Synthetic Food Colors and Hyperactivity in Children: A Double-Blind Challenge Experiment

J. Preston Harley, Ph.D., Charles G. Matthews, Ph.D., and Peter Eichman, M.D.

From the Department of Neurology, University of Wisconsin, Madison

ABSTRACT. Nine hyperactive male subjects, selected on the basis of showing a favorable "response" to the Feingold diet in an earlier study, were maintained on a strict elimination (Feingold) diet for 11 weeks, and were given multiple trials of placebo and challenge food materials. Parental and teacher ratings, classroom behavior observations, and neuropsychological test scores obtained during baseline, placebo, and challenge conditions, in general, were not found to be adversely affected by the artificial color challenge materials. As expected, comparable data gathered on a matched control group showed them to receive substantially better ratings than the hyperactive subjects on the majority of the comparison measures employed. Possible explanations for the discrepancy between the dramatic clinical-anecdotal reports that have been given and the much more equivocal findings from formal experimental projects are presented. Pediatrics 62:975-983, 1978, hyperactivity, Feingold hypothesis, food additives, diet and behavior.

A possible link between the ingestion of foods containing synthetic food flavorings and colors and the presence of hyperactive behaviors in children has been suggested by Feingold. He has advocated an elimination diet which removes foods containing salicylates and artificial food colors and flavors as a treatment modality for hyperactive children. According to Feingold, approximately 50% of the hyperactive children demonstrate positive and often quite dramatic behavioral improvement after being placed on the elimination diet. Several reviews and current status reports related to the Feingold diet have been published recently that delineate the complex methodological and data interpretation issues to be resolved before Dr. Feingold's assertions can be scientifically supported or refuted.

The present investigation represents phase 2 of an earlier study reported in Pediatrics that was concerned with determining whether hyperactive children benefit from an elimination (Feingold) diet regimen. The goal of the phase 2 study was to select those individual children who showed the "best" response to diet manipulation in the phase 1 investigation, and then to repeatedly challenge them with specified amounts of synthetic food colors during a nine-week period in which they were otherwise maintained on the elimination diet.

METHODS

Subject Selection

The criteria for subject selection for the initial study (phase 1) has been described previously. Selection of subjects for the present experiment (phase 2) was made by rank ordering the 46 male subjects from the phase 1 study on their behavioral ratings and classroom observations in terms of a favorable response to the elimination diet. The parent and teacher ratings, classroom behaviors, and laboratory observations were assigned equal weights in computing a composite diet response index for each child. The nine subjects selected for this study were those receiving the highest rank order from these combined sources and available for participation. In all instances, these subjects exceeded the 50th percentile of the total rank order of the original phase 1 sample in terms of a favorable response to the elimination diet. A control subject for each hyperactive subject was identified by the classroom teacher and matched with the hyperactive subject on sex, grade, and academic ability. The control children were selected on the basis of being "average" rather than "model" or "excellent" students. The mean ages for the hyperactive and control groups were 111 (SD = 21) and 113 months (SD = 21), respectively. None of the subjects were receiving medication for their behavior problem immediately prior to the study.

Received April 7; accepted for publication May 10, 1978.
ADDRESS FOR REPRINTS: (J.P.H.) Neuropsychology Laboratory, Department of Neurology, Clinical Sciences Center, 600 Highland Avenue, Madison, WI 53706.
Procedure

After obtaining parental consent, each child was observed over four weeks of baseline; in the first two weeks (baseline 1), social-psychological, classroom behavior, and dietary information was gathered with the child and his family continuing on their regular diets. In the next two weeks of baseline (baseline 2), the entire family was placed on the elimination diet, and this diet was strictly maintained throughout the duration of the study. All foods were provided for the entire family and specially prepared treats were delivered to the school when the experimental child or any of his classmates had a birthday or wished to bring treats to school. Following the four-week baseline interval, the nine experimental subjects were administered challenge and placebo material with multiple cross-overs during the experimental phase (nine weeks) of the project. The observers, parents, teachers, and project staff members did not know the placebo-challenge code. A week’s supply of cookies or candy bars, two items per day, were labeled for each day of the week and were included in the weekly groceries supplied to the family. Five of the subjects received the following sequence after baseline: placebo material for a two-week period, followed by two weeks of challenge items, then another two weeks of placebo material, terminating with three weeks of challenge materials. The other four experimental subjects received the opposite sequence after baseline, i.e., two weeks of challenge, followed by two weeks of placebo, then another two weeks of challenge, concluding with three weeks of placebo.

