IACM
International Association of Color Manufacturers

1101 17th St, NW, Suite 700 | Washington, DC 20036
202-293-5800 | info@iacmcolor.org
www.iacmcolor.org

November 12, 2020

Dr. Mark Miller
Air, Community, and Environmental Research Branch
Office of Environmental Health Hazard Assessment
California Environmental Protection Agency
1001 I Street
Sacramento, CA 95812-4010

Re: Comments in Response to Public Review Draft of "Health Effects Assessment: Potential Neurobehavioral Effects of Synthetic Food Dyes in Children" (August 28, 2020)

Dear Dr. Miller:

On behalf of the International Association of Color Manufacturers (IACM), we appreciate the opportunity to submit comments in response to the California Office of Environmental Health Hazard Assessment (OEHHA) public review draft of "Health Effects Assessment: Potential Neurobehavioral Effects of Synthetic Food Dyes in Children."

IACM is the trade association that represents the global color industry, comprised of manufacturers and end-users of coloring substances that are used in foods, including certified and exempt from certification colors. IACM members create and use colors for a wide variety of food and beverage products. Color additives play an important role in food, and they do so without posing a health risk to consumers.

1

Executive Summary

In October 2018, OEHHA sought information on the neurologic and neurobehavioral impacts of synthetic food dyes to conduct a risk assessment based on a request from the California Legislature. IACM submitted substantive information to OEHHA's request on February 19, 2019[1]. IACM has reviewed OEHHA's draft report, which assesses potential neurobehavioral effects of synthetic colors on children, and respectfully requests that the agency reconsider its inclusion of certain studies as detailed further below that are not relevant to the neurotoxicity endpoint. Since these studies are not appropriate to OEHHA's investigation's focus, the conclusions reached based on these studies should be reconsidered. Otherwise, the findings as drafted are not valid for consideration by the California Legislature.

Additionally, IACM emphasizes the following points:

- Several international expert bodies have drawn a different conclusion regarding the potential causal link suggested by OEHHA. Beginning with the European Food Safety

[1] https://oehha.ca.gov/media/dockets/11132/11336-international_association_of_color_manufacturers/iacm_final_comments_2-19-2019.pdf

Authority (EFSA) in 2008, numerous risk assessment authorities, including the U.S. Food and Drug Administration (FDA) in 2011, have specifically evaluated the clinical evidence purportedly linking consumption of color additives to neurobehavioral effects in children with ADHD and the general population. All of these assessments have concluded that there is no causal relationship. The difference between OEHHA's outcome and those of expert bodies raises serious questions about the validity of OEHHA's investigation, assessment, conclusions, and report.

- Without identifying an apparent neurobehavioral hazard for food colors, it is impossible to conduct a risk assessment for that endpoint for food colors. OEHHA's analysis included reported neurobehavioral impacts in human trials, but significant limitations in methodology reduce these studies' value.
- OEHHA's draft report often emphasizes results from select *in vitro* studies that help support a presumed conclusion (i.e., that color additives affect behavior). At the same time, OEHHA dismisses other *in vitro* and *in vivo* data (e.g., Lok et al., 2013) that indicate a lack of evidence for neurobehavioral impacts.
- There is an overall lack of connection between mechanistic data/experimental animal behavioral studies and human disease outcomes in OEHHA's report.
- In drawing its conclusions, OEHHA gives significant weight to non-guidance studies where weak statistical analysis is used to accentuate inconsistent signals. OEHHA also draws conclusions from "noise" in animal or *in vitro* studies and prioritizes such findings despite overwhelming evidence that supports a conclusion of no effect. Conversely, a lack of consistent results among studies generally leads to a weight-of evidence conclusion that an identifiable hazard does not exist.
- The majority of meta-analyses and systematic reviews of those meta-analyses published in the last 5-7 years have concluded that dietary intervention methods, including diet-restriction approaches (including color restricting) and those that are pro-nutrient, do not significantly alter children's behavior. These conclusions do not support an association between food colors and neurobehavioral endpoints and should be appropriately considered within OEHHA's analysis and report.
- The OEHHA report has a significant flaw regarding its inclusion of studies. While the report suggests that it has taken a systematic approach in its literature search and review, it does not describe the criteria used to qualify or exclude studies. This leaves the impression that OEHHA's weighting of studies in drawing conclusions is either arbitrary, selective for those that fit a narrative, or both.
- The reliability (i.e., quality of methods and reporting) of the studies reviewed and included in the OEHHA report was evaluated only for human epidemiology studies, while study reliability was not formally reviewed for experimental animal studies or mechanistic study data.
- OEHHA reported data for colors that were not related to the objective of the report, which was specific to neurobehavioral effects of synthetic food colors. It was not clearly defined how these data were used or will be used, beyond listing the studies in the narrative.
- Studies of mixtures of food colors are not appropriate for hazard identification. They do not allow the identification of specific food colors that might pose a hazard, if such a hazard exists. Additionally, many of the studies include color additives within the mixtures that are not approved for use in the United States nor within the scope of OEHHA's review. In fact, by considering combinations of colors, OEHHA has, in many

cases, asserted effects for color additives that likely have no contribution to the identified hazard, if such a hazard exists at all within the study.

- OEHHA prematurely concludes that food colors may cause or exacerbate neurobehavioral problems in some children and suggests that current acceptable daily intakes (ADIs) are not protective of children and should be lowered.
- However, OEHHA does not provide any conclusive evidence of causality that would warrants any risk management actions, including lowering ADIs.

OEHHA's Conclusions are Not Based on Sound Science

Regulatory bodies worldwide have evaluated the same studies included in OEHHA's report and all previous assessments have come to a different conclusion than OEHHA. The US FDA, EFSA, and the Joint FAO-WHO Expert Committee Report on Food Additives (JECFA) have all recently conducted risk assessments of and/or reviewed the existing science supporting the continued safe use of food colors. While neurobehavioral endpoints were included in their respective assessments, the pivotal conclusions or recommendations were based on endpoints other than neurobehavioral effects. Even though these risk assessment authorities have reviewed all the available studies on neurobehavioral effects, none have concluded that food colors cause neurobehavioral effects or that the relationship can be established with the available evidence, and therefore do not derive their safety conclusions based on neurobehavioral endpoints.

