and Gittelman, 1981 ¹²	1	2	1
McCann et al, 2007a ¹³	3, 4	3,4	NA
McCann et al, 2007b ¹³	3, 4	3,4	1
Pollock and Warner, 1990 ³	4	NA	NA
Rapp, 1978 ¹	2	NA	2
Rowe, 1988 ¹⁴	2, 4	NA	NA
Rowe and Rowe, 1994 ¹⁵	2, 4	NA	NA
Sarantinos et al, 1990 ¹⁶	2	NA	NA
Spring et al, 1981 ¹⁷	1	1	NA
Swanson and Kinsbourne, 1980 ¹⁸	NA	NA	1
Thorley, 1994 ¹⁹	NA	2	NA
Weiss et al, 1980 ²⁴	2	NA	NA
Williams et al, 1978 ²¹	4	1	NA
Wilson et al, 1989 ²²	2	NA	NA
Total possible	20	10	9

Note: Codes for Inclusion in Parent and Teacher/Observer Based analyses: 1 = data used in all analyses; 2 = data used, but excluded from analyses of high quality outcome measures; 3 = data used, but excluded from analyses of "dyes only" (has preservatives); 4 = data used, but excluded from analyses of Food and Drug Administration (FDA) only (has non-FDA colors).

Codes for Inclusion in Test based analyses: 1 = Attention test available (Matching Familiar Figures test, Partially Accelerated Life Test, Continuous Performance Test, Zero Input Tracking Analyzer, or other test); 2 = motor test available; 3 = language test available; 4 = other test available—not classified, excluded, or detailed in Table S3, available online.

^aConners, Newcorn, and Goyette, 1980; same study as Conners, 1980, chapter 6.

^bSame as Conners, 1980, chapter 3

^cSame as Conners, 1980, chapter 4

TABLE S3 Detailed Data Corresponding to Figure 2 for Each Domain of Outcome

Study	Domain	g	SE	Lower	Upper	z	p
Adams, 1981 ²³	Parent	0.692	0.196	0.308	1.076	3.531	.000
Bateman et al, 2004 ⁵	Parent	0.119	0.060	0.001	0.236	1.968	.049
Conners, 1980, Ch 5 ⁶	Parent	-0.159	0.099	-0.353	0.035	-1.604	.109
David, 1987 ²	Parent	0.000	0.153	-0.300	0.300	0.000	1.000
Goyette et al, 1978a ⁸	Parent	-0.188	0.157	-0.495	0.119	-1.199	.231
Goyette et al, 1978b ⁸	Parent	0.383	0.288	-0.182	0.948	1.329	.184
Harley et al, 1978 ¹⁰	Parent	0.219	0.306	-0.380	0.818	0.718	.473
Levy and Hobbes, 1978 ¹¹	Parent	0.595	0.348	-0.086	1.277	1.712	.087
Mattes et al, 1981 ¹²	Parent	-0.165	0.281	-0.715	0.385	-0.588	.557
McCann et al, 2007a ^{12 a}	Parent	0.198	0.098	0.005	0.391	2.015	.044
McCann et al, 2007b ^{13 a}	Parent	0.118	0.096	-0.070	0.306	1.230	.219
Pollock and Warner, 1990 ³	Parent	0.343	0.092	0.163	0.523	3.736	.000
Rapp, 1978 ¹	Parent	0.478	0.209	0.068	0.888	2.286	.022
Rowe and Rowe, 1994 ¹⁵	Parent	0.211	0.136	-0.055	0.477	1.555	.120
Rowe, 1988 ¹⁴	Parent	0.584	0.347	-0.095	1.263	1.685	.092
Sarantinos et al, 1992 ¹⁶	Parent	0.181	0.262	-0.333	0.695	0.691	.490
Spring et al, 1981 ¹⁷	Parent	0.431	0.957	-1.444	2.306	0.450	.652
Weiss et al, 1980 ²⁴	Parent	0.095	0.206	-0.309	0.499	0.461	.645
Williams et al, 1978 ²¹	Parent	0.522	0.204	0.123	0.920	2.563	.010
Wilson and Scott, 1989 ²²	Parent	0.000	0.358	-0.701	0.701	0.000	1.000
Adams, 1981 ²³	Tchr/Obs ^b	0.191	0.176	-0.155	0.536	1.082	.279
Goyette et al, 1978a ⁸	Tchr/Obs ^b	0.000	0.233	-0.457	0.457	0.000	1.000
Harley et al, 1978 ¹⁰	Tchr/Obs ^b	0.068	0.270	-0.461	0.596	0.251	.802
Mattes et al, 1981 ¹²	Tchr/Obs ^b	0.175	0.317	-0.447	0.797	0.551	.582
McCann et al, 2007a ¹³	Tchr/Obs ^b	0.043	0.098	-0.149	0.234	0.439	.660
McCann et al, 2007b ¹³	Tchr/Obs ^b	0.039	0.096	-0.149	0.227	0.407	.684
Spring et al, 1981-T ¹⁷	Tchr/Obs ^b	0.378	0.949	-1.482	2.238	0.398	.690
Williams et al, 1978-T ²¹	Tchr/Obs ^b	0.512	0.203	0.114	0.910	2.524	.012
Adams, 1981 ²³	Tests	0.005	0.226	-0.437	0.448	0.024	.981
Conners et al, 1980 ⁶	Tests	0.801	0.355	0.105	1.498	2.255	.024
Goyette et al, 1978a ⁸	Tests	0.318	0.244	-0.160	0.796	1.304	.192
Mattes et al, 1981 ¹²	Tests	0.099	0.359	-0.605	0.803	0.276	.783
McCann et al, 2007b ¹³	Tests	0.177	0.108	-0.036	0.389	1.630	.103
Swanson et al, 1981 ¹⁸	Tests	0.590	0.234	0.131	1.049	2.522	.012