Placebo-Challenge Materials

The placebo and challenge food were in two forms: candy bars and cookies. Each food item contained a blend of half the “average” daily intake of 27 mg of the certified food colors. These figures were derived from national consumption data. The placebo items contained no synthetic food colors. Cookies and candy bars were alternately provided each week.

The placebo and challenge materials were developed under the auspices of the Nutrition Foundation. A group of eight experienced judges evaluated the placebo and challenge forms of the candy bars and cookies. Twenty-four separate judgments were made in each of the three separate evaluations of the unlabeled products. Major attributes of the candy bars judged included the following: appearance, aroma, flavor, and texture. The cookies were rated on frosting appearance, cookie appearance, aroma, flavor, and texture.

The results of this testing conducted by the Quaker Research Laboratories indicated that the chocolate challenge cookies were not substantially different from the placebo on any of the attribute scales measured by the panel. The challenge and placebo chocolate candy bars were found to be different on several attributes; however, these findings were described as “fine differences perceived by a panel with repeated exposure to the candy bars” and more attributable to manufacturing procedures than to the food colors, per se. A nonexpert adult taste panel was not able to distinguish any differences in the placebo vs. challenge materials in either the cookie or candy bar materials. Postexperimental interviews revealed that none of the parents and/or children in the phase 2 study correctly identified the placebo vs. challenge materials during the investigation.

Placebo-Challenge Code

The specially prepared cookies and candy bars were labeled with four digit codes designated for each subject. A series of sealed envelopes, each containing the weekly code for each subject, was kept in the Department of Neurology in order to have access to the code assignment if it became necessary for medical or behavioral reasons to identify the type of materials being ingested by a subject. It was not necessary to secure the identity of the code for any subject during the study.

A supply of placebo materials was retained in the laboratory. Thus, if a child had a severe adverse reaction, his regularly scheduled materials (placebo or challenge) could be replaced with known placebos for the remainder of the week, in this way taking corrective steps but avoiding code disclosure. One child showing extreme behavioral disruption during the experimental phase of the project was placed on a regimen of the known placebo cookie for one week, and at the parents’ request his preexperimental medications schedule was reinstated for the final seven weeks of the study. Postexperimental breaking of the code revealed that this child had also been receiving the placebo cookie during the period of the disruptive behavior that prompted this “emergency” maneuver on the part of the experimenters. A second child had his medication schedule reinstated at week 9 of the 13-week study and received known placebo materials during this final phase of the study. All subsequent data analyses were adjusted to reflect these procedural changes.

Dietary Compliance

Daily dietary records were maintained by the parents of the experimental subjects throughout
the 11-week diet phase of the study to help assess maintenance of dietary compliance. Two subjects had no reported dietary deviations. The largest number of dietary infractions for the entire 11-week period (while being maintained on the elimination diet) for an individual was six. These data suggest that comparison of placebo and challenge periods was not seriously jeopardized by an unacceptable number of dietary infractions.

**Dependent Variables**

The Conners’ ten-item parent-teacher questionnaire (P-TQ) was independently completed by the subject’s mother, father, and teacher on each Tuesday and Thursday for the 13 consecutive weeks.

In addition to subjective parent and teacher ratings, behaviors in the classroom setting were recorded by trained observers. The experimental subjects and their matched, nonhyperactive control subjects were observed each Tuesday and Thursday during the baseline and experimental periods, i.e., a total of 26 consecutive observations.