For example, neurobehavioral effects observed in animal studies have been considered by regulatory and scientific expert bodies. Unlike OEHHA, these expert groups have concluded that these studies or the magnitude of effects within them do not provide evidence that warrant the revision of their respective ADIs. Both JECFA and EFSA concluded that while some animal studies were well-designed and methods were generally well described (Tanaka 2006; Tanaka et al. 2008), the results did not demonstrate any adverse effects on neurobehavioral development, could not be used in risk assessments, and therefore revision of the existing ADIs was not warranted. Neurobehavioral effects as an endpoint for hazard identification have consistently been determined to be insufficient in a risk assessment or to warrant consideration in the derivation of the ADIs for colors. All the relevant expert risk assessors separately concluded that available data on neurobehavioral effects provided insufficient evidence to base or revise an ADI.

Between 2009 and 2014, EFSA conducted detailed evaluations of six of the nine food colors of interest to OEHHA. None of the six EFSA risk assessments identified neurobehavioral effects as a critical endpoint nor relied on neurobehavioral effects to revise the ADI. EFSA continues to conclude there is insufficient evidence for a causal link between any food color and neurobehavioral outcomes in children. In no case was a neurobehavioral effect considered to sufficiently demonstrate a hazard to warrant establishing an ADI on that basis.

Like EFSA, JECFA based its risk assessments on endpoints other than neurobehavioral effects. As a result of their evaluations, JECFA has maintained or increased ADIs for these food colors and did not consider evidence presented for children's neurobehavioral effects to be sufficient for revising the ADI. Additionally, for all colors, JECFA concluded the McCann et al. (2007) was of limited value, and animal neurobehavioral studies were not considered sufficient or robust enough to be included in the risk assessment. JECFA did identify a one-generation reproductive toxicity study for FD&C Blue No. 1 that evaluated neurobehavioral development in mice (Tanaka

et al. 2012) but concluded that the "findings were not robust enough" for purposes of risk assessment.

Lastly, the US FDA Food Advisory Committee (FAC) was asked to evaluate the totality of the evidence and concluded that children's undesired behaviors appeared to result from food intolerance generally, including an individual intolerance to such substances and not a result of neurotoxic properties of the color additives included in the review. During the 2011 FAC meeting, representatives from the FDA stated that none of the agency's ADIs are based on neurological endpoints. However, it was also stated in the FDA's presentations that neurological effects of colors appear to be secondary symptoms of hyperactive behaviors (food intolerance being primary) and there is no indication that colors are directly impacting neurological processes. It was noted that the ADIs of five of the six most commonly used colors were based on two-year animal studies, which included *in utero* exposure as well as dose levels up to the maximum tolerated dose in order to capture the developmental period from the point of conception through end of life at very high dose levels. During examination of those trials, several neurotoxicity screening issues are observed or examined to determine if there is a need to do further specific neurotoxicity testing. The studies did not include specific neurobehavioral testing, but did include clinical observations of lacrimation, clinical observations of behaviors in the normal cage setting.

OEHHA's conclusion on page 20 of the draft report that the current FDA ADIs are not adequately protective of children, is not based on sound science. The ADI represents a conservative upper daily intake that is not expected to result in an adverse effect in the most sensitive individuals regarding general and organ-specific toxicity, including reproductive, developmental, neurotoxicity, genotoxicity, and other forms of toxicity. Authoritative bodies develop ADIs based on points of departure, frequently the no observable adverse effect level (NOAEL) or Benchmark Dose level (BMDL). The ADIs for FD&C colors account for the most sensitive 10 percent of individuals, assuming that humans are 10 times more sensitive than the most sensitive test species. As noted previously, while the animal studies examining the potential effects of food colors on neurobehavioral endpoints were included in the recent regulatory safety reviews of EFSA and JECFA, none of those studies presented evidence of adverse effects that would warrant revision of the established ADIs.

If OEHHA's assertions that all synthetic colors cause neurobehavioral effects were accurate, which the evidence does not support, then there must be realistic and shared molecular mechanisms that would provide some explanation for these effects to be caused by all such colors. The allowed (certified) color additives in the U.S. include azo dyes, a triarylmethane derivative, an indigotine derivative, and an iodofluorescein derivative. These chemical classes are significantly distinct in chemical structure. Therefore, ascribing behavioral effects to all FD&C colors via a unified mechanism is not scientifically supported, is not addressed nor demonstrated in OEHHA's draft report, and should be carefully questioned.

Additionally, as other regulatory bodies have noted, the human studies purporting an association between food color consumption and adverse behavior all suffer from protocol limitations, demonstrating statistically insignificant or inconsistent associations. It is not clear how OEHHA reviewed the same studies as other expert bodies and drew such different conclusions. Other regulatory bodies have repeatedly concluded that there is no consistent evidence of an association between food colors and neurobehavioral effects, much less a causal relationship.

It appears OEHHA has not seriously considered if test articles used in reported studies are comparable to purity as well as impurity standards set forth by regulations for batch-certified food colors that industry uses. Declaring only the purity information with test articles bought from Sigma is not sufficient to assess compliance to batch certified standards. The specific impurity information becomes vital since the minimum purity requirements for food colors under FDA regulations may be lower than the 99.5% material provided by Sigma, and also come with restrictions on specific impurities. Further, widespread accessibility to FDA testing procedures used in batch certification process is limited, and unless the study specifically indicates that the food colors used meet the identity and specifications as regulated by FDA, it is not possible to determine if a test article is comparable to an FDA batch-certified color. This is important contextually when reviewing experimental data because color test articles used in studies are generally not reported to be batch certified under the Federal Food Drug & Cosmetic Act, and many studies lack sufficient information about the purity and extent of critical impurities and reaction by-products in test articles for which maximum limits are set by the regulation of the color additive.