Note:
^aMcCann et al, 2007a is their study of 3 year olds; 2007b is their study of 8-9 year olds.
^bTeacher ratings (Tchr) or observer coding (Obs), or both, were available. Here, they are pooled within study.

TABLE S4 Task-based Studies, by Individual Study and by Type of Task

1 st Author, Year	Task	n	g	SE	Lower	Upper	z	p
Attention^a								
Bateman, 1994 ⁵	ATH composite	277	0.07	0.06	-0.05	0.19	1.2	.2300
McCann, 2007 ¹³	CPT	86	0.18	0.11	-0.04	0.39	1.63	.1000
Mattes, 1981 ¹²	Distraction task	5	0.1	0.36	-0.6	0.8	0.28	.7800
Adams, 1981 ²³	Memory tasks	18	0.01	0.23	-0.44	0.45	0.02	.9800
Goyette, 1978 ⁸	ZITA/ADT	16	0.32	0.24	-0.16	0.8	1.3	.1900
Conners, 1980 ⁶	PALT	9	0.8	0.36	0.1	1.5	2.25	.0200
Swanson, 1981 ¹⁸	PALT	20	0.59	0.23	0.13	1.05	2.52	.0100
Motor								
Adams, 1981 ²³	Motor battery	18						
Adams, 1981 ²³	Balance (hop)	18	0.54	0.24	0.06	1.01	2.22	.0300
Adams, 1981 ²³	Coordination (ball)	18	-0.3	0.16	-0.62	0.02	-1.86	.0600
Adams, 1981 ²³	Fine motor (draw)	18	-0.03	0.13	-0.28	0.23	-0.2	.8400
Rapp, 1978 ¹	Ayres motor	24	0	0.2	-0.39	0.39	0	1.0000
Language								
Adams, 1981 ²³	Picture Vocabul	18	0.40	0.24	-0.06	0.87	1.72	.0852
Adams, 1981 ²³	Handwriting		-0.14	0.23	-0.58	0.30	-0.62	.5358
Physiology^b								
Conners, 1980 ⁶	Actigraph foot	9	0.90	0.37	0.18	1.63	2.45	.0143
Conners, 1980 ⁶	Actigraph hand	9	0.06	0.30	-0.53	0.65	0.20	.8445
Physiology								
Restriction diet								
Uhlig, 1997 ²⁰	EEG Central	12	0.37	0.28	-0.18	0.91	1.31	.1908
Uhlig, 1997 ²⁰	EEG Frontal	12	0.71	0.31	0.11	1.31	2.31	.0207
Uhlig, 1997 ²⁰	EEG Posterior	12	-0.21	0.14	-0.49	0.07	-1.45	.1467
Uhlig, 1997 ²⁰	EEG Temporal	12	0.43	0.30	-0.15	1.01	1.45	.1478
Summary Food colors								
Motor overall		42	0.02	0.15	-0.27	0.32	0.15	.8800
PALT overall		29	0.65	0.2	0.27	1.04	3.35	.0008
Attention pooled		154	0.27	0.10	0.07	0.47	2.67	.0074
Attention not memory		136	0.32	0.11	0.11	0.54	2.90	.0037
Language overall		18	0.13	0.23	-0.32	0.58	0.57	.5668
Study	Task(s)	g Removed	SE	Lower	Upper	z	p	
One-Study-Removed Analysis for Attention Tasks								
Attention Studies (All)								
Adams, 1981 ²³	Combined	0.323	0.111	0.105	0.541	2.902	.0037	
Conners et al, 1980 ⁶	PALT	0.217	0.082	0.056	0.378	2.637	.0084	
Goyette et al, 1978 ⁸	ZITA/ADT	0.279	0.127	0.030	0.528	2.196	.0281	
Mattes and Gittleman, 1981 ¹²	distraction task	0.298	0.118	0.068	0.529	2.535	.0112	
McCann et al, 2007 ¹³	Combined	0.340	0.142	0.061	0.618	2.389	.0169	
Swanson and Kinsbourne, 1981 ¹⁸	PALT	0.201	0.085	0.034	0.368	2.357	.0184	
Overall		0.272	0.102	0.073	0.471	2.677	.0074	