A battery of neuropsychological measures was administered on four separate occasions: at the end of baseline (week 4); following completion of the first two-week experimental conditions (weeks 6 and 8); and at the conclusion of the study (week 13). The neuropsychological dependent variables have been described previously.

**Classroom Observations**

The behavior of the nine hyperactive subjects was recorded in the classroom situation using a procedure similar to the method described by Werry and Quay. Observations were made during designated academic seat work by the classroom observation coordinator and by a second observer. As previously noted, a control subject was identified for each hyperactive subject, matched on sex, grade, and academic ability by the teacher. The identities of the hyperactive and control subjects were not revealed to the observers, the observers having been instructed only to observe “these two children.” The observers were trained via video-tape classroom training sessions. A predetermined level of agreement was reached by all observers (75% agreement) prior to initiation of data collection.

The observer was positioned close enough to both children to allow the observer to hear clearly what was being said and to see what the children were doing at their desks. Each child was observed for a 20-second interval. A ten-second period followed for the recording of the appropriate symbols on the score sheet. Behaviors occurring during this ten-second period were not recorded. Thus, two observations were made per minute. The second child was observed for the next minute. There are 60 separate cells of observation, giving a total observing time of 30 minutes per child for each of the two weekly sessions.

**Definition of Behavioral Categories**

(No) Deviant Behavior. Deviant behavior was defined as any behavior that violates any explicit or implicit rule by which the class and teacher operate. Thereafter, it was necessary to determine the rules in a given classroom before any observations were undertaken. The observer questioned the teacher as to the conditions under which it was permissible for a child to leave his seat or to speak. A count was made of the number of observational intervals without the occurrence of deviant behavior; hence the designation of the category as (no) deviant behavior.

Gross Motor Activity. This was defined as any repetitive movements of the trunk, arms, or legs including walking, running, hopping, skipping, jumping, and rocking in seat. Also, gross physical movements such as arm flailing, bouncing in seat, leg swinging, rocking, and head swaying were included. Foot movements and hand movements alone were not counted.

Nonwork. Nonwork was defined as activity that the child engages in for self-amusement, -entertainment, or -stimulation. To be scored as nonwork, the activity must include, although not be limited to, the child’s use of his hands to manipulate his own or community property so that such behavior is incompatible with learning. Examples are as follows: pencil play, rolling pencil on desk, waving, hitting on desk, marking with pencil, tearing paper, crumpling paper, hand movements near face, waving things near face, fiddling with clothes, scribbling on paper, patting body, daydreaming, staring, etc.

Disturbing Behavior. This was defined as any physical contact initiated or reciprocated by the child under observation with another person independent of the intent of the child (aggressive, affectional, or other), plus vocalizations or other voluntary respiratory noises such as whistling or grunting that are not task related. Examples are as follows: calling the teacher without raising hand, talking to others without permission, swearing, and any noise that seems produced for the sake of noise.

Isolation. This category is scored if the child has been sent out of the room as a punishment or
has been placed in a corner of the room. If the child is isolated in the classroom, other deviant behaviors that can be noted such as vocalizations and other noises continue to be recorded.

**On- and Off-Task Activity.** This is a measure of the child’s attention to designated task material. Attending was defined as the eyes being directed to the task at hand or to the teacher for a period of not less than 15 of the 20-second observation period. The child is also considered to be on task when he can be clearly seen to be doing one task even though his eyes may be off his work, e.g., counting on his fingers, verbally working out math problems, etc. Hand raising is considered on-task behavior. A child was considered to be off task when he was engaged in some deviant behavior, as previously defined.

**Scoring.** The number of observations for each category was summed for each half-hour scoring period and recorded for the two children in the classroom.

**Reliability.** A ratio agreement was calculated between each of the nine observers (one for each hyperactive and his respective classroom control subject) and the chief classroom observer who visited each classroom and completed observational ratings concurrently with the second observer. The mean percent of agreement for all rating categories combined was 82.3 for the first reliability check and 81.2 for the second check.