For example, FD&C Red No 3 under 21 CFR 74.303[2] not only needs to meet the purity requirement of 87%, but contains specifications that restrict the levels of other iodinated compounds like sodium iodide, (≤ 0.4%), triiodoresorcinol (≤ 0.2 %); 2(2',4'-Dihydroxy-3', 5'-diiodobenzoyl) benzoic acid (≤ 0.2 %); monoiodofluoresceins (≤ 1.0 %); and other lower iodinated fluoresceins (≤ 9.0 %). It is vital that the test articles used in studies meet these specification requirements as excess iodide from highly bioavailable impurities like sodium iodide or other iodinated compounds can lead to confounding outcomes. In this case, even with presence of 13% sodium iodide, a test article would still meet FDA purity criteria but not impurity restrictions. This can result in a biological outcome due to this impurity rather than due to the dye. Like FD&C Red No 3, restricted impurities and limits for all FD&C colors can be found in 21 CFR Title 74.

Furthermore, we note that bioavailability of the colors was generally not addressed in the OEHHA report, except for including absorption and bioavailability as important research needs for future studies. It should be noted, however, that bioavailability is low for all seven colors (EFSA, 2009a,b,c, 2010, 2011, 2014; WHO/FAO, 2017a,b, 2019).

Observations on Study Quality/Conclusions

*Epidemiologic Studie*s

Overall, design problems exist across epidemiologic studies that attempted to assess the causal relationship between food dyes and potential neurobehavioral concerns such as ADHD in children. In fact, OEHHA stated (page 82, Section 2.4.2), “It appears that the recruitment of participants in most of the studies we reviewed involved convenience sampling, and few studies provided enough information for us to calculate or estimate participation rates. Because of this, we could not use overall participation rates or other subject selection criteria as an indicator of study quality.” Other shortcomings in study design included proper assessment of the exposure (e.g., exposure to different dyes and/or mixtures of multiple dyes, different purity standards across countries); temporality (adequate washout period, timing of exposure vs. testing, whether the exposure to food dye potentially results in either transient or long-term effects);

[2] https://www.ecfr.gov/cgi-bin/text-idx?SID=eb4bb165720bc22df4dda6637c022217&mc=true&node=se21.1.74_1303&rgn=div8

consideration of bias, chance, and confounding issues (e.g. other potential causes of neurobehavioral problems); and/or problems with scoring methods between different assessors (teachers, researchers, or parents using different metrics or tests). Diagnosing a complicated neurobehavioral disorder like ADHD is challenging as the related symptoms and signs are quite individualized—what may be a symptom of ADHD in one person may not be a symptom in someone else. In addition, ADHD may be triggered by any number of currently unknown factors.

OEHHA noted in its report (page 29, Section 2.1) that it did not perform a full meta-analysis due to the "high quality" publication of Nigg et al. (2012). This meta-analysis identified 24 studies published from 1976 through February 2011 and evaluated behavioral effects (relative to inattention and hyperactivity) and color additives. Of the 24 studies included in Nigg et al. (2012), only 11 studies, including McCann et al. (2007), evaluated hyperactive children. The authors noted a wide variation in responders between studies. They reported that some children in the reviewed studies saw a reduction in ADHD symptoms on restriction diets. In focusing exclusively on Nigg et al. (2012), the OEHHA report notably does not consider many other meta-analyses and systematic reviews, including Sonuga-Barke et al. (2013), Stevenson et al. (2014) or Pelsser et al. (2017). Going further back into the literature, even a study by Kavale and Forness (1983), an earlier meta-analysis of studies that evaluated several elimination diets, including the Feingold diet, and challenge trials with color additives found no statistically significant association between color additives and hyperactivity in children. Further, no clear criteria were established in the OEHHA report to conclude that the Nigg et al. (2012) analysis is sufficiently high quality to disregard the full weight of evidence, and raises the question why OEHHA emphasized a study finding an effect while minimizing those that did not.

As highlighted during a meeting of the FDA's Science Board in October 2019 on the topic of Color Additives and Behavioral Effects in Children[3], and emphasized by IACM's February 19, 2019 comments, there are two additional recent publications that are critical for OEHHA to consider in its assessment, a meta-analysis of dietary interventions for ADHD by Sonuga-Barke et al. (2013), and a systematic review by Pelsser et al. (2017).

Sonuga-Barke et al. (2013) was published on behalf of the EUNETHYDIS11 European ADHD Guidelines Group, used largely the same dataset as Nigg et al. (2012) and reported similar statistically significant but small (yet clinically insignificant) effects on symptoms of ADHD from ingestion of color additives, but drew slightly different conclusions. The authors identified eight papers that evaluated food color additives which met the authors' criteria for inclusion. The meta-analysis revealed a statistically significant but weak association between food color restriction and improved behavior. Significantly, when the analysis was limited to (1) four papers that utilized a protocol with low or no pharmacological interventions (because allowing subjects to continue with taking medication to treat ADHD may reduce the ability to detect a potential effect due to food colors) and (2) protocols that were likely blinded, the association between color additives and the behavioral impact was further reduced and ceased to be statistically significant. Pelsser et al. (2017) performed a systematic review of two meta-analyses (Schab, et al., [2004] and Nigg, et al., [2012]) that evaluated the evidence associated with elimination diets for food color additives and ADHD and concluded that the current evidence does not support restriction of food color additives for the treatment of ADHD.

OEHHA also chose not to consider a second publication by the EUNETHYDIS European ADHD Guidelines Group, Stevenson et al. (2014), which focused only on dietary treatments for ADHD.

[3] https://www.fda.gov/media/135001/download

Stevenson et al. (2014) reviewed three meta-analyses related to ADHD and the purported efficacy of the artificial food color elimination diet (i.e., Schab et al., [2004]; Nigg et al., [2012]; Sonuga-Barke et al., [2013]). Stevenson et al. (2014) concluded that the effect size was too small to be of value and that the patient population that would benefit from a color additive elimination diet remains uncertain. Consistent with previous evaluations, the authors ultimately came to the same conclusions that the methodology used in most of the trials on which the meta-analyses are based were weak, limiting their ability to demonstrate an efficacious treatment for ADHD.

Taken together, reviews of the clinical trial literature associated with ADHD and the consumption of color additives have produced neither consistent nor strong association between color additive intake and adverse neurobehavioral effects. Furthermore, removal of color additives from the diet has not been demonstrated to be an efficacious treatment of ADHD. None of the studies conducted to date have individually or collectively provided evidence to support the conclusion that an association exists. Moreover, the weak evidence that is detected through systematic review and meta-analysis has been inconsistent and likely the product of subjective diagnostic criteria.