TABLE S4 Continued

Study	Task(s)	g Removed	SE	Lower	Upper	z	p
Attention Studies (Colors Only)							
Adams, 1981 ²³	Combined	0.458	0.140	0.183	0.733	3.262	.0011
Conners, et al., 1980 ⁶	PALT	0.271	0.138	0.001	0.542	1.965	.0495
Goyette et al., 1978a ⁸	ZITA/ADT	0.354	0.191	-0.022	0.729	1.848	.0647
Mattes and Gittleman, 1981 ¹²	distraction task	0.382	0.166	0.056	0.707	2.296	.0217
Swanson and Kinsbourne, 1981 ¹⁸	PALT	0.255	0.160	-0.058	0.569	1.598	.1101
Overall		0.340	0.142	0.061	0.618	2.389	.0169

Note: ADT = Auditory Discrimination Test; ATH = aggregated test hyperactivity; CPT = Continuous Performance Test; EEG = electroencephalography; PALT = Paired Associated Learning Test; SE = standard error; ZITA = Zero Input Tracking Analyzer.

^aOne-study-removed analysis for attention tasks.

^bActigraphy effects estimated from cell means in authors' table may be over- or underestimated.

TABLE S5 Studies of Children Who Were Preselected to Be Diet Responsive Versus Those Not Preselected

	Studies	N	g	CI	p	Q	p	τ	I ²
Parent									
All studies									
Not preselected	5	593	.199	.09-.31	.001	6.0	.20	.07	33%
Selected	15	221	.192	.02-.36	.023	36.4	.001	.24	61%
Difference between groups: $p = .95$ (NS)									
High-quality studies									
Not preselected	3	515	.15	.07-.24	.001	3.6	.16	.08	45%
Selected	8	123	.11	.01-.22	.041	23.8	.001	.26	71%
Difference between groups: $p = .73$ (NS)									
Teacher									
All studies									
Not preselected	4	267	.21	-.04-.46	.09	7.7	.052	.194	61%
Selected	6	78	.06	-.12-.24	.52	0.9	.96	.00	.00%
Difference between groups: $p = .33$ (NS)									
High-quality studies									
Not preselected	3	257	.27	-.05-.60	.09	7.4	.02	.23	73%
Selected	5	50	.12	-.10-.35	.29	0.6	.97	.00	.00%
Difference between groups: $p = .45$ (NS)									
Tests of Attention									
Not preselected	2	106	.33	-.06-.72	.09	2.6	.11	.23	61%
Preselected	5	67	.22	.05-.45	.01	3.9	.41	.00	.00%
Difference between groups: $p = .62$ (NS)									

Note: Two studies selected children to be hyperactive, but did not further select for diet sensitivity before giving the food color challenge. Williams et al (1978)²¹ gave a challenge to 26 children with high hyperactivity scores and obtained parent and observer ratings. Parent ratings yielded $g = 0.56$; teacher ratings yielded $g = 0.51$. Swanson and Kinsbourne (1981)¹⁸ conducted their challenge and response on the Partially Accelerated Life Test with children who were hyperactive but not previously given a dietary test, and saw a difference in response between hyperactive and nonhyperactive youngsters, although the pattern was difficult to interpret. CI = confidence interval; NS = not significant.

