Reviewer: Emily S. Barrett, PhD; Associate Professor of Biostatistics and Epidemiology at the Rutgers School of Public Health and the Environmental and Occupational Health Sciences Institute, Piscataway, NJ.

11/25/20

REVIEW OF "DRAFT HEALTH EFFECTS ASSESSMENT: POTENTIAL NEUROBEHAVIORAL EFFECTS OF SYNTHETIC FOOD DYES IN CHILDREN" (AUGUST 2020)

Based on my expertise and experience, I am reviewing the findings, assumptions, or conclusions I agreed I could review with confidence: Conclusion 1, and to the extent possible, Conclusion 3. Taken as a whole and to the best of my knowledge as an epidemiologist who studies chemical exposures and their impact on children's health and development, I believe this proposal to be based upon sound scientific knowledge, methods, and practices. Below, I highlight the factors that led me to that conclusion and identify particular strengths (and to a lesser extent, shortcomings) of the current report.

Conclusion 1

Conclusion 1 states: "After reviewing the epidemiological literature on the neurobehavioral effects of synthetic food dyes, OEHHA concludes that the data suggest an effect of artificial food dyes on children's neurobehavior."

The first section of the report consists of a systematic review of the scientific literature on seven synthetic food dyes that are approved for use by the US Food and Drug Administration (US-FDA) and are in commonly found in foods, beverages, over the counter medications, and vitamins: FD&C Blue No. 1, Blue No. 2, Green No. 3, Red No. 3, Red No. 40, Yellow No. 5, and Yellow No. 6. Based on these results, Conclusion 1 indicates that: (1) there is solid evidence that synthetic food dyes are associated with neurobehavioral measures (e.g. inattentiveness, hyperactivity, and restlessness) in children; and (2) some children may be particularly vulnerable to neurobehavioral outcomes following food dye consumption. However, it is also noted that the literature is variable, with associations observed in some studies but not others. These conclusions are consistent with the result of a 2012 meta-analysis on this topic as well (Nigg et al. 2012).

The process for undertaking the systematic review was overall sound and thoroughly described, though the inclusion of a "PECO" statement, a common element of systematic reviews, would have been useful. The reviewers used appropriate steps to identify publications of interest including searches in several of the largest biomedical literature databases as well as government reports. They justifiably chose to focus the systematic review on the results of clinical trials on this topic as they are considered the gold standard for strength of epidemiological evidence. Importantly, all of the studies reviewed employed a crossover design such that participants acted as their own controls, reducing potential confounding by relatively stable factors like socioeconomic

status. The strength of this design and applicability to this particular research question are explained well (for instance on p. 44) and the search strategy was well documented including the specific key words used in the search process, facilitating future replication (Section 1.3). Ultimately 27 studies were identified that met inclusion criteria for the review, and I am not aware of any additional studies that should have been included. The only lack of clarity noted in the inclusion criteria was #4 (p. 30) regarding "a neurobehavioral outcome related to hyperactivity or inattention was assessed"; a more comprehensive list of outcomes potentially "related to" hyperactivity and inattention would be preferable.

The quality of the 27 included studies was evaluated through a list of key factors to consider and a simple scoring system (Section 2.4). The list of included factors (2.4.1) is quite comprehensive, however it might have been useful to work within the framework of an existing Risk of Bias (RoB) tool intended for epidemiological studies. Such RoB tools (including, but not limited to, the Office of Health Assessment and Translation [OHAT] tool, Program on Reproductive Health and the Environment [PRHE]'s Navigation Guide, and the Integrated Risk Information System [IRIS] Tool) are specifically designed to assess internal validity by evaluating the extent to which elements of study design and conduct may have influenced results. While many such factors are captured in the scoring system devised by the authors, starting with and adapting an existing RoB tool might have added to the rigor of the systematic review. Alternatively, if extant RoB tools were considered but ultimately not used, it would have been helpful to explain the choice to instead create a new scoring system. Nevertheless, it is important to note that the factors used in study quality assessments (2.4.1) largely overlap with domains covered by RoB tools, thus the decision not to use an extant RoB tool is considered only a minor limitation and does not detract from the conclusions of the report. Overall Section 2.4 is an excellent summary of the decision making process around inclusion and exclusion of individual studies as well as the study elements that were then abstracted. In particular, I would like to note Section 2.4.3.9 in which the authors discuss consideration of magnitude of association as well as statistical significance as important evidence of causation. This is particularly important given the very small size of many of the studies considered, which may have been underpowered to detect effects. In fact, this reviewer questions the value of including extremely small studies (such as those with n=1), however this concern is ameliorated by the greater attention to and discussion of the larger and more rigorous studies.