The only challenge study that attempted to replicate the findings of McCann et al. (2007) in a different population was published by Lok et al. (2013) which, OEHHA notes (page 43, Section 2.7.4), was not included in the Nigg et al. (2012) meta-analysis. Lok et al. (2013) replicated the design of the McCann et al. (2007) study including the use of a randomized double-blind placebo-controlled design with a within-subject cross-over protocol, to assess hyperactivity in eight to nine year-old children in Hong Kong. The authors stated that this study "does not attempt to negate or contest the findings of the Southampton study but to build on this study in a sample of Chinese children because food safety in China is a major public health issue," hence the study adhered to a protocol very similar to that used by McCann et al. (2007). Lok et al. (2013) used the same doses of the same color additives used in McCann et al. (2007) in Mix A (FD&C Yellow No. 6, FD&C Yellow No. 5, carmoisine, and ponceau 4R) with a few differences such as attempting to obtain a study sample including children from a wide spectrum of socioeconomic backgrounds. More notably, Lok et al. (2013) excluded children with ADHD and currently treated with medication, diagnosed with diabetes, with phenylketonuria (in which interference with diet can have adverse health effects), and/or other mental health problems (e.g., learning disability, Down syndrome) from the study population. These exclusion criteria were not applied in the McCann et al. (2007) study. Other differences include administration of treatment given as capsules instead of juice; timing of administration was specified to be in the morning before school rather than anytime of the day; the preservative sodium benzoate was not included in the same treatment as food colors but was tested separately; and assessment was based on teachers and parents only but with no independent assessor. Lastly, Lok et al. (2013) assigned two types of scores: a) strengths and weaknesses of ADHD symptoms (positive and negative scores) and normal behaviors (SWAN) rating scale (based on the DSM-4 diagnostic criteria) that teachers and parents filled in; and b) child behavior checklist (CBCL), which only the teachers filled in, both of which have been validated with local norms in Hong Kong.

In contrast to McCann et al. (2007), Lok et al. (2013) did not detect an association between color additive intake and behavior, even though evaluation was also based on parent and teacher assessments. Because Lok et al. (2013) used essentially the same protocol as McCann et al. (2007), IACM recommended in our February 19, 2019 comments that OEHHA carefully

review the study protocol and findings from Lok et al. (2013) to assess the lack of reproducibility of McCann et al. (2007).

While the results of McCann et al. (2007) were the impetus for a renewed interest in the potential for food color additives to negatively affect behavior in children, regulatory agencies worldwide have dismissed and discounted the findings from McCann et al. (2007) in their own subsequent reviews. Lok et al. (2013) studied a different population using a very similar protocol and larger doses of the food color additives but was unable to reproduce the results, which should raise significant concerns for OEHHA as it has with other regulatory agencies, diminishing the confidence in results from McCann et al. (2007), and be viewed as offering limited value due to the lack of reproducibility.

Unfortunately, the OEHHA report provides only a cursory review of Lok et al. (2013) and ultimately dismisses the study, seemingly because independent observers and computerized tests like the Conners tests were not used as in McCann et al. (2007), even though the study uses two behavioral metrics based on parent and/or teacher reporting that have been validated in China. It is unclear why OEHHA allows for one psychological test and not the other, when both have been validated as viable tools for clinical assessment. OEHHA also notes that Lok et al. (2013) did not report an association for any outcome metric and proceeds to dismiss the study's validity due to no observed effect.

Even without considering Lok et al. (2013), reviews of the clinical trial literature associated with ADHD and the consumption of color additives have not produced consistent associations between color additive intake and undesired neurobehavioral symptoms. While there are numerous clinical studies that attempt to investigate the relationship between food colors and potential neurobehavioral effects, none of the studies conducted have succeeded in providing the evidence that would support the conclusion reached by some meta-analyses and OEHHA that an association exists. Unfortunately, OEHHA's report lacks transparency in the criteria used to assess the quality of these studies.

Animal Neurotoxicity

The animal toxicity studies that were considered in the OEHHA draft report were not evaluated by OEHHA using any type of published methods for study quality or reliability. OEHHA reports changes in all neurobehavioral outcomes without consideration of study quality, design, consistency in the test methods employed, or interpretation of findings across studies. All developmental neurotoxicity study designs were considered with neurobehavioral measures and exposure throughout development. Temporal concordance of rat and human brain development was not provided when assessing the behavioral effects. As we noted previously, in many cases, the purity of the test article was not reported in many of the evaluated studies. Observed effects were attributed to the color only, and the influence of potential impurities of the test article was not considered. Types of behavioral tests (behavior ontogeny, motor activity [including habituation], motor and sensory function, or learning and memory) were scored using different tools/procedures (automated versus manual, real-time versus post-test, single versus repeat testing of animals) and varied across studies. All statistically significant outcomes appeared to be incorporated into the OEHHA summary of animal toxicology without evaluation or discussion regarding the lack of consistency of effects across studies, between sexes or life stages within the same study, or of the observed animal-to-animal variation in effects upon exposure to colors versus those in the control groups. Statistically significant behavioral results were provided as evidence when pairwise or trend tests were positive, yet without associated

specific reference to the importance of litter randomization prior to testing and litter-based statistical analysis. Moreover, the results presented encompassed different species (rat and mouse) and strains, with concurrent control groups but without positive controls or reference to historical control data.

Regarding the studies conducted predominantly in the Tokyo Metropolitan Laboratory of Public Health (Tanaka 1994, 1996, 2001, 2006; Tanaka et al. 2008, 2012), OEHHA noted (page 89, Section 3.2.4):

> It is tempting to compare across the food dyes for developmental neurotoxicity using the studies from the Tokyo lab. However, these data alone are not adequate to conclude that some dyes are more toxic than others by comparing across studies. While the design of the studies is similar there were changes in procedures, equipment, and statistical analysis over time (1994-2012). An effort was made by the laboratory to test comparable dose ranges in this set of studies by using multiples of the JECFA ADIs for the low dose. However, different multiples (100X, 10X) were used and the JECFA Red No. 3 ADI is notable for being two orders of magnitude lower than the FDA ADI. In addition, reproductive and developmental toxicity varied across studies and could influence the later behavioral assessments.