### RESULTS

**Parent-Teacher Questionnaire (P-TQ) Ratings**

**Baseline.** The mean behavioral ratings made during the first two weeks of baseline on the hyperactive and control groups (during which time the families of the hyperactive children continued on their regular diet regimen) were compared. The hyperactive subjects were rated as significantly more hyperactive than were the control subjects by mothers \((t = 5.67, df = 16, P < .001)\), fathers \((t = 5.95, df = 14, P < .001)\), and teachers \((t = 3.27, df = 16, P < .01)\). Similar results were found when the mean P-TQ ratings made by the mothers \((t = 3.59, df = 16, P < .01)\), fathers \((t = 5.08, df = 14, P < .001)\), and teachers \((t = 4.12, df = 16, P < .001)\) were compared for the hyperactive and control groups during the last two weeks of baseline. During these latter two weeks of baseline, families of the hyperactive subjects were provided with all foods needed for the strict maintenance of the elimination diet.

**Challenge vs. Placebo Conditions.** As shown in the Table, mean P-TQ ratings for the hyperactive children given by mothers, fathers, and teachers under the combined placebo conditions were not found to be significantly different from those made during the combined challenge periods.

The mean P-TQ rating made during all placebo periods for the hyperactive subjects was compared to the control group rating for the same time periods. The behavior of the hyperactive group was judged to be significantly more disruptive than the control group by the mothers \((t = 4.18, df = 16, P < .001)\), fathers \((t = 5.01, df = 14, P < .001)\), and teachers \((t = 3.20, df = 16, P < .01)\). Similarly, the P-TQ ratings made by the mothers \((t = 7.83, df = 15, P < .001)\), fathers \((t = 5.47, df = 14, P < .001)\), and teachers \((t = 3.33, df = 15, P < .01)\) for the combined challenge periods indicated the presence of more severe “hyperactive” behaviors in the hyperactive than in the control group.

The mean P-TQ ratings for subjects 1 to 5 (receiving the placebo-challenge-placebo-challenge sequence) by mother, father, and teacher are shown in Figure 1. Inspection of Figure 1 does not reveal a striking challenge effect, and the ratings that seem most in accord with the hypothesized challenge condition are in large part supplied by the mothers. Inspection of Figure 2 is even less suggestive of any marked or consistent challenge vs. placebo effect upon behavioral ratings for subjects 6 to 9 who received the opposite experimental sequence.

Figure 3 shows the consistently low ratings given to the nine nonhyperactive control children by mothers, fathers, and teachers over the 13 weeks of the experiment, their average rating being one third or less of that received by the hyperactive sample. This difference between the hyperactive and control subjects demonstrates the sensitivity of the Conners P-TQ rating scale to the identification of the behaviors under study.

### Classroom Behavior

For purposes of concise presentation of the multiple behavioral categories observed in the
classroom, several of the individual categories (i.e., gross motor activity, nonwork, disturbing behavior, and off-task behavior) were collapsed into an overall "disruptive behavioral index" (Fig. 4). The "isolation" variable was excluded from the composite disruptive behavior index because of the very low frequency of tallied behaviors in this category; the "no deviant behavior" category ratings are presented separately in Figure 5. Detailed analysis and graphic presentation of the
separate behavioral classifications comprising the "disruptive behavior index" and of the no deviant behavior and isolation behavior classifications have been reported elsewhere.12

Baseline: Hyperactive vs. Control Group. The magnitude of the disruptive behavior index in the hyperactive group was closely comparable during the first two weeks (mean, 29.9; SD, 10.9) and the last two weeks (mean, 34; SD, 16.7) of baseline. These behaviors were more frequently recorded in the hyperactive than in the control group for the combined four weeks of baseline ($t = 2.36$, $df = 16$, $P < .05$).

Challenge vs. Placebo Condition. The mean disruptive classroom behavior index of the hyperactive group was not differentially affected by the challenge (mean, 28.4; SD, 18.4) and placebo (mean, 27.3; SD, 19.5) manipulations ($P > .05$). The indices of disruptive classroom behaviors earned by the hyperactive subjects under the placebo-challenge conditions are shown in Figure 4, as are the disruptive behavioral indices of the control group for corresponding time periods.