The report is quite comprehensive in its data extraction and summaries. Tables 2.1-2.3 are helpful in ensuring transparency regarding excluded papers as well as data extraction and coding relevant to the 27 included papers. The overall approach utilized to select studies for inclusion and assessment of study quality was methodologically sound, however there was a lack of clarity on several minor points in Section 2.4.1, which explains the factors used to assess study quality. Clarifications needed include:

- On what basis was ≥50 mg/day used as a cutoff for a "high" dose?

- What constitutes an “adequate” washout period?

The Results section (2.6) is comprehensive and thoughtfully written, with consideration of a number of factors that might explain disparate results across studies including age of the study and the source of behavioral data (e.g. parent or teacher report, direct observation). However, it was somewhat surprising that differences in results across studies were not examined in relation to other factors, such as neurodevelopmental domain. While the studies focused on outcomes “related to attention”, some more granularity could be useful (for instance distinguishing between studies examining memory vs. activity). This was somewhat ameliorated by the recent Nigg et al. (2012) meta-analysis, in which neuropsychologists identified studies using tasks that specifically and directly measured attention; importantly, the effect size was stronger when including only those studies with that specific outcome.

Similarly, there was little consideration of whether results might vary based on the particular food dye used in the challenge, possibly because many studies used a mixture of several dyes making it hard to distinguish between their relative impacts. This omission may have been due to the paucity of studies examining a single, clear food dye exposure, as explained elsewhere in the report. Finally, the considerable differences in timing of exposure (as well as age at exposure) and latency until outcome measurement may contribute to inconsistent findings. Direct comparisons of studies with very similar designs (such as the Lok et al 2013 vs McCann 2007 comparison on pp. 43-44) are useful in parsing disparate results and could be employed more extensively in the report.

Despite these minor limitations, the Conclusion 1 remains well-supported, with the majority of studies reporting some evidence of association between food dye exposure and adverse neurobehavioral outcomes, despite differences in design elements, populations studied, and quality of research. Importantly, several of the more recent studies (which are among the highest quality and largest studies, including McCann et al 2007 and Bateman et al 2004) reported associations and went on to identify polymorphisms in histamine degradation genes that may underlie susceptibility to the adverse behavioral impacts of food dyes. The report appropriately highlights the results of these studies in multiple sections as they are among the most rigorous studies on the topic.

Several important elements of the current review that represent an advance beyond prior reviews (by the FDA and others) should be noted with regard to Conclusion 1. First, although prior evaluations focused particularly on the potential associations between food dyes and hyperactivity in children, in the current review, the committee also considers additional behavioral outcomes of interest. Second, recognizing that all children may be at risk, the committee evaluated studies in the general population as well as children with neurodevelopmental or behavioral disorders. Finally, although this external reviewer will not evaluate Conclusion 2, it is important to note that the

committee conducted an extensive review of the relevant animal toxicology literature that far exceeded prior reviews by the FDA.

In addition, this review points out several important limitations of the current epidemiological research in this area:

1) The majority of studies on this topic are quite old, which presents some issues. For instance, only two studies reported disclosures and source of funding, which is now common practice. There is potential for inherent conflicts of interest in industry funded research on this topic.
2) Similarly, a number of the studies were quite small. Of the 27 included in this analysis, 21 had samples sizes under 30 children, many of them less than 10 children. Although the report does a good job of considering both significant results and large effect sizes, there is a clear need for future work that is adequately powered.
3) There was considerable variation in the age of the children studied, and overall, there was some indication that effect sizes might be larger in younger children (e.g. preschool age) suggesting a need for additional study in this potentially vulnerable age group.
4) Most of the 27 included studies considered the combined effects of multiple food dyes, making it difficult to pinpoint which one or ones might be most strongly associated with behavioral issues. Additionally in some studies, another "agent" such as benzoic acid was used, potentially obscuring the true impact of the food dyes themselves (though importantly associations between food dye consumption and adverse behavior were reported in a number of studies that did not include such agents). Results of several studies of Yellow No. 5 alone (summarized in Table 7.10) suggest the need to conduct and compare studies of single food dyes to better identify those that might impact neurobehavioral outcomes.
5) There was a lack of blinding in many studies, which impact the child's own behavior as well as parental or researcher reports. Moving forward, direct observation by a psychologist who is blinded to the study arm (treatment vs placebo) would be the gold standard for outcomes measurement in this area.
6) Timing between exposure and outcome assessment was quite variable (and in some cases unclear) and there is a lack of clarity as to whether there may potentially chronic or long-lasting impacts of food dye exposure (particularly during sensitive developmental periods) on child neurobehavioral outcomes, as opposed to strictly adult impacts. While animal evidence suggests transient impacts, timing and type of exposure (acute vs chronic) clearly needs additional consideration in humans.
7)