However, a review of the Tanaka studies found a lack of evidence of consistent or sustained adverse changes within studies or across studies with individual colors. A statistical evaluation of the results for specific endpoints reported for control groups across the individual studies from this laboratory, where possible, indicated statistically significant differences in endpoints in the control groups (absent any treatment) over time. This demonstrated significant variability in the background response to these neurobehavioral assays within this laboratory. Further, combination of control results for comparison across treatment groups resulted in the loss of statistical significance in selected responses, because of the significant variability in control responses over time, hindering the ability to draw any meaningful conclusions from the reported observations upon exposure to any of the colors evaluated. Also, while OEHHA notes that the JECFA ADI for Red No. 3 is two orders of magnitude lower than the FDA ADI, it is worth noting that it is actually only one order of magnitude lower.

Summary information is provided for each of the colors as follows, with complete contextual analysis available in Appendix A.

FD&C Blue No. 1/Brilliant Blue

Only one study was identified for Brilliant Blue (Tanaka et al. 2012). OEHHA concluded the lowest observed adverse effect level (LOAEL) from Tanaka et al. (2012) was 127 mg/kg/day. However, Gentry et al. (submitted) concluded that the results of Tanaka et al. (2012) provide only limited evidence of isolated changes that are not sustained over time despite chronic exposure to the test color and are limited to one or two endpoints measured from a large number intended to evaluate neurobehavioral domains. While the results of Tanaka et al. (2012) are adequate for quantitative risk assessment, the study did not provide consistent evidence of effects, even following doses higher than ADI.

FD&C Red No. 3/Erythrosine

Overall, four studies were identified for review (Tanaka [2001]; Vorhees et al. [1983a]; Dalal and Poddar [2009, 2010]). OEHHA concluded that the Tanaka (2001) study had a no observed adverse effect level (NOAEL) of 28 mg/kg/day which is lower than the FDA NOAEL of 250 mg/kg/day based on the observation of "distended cecum" in rats, suggesting that the ADI should be protective against neurobehavioral effects. However, this assumes Red No. 3 causes neurobehavioral effects in the absence of sufficient evidence.

Gentry et al. (submitted) concluded that the results of Dalal and Poddar (2009, 2010) would not be considered for any quantitative evaluation of neurobehavioral effects in children because both studies were conducted in mature adult rats. In addition, as noted earlier, the specifications for the test substance reported in Dalal and Poddar (2009, 2010), Erythrosine B, purchased from Sigma is not comparable to FDA specification for FD&C Red No. 3 under 21 CFR 74.303. Therefore, critical information about purity of the color additive, as well as the extent of critical impurities and reaction by-products in the color additive for which maximum limits have been set by regulation are unknown.

Gentry et al. (submitted) concluded that the study conducted by Vorhees et al. (1983a) would also not be considered for quantitative risk assessment because specific study quality and reliability criteria were not met (i.e. ToxRTool score of 3 – not reliable). Finally, while the results of Tanaka (2001) are deemed to be adequate for quantitative risk assessment, the study did not provide consistent evidence of effects, including following doses higher than the JECFA or FDA NOAELs.

FD&C Red No. 40/Allura Red

Overall, three studies were identified for review (Tanaka [1994]; Vorhees et al. [1983b]; Noorafshan et al. [2018]). Based on the results of Noorafshan et al. (2018), OEHHA proposed a LOAEL for Allura Red of 7 mg/kg/day, which is the same as the FDA and JECFA ADI of 7 mg/kg/day, suggesting that the ADI may not be protective against neurobehavioral effects. OEHHA concluded that if the results of Noorafshan et al. (2018) were to be used as the basis for setting an ADI the resulting ADI would be 100 to 1000 fold lower than the existing ADI, depending on the method used to derive the point of departure.

However, the results of Noorafshan et al. (2018) should not be considered for any quantitative evaluation of neurobehavioral effects in children because the study was conducted in adult rats. In addition, the Radial Arm Maze test conducted by the authors to assess learning and memory included a forced food restriction prior to testing so the animals would lose 15% of body weight. Food-based reward was used to encourage the animals to complete the test. The authors offered no details on the potential effects of the forced food restriction and reduction of body weight on the bioavailability of the test substance, the effect of 15% body weight reduction on the overall outcome compared to normal weight control animals, or the relevancy of the forced food restriction to normal human behavior.

The study conducted by Vorhees et al. (1983b) should not be considered for quantitative risk assessment because of limited data presented for quantitative risk assessment and because specific study quality and reliability criteria were not met. Specifically, the test substance purity and the housing and feeding conditions of animals were not reported in the study (i.e. ToxRTool score of 3 – not reliable).

While the results of Tanaka (1994) are adequate for quantitative risk assessment, the study did not provide consistent evidence of effects even using doses higher than JECFA of FDA NOAELs.

FD&C Yellow No. 5/Tartrazine

Overall, five studies were identified for review (Tanaka [2006]; Tanaka et al. [2008]; Sobotka et al. [1977]; Gao et al. [2011]; Rafati et al. [2017]). OEHHA concluded that the lowest NOAELs identified of 175 mg/kg/day in mice and 125 mg/kg/day in rats (Gao et al. 2011) were lower than the NOAELs identified by FDA of 500 mg/kg/day in dogs and 1000 mg/kg/day in rats, and 1.5 percent in diet estimated at 2250 mg/kg/day in mice. It is worth noting that the endpoints used by OEHHA to identify NOAELs went beyond neuro-related endpoints, bringing into question the relevancy of the conclusions reached.

Additionally, the results of Sobotka et al. (1977) did not provide consistent evidence of effects. The study conducted by Rafati et al. (2017) would not be considered for quantitative risk assessment because the study also did not provide consistent evidence of effects. In addition, the advanced age of rats is a confounder in the memory tests conducted as age alone has been shown to affect radial arm maze test performance. The results of Gao et al. (2011) should not be considered because specific study quality and reliability criteria were not met (i.e. ToxRTool score of 3 – not reliable). Tanaka (2006) and Tanaka et al. (2008) also do not provide consistent evidence of effects, including following doses higher than the JECFA or FDA NOAELs.