SUPPLEMENTARY REFERENCES

1. Rapp DJ. Does diet affect hyperactivity? *J Learn Disabil.* 1978;11:383-389.
2. David TJ. Reactions to dietary tartrazine. *Arch Dis Child.* 1987;62:119-122.
3. Pollock I, Warner JO. Effect of artificial food colours on childhood behaviour. *Arch Dis Child.* 1990;65:74-77.
4. Adams W. Lack of behavioral effects from Feingold diet violations. *Percept Mot Skills.* 1981;52:307-313.
5. Bateman B, Warner JO, Hutchinson E, *et al.* The effects of a double blind, placebo controlled, artificial food colourings and benzoate preservative challenge on hyperactivity in a general population sample of preschool children. *Arch Dis Child.* 2004;89:506-511.
6. Conners CK. Foods, food dyes, and allergies. In: *Food Additives and Hyperactive Children.* New York: Plenum 1980;77-85.
7. Conners CK, Goyette CH, Newman EB. Dose-time effect of artificial colors in hyperactive children. *J Learn Disabil.* 1980;13:512-516.
8. Goyette GH, Connors CK, Petti TA, Curtis LE. Effects of artificial colors on hyperkinetic children: a double-blind challenge study [proceedings]. *Psychopharmacol Bull.* 1978;14:39-40.
9. Gross MD, Tofanelli RA, Butzirus SM, Snodgrass EW. The effect of diets rich in and free from additives on the behavior of children with hyperkinetic and learning disorders. *J Am Acad Child Adolesc Psychiatry.* 1987;26:53-55.
10. Harley JP, Ray RS, Tomasi L, *et al.* Hyperkinesis and food additives: testing the Feingold hypothesis. *Pediatrics.* 1978;61:818-828.
11. Levy F, Hobbes G. Hyperkinesis and diet: a replication study. *Am J Psychiatry.* 1978;135:1559-1560.
12. Mattes JA, Gittelman R. Effects of artificial food colorings in children with hyperactive symptoms. A critical review and results of a controlled study. *Arch Gen Psychiatry.* 1981;38:714-718.
13. McCann D, Barrett A, Cooper A, *et al.* Food additives and hyperactive behaviour in 3-year-old and 8/9-year-old children in the community: a randomised, double-blinded, placebo-controlled trial. *Lancet.* 2007;370:1560-1567.
14. Rowe KS. Synthetic food colourings and 'hyperactivity': a double-blind crossover study. *Aust Paediatr J.* Apr 1988;24:143-147.
15. Rowe KS, Rowe KJ. Synthetic food coloring and behavior: a dose response effect in a double-blind, placebo-controlled, repeated-measures study. *J Pediatr.* 1994;125:691-698.
16. Sarantinos J, Rowe KS, Briggs DR. Synthetic food colouring and behavioural change in children with attention deficit disorder: a double-blind, placebo-controlled, repeated measures study. *Proc Nutr Soc Australia.* 1990;15:233.
17. Spring C, Vermeersch J, Blunden D, Sterling H. Case studies of effects of artificial food colors on hyperactivity. *J Spec Educ* 1981;15:361-372.
18. Swanson JM, Kinsbourne M. Food dyes impair performance of hyperactive children on a laboratory learning test. *Science.* 1980;207:1485-1487.
19. Thorley G. Pilot study to assess behavioural and cognitive effects of artificial food colours in a group of retarded children. *Dev Med Child Neurol.* 1984;26:56-61.
20. Uhlig T, Merckenschlager A, Brandmaier R, Egger J. Topographic mapping of brain electrical activity in children with food-induced attention deficit hyperkinetic disorder. *Eur J Pediatr.* 1997;156:557-561.
21. Williams JI, Cram DM, Tausig FT, Webster E. Relative effects of drugs and diet on hyperactive behaviors: an experimental study. *Pediatrics.* 1978;61:811-817.
22. Wilson N, Scott A. A double-blind assessment of additive intolerance in children using a 12 day challenge period at home. *Clin Exp Allergy.* 1989;19:267-272.
23. Adams W. Lack of behavioral effects from Feingold diet violations. *Percept Mot Skills.* Feb 1981;52:307-313.
24. Weiss B, Williams JH, Margen S, *et al.* Behavioral responses to artificial food colors. *Science.* 1980;207:1487-1489.