I would also add, though it was not explicitly noted in the report, that given increasing evidence that chemical exposures may impact neurodevelopmental outcomes differently in males and females, sex differences in response to food dyes should be considered in future work. This hypothesis of potential sex differences in response to food dye exposure is further supported by some of the animal studies reviewed in Conclusion 2 (e.g. Tanaka et al 2001).

Regarding publication bias (discussed in 2.7.7), I concur with the reviewers that it is unlikely that publication bias would significantly skew the overall conclusions from this body of literature. While it is possible that some smaller studies with null or unexpected findings might not have been published, one would imagine that would be less of an issue with larger, well-designed trials. One possible exception would be the potential for large industry-sponsored trials showing associations between food dyes and problem behaviors being left unpublished. The addition of those studies, however, would only strengthen the overall body of evidence linking this exposure and outcome.

In summary, this reviewer affirms the quality of the systematic review of the epidemiologic literature, the results of which support Conclusion 1.

Conclusion 3

Conclusion 3 states: "Our estimates of exposure indicate widespread exposure to artificial food dyes in children, that children are exposed to larger amount per body weight than women, and that the highest exposures were from over-the-counter medications in a single day."

Conclusion 3, which evaluates children's level exposure to food dyes, is based on studies measuring food dye levels in foodstuffs, medications, and vitamins considered in concert with NHANES data on food consumption in children. It further examines exposure by demographic characteristics including poverty level, race/ethnicity, and maternal education. Based on the evidence presented, this reviewer concurs with the report's authors regarding Conclusion 3, namely that intake of synthetic food dyes is likely to be higher among children than adults and comes from disparate food sources including beverages, breakfast cereals, and desserts as well as from over-the-counter (OTC) medications and vitamins. Importantly, in novel analyses performed for this report, OTC medications were estimated to result in acute exposures that could exceed the FDA and Joint FAO/WHO Expert Committee on Food Additives (JEFCA) Acceptable Daily Intakes (ADI) even when used as recommended.

Conclusion 3 is supported by evidence from a variety of sources. Six recent studies have examined exposure to food dyes in U.S. and Canadian food stuffs either through: (a) dietary logs combined with ingredient lists or manufacturer information; or (2) direct chemical analysis of food items. Methods varied quite considerably across the six cited studies making it difficult to directly compare them, however in general, food dyes were commonly found in children's diets (or foods commonly consumed by children) and were particularly prevalent in certain food groups (e.g. fruit snacks, juices and soft drinks, candy). Of greatest relevance for estimating exposure in the general U.S. population are studies (e.g. Bastaki et al 2017) linking NHANES dietary data to estimated food dye content in those foodstuffs.

To complement and extend existing work, the authors chose to conduct an additional novel analysis for this report, which was well-justified and important for several reasons: (1) there are few population based studies on exposure to food dyes; (2) most exposure data are old and may not reflect current exposures among American children; (3) prior research didn't include additional potentially vulnerable populations like pregnant women; (4) prior research did not sufficiently consider additional sources of food dyes such as vitamins and medications. Novel chemical analyses conducted in a U.C. Davis laboratory in preparation for this report measured FD&C batch-certified food dye exposures in over the counter medications and vitamins. To my knowledge, this novel work is not yet peer-reviewed, and thus has not gone higher scrutiny by independent exposure scientists; nevertheless it is this reviewer's opinion that the new analyses greatly strengthen the overall conclusion due to the significant gaps in the prior literature.

In the new analysis, the researchers linked 2015-2016 NHANES demographic and 2-day dietary recall data (focusing on pregnant women, non-pregnant women of reproductive age, and children by age group) and food dye concentrations measured by the US FDA (Doell et al 2016, Harp et al 2013), to estimate food dye consumption (in mg/kg body weight/day) among NHANES participants using both typical-exposure and high-exposure scenarios. The estimates suggested the highest exposure occurred for FD&C Red No. 40 in children 9-16, 16-18, and pregnant women, with food dye consumption generally highest in children age 5-18, though for some dyes, like Blue No. 1 and Blue No. 2, estimates were highest for children ages 0-9). It should be noted that within each age group, only a fraction of NHANES participants actually consumed foods containing a particular food dye. For instance, among the 186 children under age 2, 108 (58%) consumed a food item containing Blue No. 2, while only 17 (9%) consumed a food item containing Green No. 3, thus for some groups and dyes, estimates were based on very small sample sizes. This was most notable for Green No. 3, which was consumed least frequently. The primary dietary sources of food dye exposure varied by dye and age group. For example, among the youngest children (0-<2), white icing was the predominant source of Blue No 1, whereas in older children, ice cream cones and soft drinks were more common sources. The food dye with highest exposure, Red No. 40, was most frequently consumed in fruit juice in children under 5 and in soft drinks in children 5-16.