FD&C Yellow 6/Sunset Yellow

OEHHA only identified one study for Sunset Yellow (Tanaka 1996) and concluded that while some neurobehavioral effects in offspring were reported for preweaning development and maze learning in Tanaka (1996), conclusions could not be drawn due to the statistical approach and varying group sizes in the study. OEHHA stated that a NOAEL without a LOAEL in the same study is not suitable for risk assessment. It is worth noting that a study does not need a LOAEL to have a NOAEL. If the highest dose shows no adverse effect, that is the NOAEL of that study. However, it does not mean it is the NOAEL of the material as well. Instead, a weight of evidence of all studies is needed with the lowest adverse effect dose in the most sensitive species representing the pivotal NOAEL for deriving an ADI. While the results of Tanaka (1996) may be adequate for quantitative risk assessment, the study did not provide sufficient or consistent evidence of effects, including following doses higher than the JECFA or FDA NOAELS.

Mixtures

OEHHA also considered animal neurotoxicity studies in its draft report that considered mixtures of colors. The eleven studies identified would not be appropriate for inclusion in a quantitative risk assessment. For six studies, there is no way of determining whether the effects were from one color or a combination of two or more colors or any other potential confounder or impurities, especially if the purity of test substance was not reported (Doguc et al. [2013]; Doguc et al. [2019]; Doguc et al. [2015]; Basak et al. [2014]; Basak et al. [2017]; Ceyhan et al. [2013]). The results of Ceyhan et al. (2013) also indicate exposure to synthetic food colors may lead to alterations in expressions of N-methyl-D-aspartate (NMDA) receptors and nicotinic acetylcholine receptors (nAChRs) in adulthood with possible differences in males and females noted; however, the toxicological meaning of these changes is not completely understood.

In the five other mixture studies (Kantor et al. [1984]; Shaywitz et al. [1979]; Goldenring et al. [1980]; Reisen and Rothblat [1986]; and Erikson et al. [2014]), it is impossible to assess the results of the study on an individual dye basis or establish any potential dose-response relationship for individual dyes. Each study also introduces significant uncertainties. In Kantor et al. (1984), only one sex was tested and only locomotor activity in adult animals was evaluated. In Shaywitz et al. (1979) and Reisen and Rothblat (1986), exposure was via oral gavage and not diet which would have otherwise emulated human exposure. In Goldenring et al. (1980), the exposure route was by gastric cannula and not through the diet further limiting the use of the study results. Additionally, due to the limited number of animals, only one dose group was tested. In Erikson et al. (2014), exposure was only during a limited time during childhood development. Therefore, these studies should not factor into a quantitative risk assessment.

Toxicokinetic and Mechanistic Data

While the OEHHA report mentions various *in vivo* and *in vitro* studies of mechanistic data for each color, there is no overall conclusion regarding how such data contribute to the understanding of underlying mechanisms of neurobehavioral effects, nor are such data considered in the context of plausible modes of action or adverse outcome pathways related to neurobehavioral outcomes. There is an overall lack of commentary on dose (relevance to human exposure), duration/timepoints, models used (predictability or consistency of models used, either *in vivo* or *in vitro*), or "bridging" of molecular or cellular signals to behavioral effects in the OEHHA report. It is important to note that a single signaling event is not enough for an adverse outcome pathway to be relevant, but rather multiple key events (KEs) are required to get from the molecular initiating event (MIE) to the adverse outcome.

The OEHHA report considered the high-throughput screening (HTS) assay data to provide 'limited support for *in vivo* 'neurotoxicity,' based on the fact that the colors had some, albeit inconsistent, activity in HTS assays deemed relevant to neurobehavioral effects. Further, clarity was not provided on how such *in vitro* data were specifically related to outcomes in humans and/or laboratory animals. Although the OEHHA report included an extensive assessment of HTS data, such as bioactivity of HTS assays, from the ToxCast/Tox21 program for the seven colors, the analyses included many non-specific assays for neurobehavioral effects (e.g., markers of oxidative stress or inflammatory response in non-neuronal cell lines), and did not account for confounding by refraining from integrating potential cytotoxic interference or other data quality issues (such information is provided in the ToxCast/Tox21 database). OEHHA did not attempt to elucidate underlying mechanisms for neurobehavioral effects with the use of HTS data. Instead, the report concludes generally that the HTS data are limited yet supportive of *in vivo* neurotoxicity observations for the food colors, without specifying connections between HTS results and *in vivo* outcomes.

High-Throughput Screening Data Review

OEHHA states (page 174, Section 5.4):

> Based on the subset of assays we evaluated here, the ToxCast assay results provide limited support for *in vivo* neurotoxicity observations for the food dyes. It should be noted that the assays explored here are intended to provide initial information about the capacity to associate *in vitro* work with the ability for a food dye to promote a biological response. However, these assays are limited for predicting long term or indirect adverse effects in complex biological systems, in part, due to the complexity of the *in vivo*

> pathway interactions leading to neurotoxicity (including neurobehavioral effects) and DNT [developmental neurotoxicity] compared to the current limited spectrum and range of the ToxCast assays. Evaluation of these chemicals in future iterations may offer more refined results and validate that these gene markers play a critical role in chemically-induced mechanisms of neurotoxicity."

While we appreciate the utility of HTS testing, any response noted in HTS assays would ultimately only help to formulate hypotheses that would require further testing to determine whether support for a particular adverse neurobehavioral outcome pathway exists. HTS assays, by definition, represent a screening and prioritization strategy but cannot be used to 'validate' any particular signal as critical to chemically induced mechanisms of neurotoxicity or assign hazard identity. Additionally, at this time, these individual assays can only be reliably used when information on assay interference such as cytotoxicity and chemical purity and stability is available and incorporated into the assessment. This information would increase confidence in activity calls and would enable integration into a reliable evidence base that provide support (or not) for an association with a particular neurobehavioral outcome or with specific events within a neurobehavioral adverse outcome pathway.