Figure 5 shows the frequency of no deviant behavior ratings for the hyperactive and control subjects under baseline, placebo, and challenge periods. In each instance, the control subjects showed a greater, albeit statistically nonsignificant, frequency of no-deviant behaviors than did the hyperactive subjects.
Neuropsychological Measures

No statistically significant differences were obtained between the hyperactive and control groups for 13 neuropsychological dependent variables taken at the end of the baseline period. Better baseline performance, however, was demonstrated by the control than by the hyperactive subjects on the continuous performance test\(^\text{\textsuperscript{13}}\) of attention/vigilance \((t = 2.44, \ df = 16, \ P < .05)\). The performance of the hyperactive children on the neuropsychological measures was not differentially affected by the placebo-challenge treatments.

Individual Profile Analysis

Parent-teacher ratings and classroom observer data for each individual hyperactive subject across the 13 weeks of the study are presented in graphic form in an extended report of the present investigation.\(^\text{\textsuperscript{12}}\) Only one subject displayed a behavioral profile of parental ratings and classroom observational data that even approximated the predicted on-off effect of the challenge and placebo materials. The other eight individual graphs showed no discernible relationship between challenge-placebo conditions and either the Conners P-TQ ratings or classroom observational behavior ratings.

DISCUSSION

Feingold has observed that if a child deviates from his elimination diet (i.e., one free of foods containing synthetic flavors, colors, and salicylates), ingestion of even minute amounts of the prescribed foods (e.g., a stick of gum, a cookie, or soft drink) "causes a recurrence of the complete behavioral pattern within two to four hours which persists for one to four days."\(^\text{\textsuperscript{12}}\) The overall results of the present study, in which the subjects were challenged up to 21 consecutive days with food items containing the average daily intake of artificial colors, do not provide confirmatory experimental data for these clinical observations.

Negative results on another challenge study that incorporated food containing the certified food colors have also been reported by Goyette and colleagues.\(^\text{\textsuperscript{14}}\) Those authors did not find an associated change in hyperactive symptoms related to the challenge vs. placebo treatments. A performance decrement on an attention task, however, was found shortly after the ingestion of the challenge materials suggesting some kind of transitory psychopharmacological effect. Additional work by Conners and Goyette using preschool children suggested a possible deleterious behavioral effect of food colorings in this younger age group, an observation that is in accord with the results of the phase I study by the Wisconsin investigative team.\(^\text{\textsuperscript{9}}\)

Following a review of the scientific literature relevant to the Feingold hypothesis, Spring and Sandoval\(^\text{\textsuperscript{13}}\) concluded that "public advocacy has far outstripped the research on which it should be based." The authors also challenged the preva-
lence estimates of hyperkinesis and the causal relationship between amount of food additives and increase in the incidence of hyperkinesis over the last decade inferred by Feingold, presenting evidence that did not support Feingold's impression that there has been a substantial increase in hyperactivity-learning disabilities over the last ten years rather than just an increased awareness of the disorder. Feingold acknowledges that "hyperkinesis is still not understood despite the tremendous volume of basic research and clinical observations on it" and that reported incidence figures vary widely, depending upon the source.

The recommended use of the K-P diet program, however, has not been limited to the treatment of hyperactive children. It has also been considered to be effective in the overall behavioral treatment/management of children with petit mal seizure disorders, mental retardation, and learning disabilities. For example, Feingold states the following:

In retardation the clinical response may be dramatic, as evidenced by improved behavior, better coordination of both fine and gross muscles, and improved learning ability. All of these gains induce a marked transformation in the patient whose expression becomes more alert and bright, his social adjustment improves, permitting him to function as a self-sufficient person who does not require one-to-one attention or instruction.