In unadjusted analyses, total food dye consumption was weakly inversely correlated with higher income and income/poverty ratio and was highest in Non-Hispanic Black participants. Among adult women, food dye intake was higher in women with a high school degree (or GED) or less, compared to women with higher levels of education. While these results are interesting and may be a first step towards identifying populations that may typically have higher food dye exposures, I would consider these results preliminary and hypothesis-generating, rather than definitive given that no multivariable modeling was conducted. The discussion of these results in the report is tempered and appropriate.

As a next step toward risk characterization, the report compares FDA food dye intake under both typical-exposure and high-exposure scenarios (based again on NHANES dietary data) in relation to the US FDA and JECFA ADIs, with a Hazard index >1 indicating food dye exposure estimates (in mg/kg/day) exceeding the ADI (without contributions from medication or vitamins). Under both the typical-exposure and high-exposure scenarios, hazard ratios exceeded 1 for FD&C Red No. 3 among multiple age groups (children and pregnant women) and for both mean and 95% percentile exposure estimates (pp. 206-261). Estimates were typically highest for the youngest age group, children 0-<2 years. By contrast, hazard indexes were below 1 for the other food dyes under consideration.

With the addition of the novel food dye intake data from over the counter medications and vitamins (which had not been previously studied in this context), a second comparison to ADIs was made (p. 269). Notably, this set of comparisons did not include dietary intake of food dyes and thus would be an underestimate of typical total food dye intake. For certain brands of cold, cough, and allergy medicines intended for children, recommended use (based on the label) would result in Hazard indices for Red No. 40 or Blue No. 1 greater than 1 in children 6<12 and 12-16 (without any consideration of diet). Intake of other dyes in medication and through vitamins, by contrast, was estimated to be low. While use of these medications is likely to be intermittent for most children, there may be a subset who chronically use allergy medications with food dyes (potentially up to several times a day per instruction labels) and therefore may be particularly at risk of adverse neurobehavioral outcomes.

Finally the results of novel testing of food stuffs for food dye content at UC Davis further resulted in Hazard indices greater than 1 for some age groups based on consumption of a single serving of certain food items (or half a serving for children under age 2). Results were particularly notable for FD&C Red No. 3, for which a single serving of a variety of food items would result in a hazard index >1 based on the JECFA ADI (though not the US FDA ADI).

In conclusion, while the novel analyses of food dye intake through diet and medication use were not exhaustive in terms of the variety of foods and medications assayed, even with the limited scope of the new analyses, there is reason to believe that some children may routinely consume FD&C food dyes in amounts that exceed the US FDA and/or JEFCA ADIs, particularly through intake of OTC medications. Overall, this reviewer agrees with several of the noted limitations of the current literature on children's exposure and by extension, regulatory policy. Of particular importance are the observations that:

1) The older age of most of the studies reviewed (35-70 years old) is an important limitation of the literature, as there have been numerous advances in neurodevelopmental assessment since then, with more sensitive and rigorous tools now widely in use in the pediatric neurodevelopment literature.

2) The US FDA ADIs are estimated based on animal studies (on dogs and rodents) conducted in the 1960s-1980s, which are mostly not available for public review. To some extent, the WHO JECFA ADIs are based on more recent animal studies and the ADI for Red No. 3 in particular, was based on a study of adult human males and changes in thyroid hormone. However critically, for the WHO JECFA ADIs, as for the US FDA ADIs, none were based on neurobehavioral endpoints, making them inadequate for this purpose.

The report concludes, and this external reviewer agrees, that were the ADIs to be updated based on more recent data (where it exists) and on behavioral outcomes (rather than general toxicity), they would be considerably lower. This further suggests that current regulation of synthetic food dyes is out of date and not based on the most current evidence. Taken as a whole, I believe this proposal to be based on sound scientific knowledge, methods, and practices and have not identified any major weaknesses or omissions that would undermine the authors' conclusions.