Chappell et al. (2020) concluded that "the results of our assessment of available *in vitro* mechanistic data collected from assays that measure signals related to MIEs or K.E.s involved in neurodevelopmental processes indicate that the seven FDA-approved food colors (when batch certified) have limited or no activity for such signals. While available information on FD&C colors and genes or enzymes that may have a role in mechanisms of neurodevelopmental alterations may be limited, FD&C Red No. 3 was the only color (of the seven assessed) that showed activity associated with neurodevelopmental pathways. Additional follow-up assays, especially with test articles that pass analytical Q.C. criteria, would provide clarity and increased confidence in these findings. Overall, the FD&C colors do not appear to alter signaling pathways related to neurodevelopmental processes on the molecular or cellular level."

Set of HTS assays used in assessments

OEHHA included 283 HTS assays in their risk assessment, 182 of which were deemed directly neuro relevant. The other assays included in the OEHHA assessment were generally related to inflammation, oxidative stress, or were selected based on the activity of known neurotoxicants (pesticides), regardless of assay target category or type.

However, OEHHA limited their identification of neuro-relevant assays to the NovaScreen (NVS), Attagene (ATG), and Tox21 vendors. OEHHA included assays with relatively broader (i.e., indirect, unclear, and/or non-specific) potential relationships to neurobehavioral adverse outcomes, such as estrogen receptor (E.R.), androgen receptor (A.R.) antagonism/agonism, and inflammation in non-neuronal cells, respectively. These should not be included as inflammatory response in other cell systems is too general to draw conclusions regarding the ability of the color tested to induce inflammation in the brain *in vivo*, and a link between androgenic or estrogenic changes and neurobehavioral outcomes has not been identified to be a key event in any of the pathways identified to date for neurobehavioral adverse effects as noted in Chappell, et al. (2020).

In the OEHHA report (page 158, Section 4.3.3.5), the following statement is incorrect, "Even with the limitations of the *in vitro* data, in contrast to a recent study published by Chappell et al. (2020), our approach resulted in significantly more active assay hits (283 compared to 116

assays)." In fact, a total of 99 HTS assays were mapped to potential mechanisms of neurobehavioral outcomes in Chappell et al. (2020). Across these 99 assays, the "coverage" of the seven colors tested culminated in a total of only 116 assay endpoints. Further, the OEHHA report lists 283 assays selected for their assessment based on their designation that these are in some way relevant to neurobehavioral outcomes. For active assays, however, according to OEHHA's Appendix C, Table 1, a total of 350 assay endpoints across the seven colors among these 283 assays were considered active (based solely on "hit-call") but did not account for all the data quality flags noted previously.

The number of active" hit-calls" are, consequently, much fewer in Chappell et al. (2020) compared to OEHHA's assessment due to a lower number of overall assays deemed relevant to neurobehavioral outcomes, as well as integration of data quality issues and cytotoxic interferences, which OEHHA did not account for nor integrate into their assessment. Therefore, OEHHA should state the following to ensure a more appropriate characterization of the difference: "Even with the limitations of the *in vitro* data, in contrast to a recent study published by Chappell et al. (2020), our approach resulted in more assay endpoints included in the assessment (283 for OEHHA compared to 99 for Chappell et al. (2020)) and more corresponding active "hit-calls" (350 for OEHHA and 8 for Chappell et al. (2020). These differences could be explained by the fact that (i) we cast the net much wider to include indirect effects that have questionable associations with neurobehavioral outcomes and (ii) we (OEHHA) did not account for data quality issues or assay interference due to cytotoxicity in determining activity calls in contrast to the approach Chappell et al. (2020) took."

Consideration of Cytoxocity and data quality flags for HTS assay activity

OEHHA (where stated) applied no filter for cytotoxic interference (i.e., the AC50 value of activity relative to the cytotoxic concentration), nor data quality flags (assigned by ToxCast, related to issues in data analysis and model fitting), nor chemical analytical quality control (Q.C.) (e.g., purity and identity). Determination of activity or inactivity was based solely on the "hit-call" provided in the ToxCast database. While Judson et al. (2016) is cited by OEHHA as stating that the cytotoxic burst should not be used as a filter, cytotoxic interference is a known and well-established factor that should be considered in the interpretation of *in vitro* data as discussed in the very same paper (Judson et al., 2016). It is worth noting that the U.S. Environmental Protection Agency (EPA), the National Center for Computational Toxicology [now the Center for Computational Toxicology and Exposure] (NCCT/CCTE) – the EPA division that works with these assays – is not using HTS assays at this time for either hazard or risk assessment. Consequently, it is imperative that the cytotoxic interference information is considered in the determination of assay activity; and, more specifically, it provides the necessary context to assign assay data as "unreliable" vs. active or inactive.

Outside of cytotoxic burst criteria, the viability assays in the Tox21 program are specific to measuring cell death related to specific individual assays. There is no reason to ignore such assay information that would otherwise help with appropriate analyses and interpretation of the results. Chappell et al. (2020) prioritized such viability assay data first relative to cytotoxic burst information for any assays for which specific viability assay data were available (such data are only available for Tox21 assays).

OEHHA did not consider data quality flags in their assessment (page 146, Section 4.3.1). Chappell et al. (2020) noted more than one data quality flag would render the assay not "active".

Examples of data quality flags include: "Noisy data," "Hit-call potentially confounded by overfitting," "Only highest conc above baseline, active," among others.

The OEHHA report makes no mention of chemical quality information in Section 4.3 on HTS assays. While Chappell et al. (2020) did not exclude data based on sub-optimal or absent chemical quality data, such information was discussed in the broader context of overall interpretation of the data.

It should be noted that other groups/programs are applying various criteria when deciding when to assign a classification of "active" to responses observed for HTS assay endpoints, and/or to include the assay data in an assessment. Relative to National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM)/Integrated Chemical Environment (ICE) curated high-throughput screening (cHTS) data:[4] Data quality is considered by omitting any assay endpoints with a "hit-call" of active from curated data when: the assay is a down-direction assay (i.e., inhibition, antagonism, loss-of-signal, etc.), and the best-fit curve was a gain-loss model; or the best-fit curve was a gain-loss model, and only a single mid-range concentration had activity above the activity cutoff threshold, among other criteria. Sample quality is considered by omitting assay endpoints from the curated ToxCast HTS data within ICE in which chemicals with a chemical Q.C. grade of "caution" was used, among other criteria.