Such promising statements, unfortunately, have not been derived from systematic data collection or experimental studies, thus making it impossible to critically evaluate the claimed efficacy of the elimination diet in these diverse diagnostic categories. Although the results of the Wisconsin phase 1 and phase 2 experiments are in large part negative with respect to supporting the Feingold hypothesis, it is not possible from our data to categorically dismiss the possibility that a small subset of children, especially those of preschool age, may demonstrate hyperkinetic symptoms following the ingestion of synthetic food flavors and colors. In contrast to these modest and guarded conclusions from controlled investigations, Feingold has stated on the basis of clinical observation that approximately 50% of hyperactive children treated with the K-P diet program show a successful response. A general description of an expected response to the K-P diet program is as follows: "a complete change in the child may be observed within a week or two and sometimes within a few days. The child becomes quiet, docile, affectionate, and cooperative. The sleep pattern improves. Improved scholastic performance usually follows very rapidly." Anecdotal case reports of dramatic and impressive improvement in the child's behavior subsequent to being maintained on the K-P diet have been presented in detail. Based upon these clinical and personal reports, Feingold believes that it is imperative to implement the use of a logo to assist the consumer in identifying products not containing artificial food colors and flavors plus the establishment of "controlled" school lunch programs. Such action has been recommended even though Feingold acknowledges the following:

It is true that the mechanisms involved are not known, nor have the specific compounds been identified. But such basic data will require many years of well controlled research. It is not necessary to await the availability of basic data. It has been demonstrated that these children respond to dietary intervention. That is the immediate and urgent need to halt and reverse the persistent rise is scholastic failures, vandalism, delinquency and crime.

The Feingold diet clearly offers an appealing, relatively simple treatment for a very large number of children who have rightly or wrongly been labeled hyperactive. Parents are typically attracted to the diet treatment program because of dissatisfaction with other therapies that have been tried. They are frequently seeking alternatives to medication and other treatment modalities because of deeply felt negative attitudes and convictions regarding synthetic substances and positive attitudes (which are strongly supported by the current social-environmental "Zeitgeist") toward natural food products. The diet may help to reduce the parents' feelings of guilt or other negative emotions involving their hyperactive child because an "outside" causative agent has now been identified that helps to minimize any threatened experience of blame on the part of the parents. The numerous Feingold associations that have been established across the country may also contribute to a positive expectancy effect and to a collective parental identity and conviction of making common cause against exploitation by a malevolent food industry complex.

These kinds of cautious, even pedestrian, possible alternative explanations of the highly positive anecdotal clinical reports have been preempted by the ready admission that placebo and expectancy effects may indeed be operative but that what is important is the fact of improvement, regardless of the actual mechanisms involved. Nonetheless, it must be emphasized that when data are collected by means of rigorous experimental designs that minimize potential sources of subjective bias and positive behavioral outcome expectancy, diet-related behavioral changes are much less spectacular than are the impressive clinical case studies and fervent parental testimonials that have been so widely disseminated.

While little research support has been found
for these clinical reports with respect to school-age children, some data (including our own phase I study) do suggest that there may be a small subgroup of younger children who display adverse behavioral reactions to particular synthetic food substances, and hyperactive children in this age group clearly deserve further study. Firm scientific conclusions regarding this entire investigative area have not yet been reached and may not be for a considerable period of time. Nonetheless, the results of controlled experimental investigations of the Feingold hypothesis presently available appear to be at sufficient variance with the positive clinical-anecdotal reports to suggest that the unqualified advocacy of sweeping, diet-related treatment recommendations is distinctly premature at this time.

REFERENCES


ACKNOWLEDGMENT

This project was supported by Grant 133-9051 from the University of Wisconsin Food Research Institute, and Grant 527 from the Nutrition Foundation.

We wish to thank our fellow team members and colleagues, M. Collins, B. Lancaster, M-L. Q. Mason, B. Read, E. Traisman, and V. Vought for their assistance. We gratefully acknowledge the technical assistance of Ms. Miyea Cohen, University of Illinois–Champaign, in establishing the classroom observational procedures.