Interestingly, and somewhat contradictory to the approach taken for synthetic colors, OEHHA itself excluded HTS assays based on data quality flags and/or chemical Q.C. issues (i.e., lacking Q.C. information or major issues with chemical analytical Q.C. results) in their evidence on the Carcinogenicity of Acetaminophen (2019).[5]

Summary information is provided for each of the colors as follows, with complete contextual analysis available in Appendix B.

FD&C Blue No. 1/Brilliant Blue

Only minimal neuro-relevant mechanistic information is available for Blue No. 1. OEHHA states (page 169, Section 5.3.5), "Both Blue No. 1 and Green No. 3 inhibit purinergic receptors"; however, the relevance of this finding to potential neurobehavioral outcomes is unclear because purinergic receptors are potentially related to hereditary neurodegenerative diseases involving neuroinflammation (e.g., Huntington's disease) as explained for Blue No. 1 by Wang et al., 2013), or have been shown to have an antinociceptive (i.e., beneficial) effect (related to exposure to Green No. 3, as described in Yang et al., 2019). The relevance of purinergic signaling to neurobehavioral effects in children is not explained by OEHHA. It is also stated that "Blue No. 1 inhibited neurite outgrowth in cultured neuroblastoma cells." As clarified by Chappell et al. (2020), the concentration at which neurite outgrowth was inhibited was also significantly cytotoxic to the cells. This diminishes the relevance of this response and its value in the evidence base of neuro-relevant mechanistic data. Taken together, there was insufficient evidence to support an association between exposure to Blue No. 1 and mechanistic changes that are potentially related to neurobehavioral outcomes.

FD&C Blue No. 2/Indigo Carmine

[4] https://ice.ntp.niehs.nih.gov/DATASETDESCRIPTION?section=cHTS

[5] https://oehha.ca.gov/media/downloads/crnr/acetaminophenhid092019.pdf

No neuro-relevant mechanistic data were identified for Blue No. 2. OEHHA (page 169, Section 5.3.5) points to studies that identify cardiac "side effects" that occur under unique exposure conditions, which have led to hypotheses that a relationship exists between these measured changes and possible serotonin-based mechanisms and histamine release into circulation (Erickson and Lauron [1960]; Jo et al., [2013]; Lee et al., [2015]). The studies that suggest this information were based on non-oral routes of exposure (intravenous, subcutaneous, or intramuscular) to Blue No. 2 for clinical diagnostics. There was no evidence to suggest that the same effects would occur following oral exposure to Blue No. 2. Overall, no *in vivo* Blue No. 2 studies provided mechanistic data relevant to neurobehavioral outcomes, and the limited HTS assays mapped to neuro-relevant mechanistic data were inactive (Chappell et al., 2020).

FD&C Green No. 3/Fast Green

Only limited neuro-relevant mechanistic information was identified by OEHHA for Green No. 3. OEHHA states (page 169, Section 5.3.5), "Both Blue No. 1 and Green No. 3 inhibit purinergic receptors"; however, the relevance of this finding to potential neurobehavioral outcomes is unclear because inhibition of purinergic receptors has potential neurotherapeutic value and would be considered not relevant for evaluating adverse neurobehavioral effects. The rationale for such a link is not explained further in the OEHHA report.

FD&C Red No 3/Erythrosine

Red No. 3 has a fuller neuro-relevant mechanistic evidence base compared to all other synthetic colors/dyes being evaluated. Nonetheless, much of the available evidence must be considered with care. OEHHA provides reference to many of these studies without conducting due diligence on study quality. OEHHA summarizes the available evidence on Red No. 3 in Section 5.3.5 (page 169): "Some studies of Red No. 3 have reported changes in neurotransmitter uptake by brain tissues, inhibition of enzymes (acetylcholinesterase, Na+/K+ ATPase), and photooxidation of enzymes. Together this *in vitro* work suggests that Red No. 3 can have many biological targets relevant to brain function and is consistent with contemporary work on Red No. 3 protein binding. As well, Red No. 3 has been shown to affect thyroid hormones in both rodents and humans. The absorption of Red No. 3 appears to be low based on very limited pharmacokinetic data. However, there is some absorption and metabolism. Deiodinated metabolites have been measured, and the time of peak Red No. 3 circulating levels corresponds to the time of peak Red No. 3 effects on behavior (activity) and impacts on neurotransmitters measured in *in vivo* studies."

Despite OEHHA's summary, and although changes in neurotransmitter function remains a K.E. that could – in combination with other K.E.s – potentially lead to neurobehavioral adverse outcomes, many of the studies (especially the *in vitro* studies) did not consider impurities, including non-specific protein binding characteristic of Red No 3, at the time these studies were conducted. The latter could potentially and erroneously produce the assay results noted. To reiterate, a single signaling event is not enough to suggest an adverse outcome pathway is implicated.

Additionally, as described elsewhere (Chappell et al., 2020), the changes measured *in vitro* with Red No. 3 are most likely due to methodological artifacts related to the amount of tissue used in incubation systems, more fully described by Mailman et al. (1980). Mailman et al. (1980) found that synaptosomal protein concentration present in the incubation medium significantly influenced the inhibitory effect of Red No. 3 on synaptosomal dopamine uptake in rat brain

preparations. The authors attributed this effect to non-specific interactions with neural membranes. This would be a concern with all the *in vitro* studies cited by OEHHA in Section 4.1.1, starting on page 134, in which the model systems were not validated. Notably, none of the critical studies conducted by Mailman (1987) and Mailman and Lewis (1987), as cited by Chappell et al. (2020), were even considered by OEHHA (no references to these studies were made). The Mailman studies investigated potential factors contributing to such effects upon *in vitro* exposure to Red No. 3. Overall, there is alignment in results from *in vivo* and *in vitro* studies, indicating that Red No. 3 can potentially disrupt neurotransmitter function at high dose levels (>100x ADI *in vitro*, which is an unlikely, if not impossible, target concentration from oral exposure) and that there is a lack of activity for Red No. 3 for thyroid pathway perturbations that may be related to neurobehavioral outcomes. Taken together, there is insufficient evidence supporting an association between exposure to Red No. 3 and neurobehavioral-related key events.

FD&C Red No. 40/Allura Red

Very limited neuro-relevant mechanistic information is available for Red No. 40 to evaluate mechanistic changes as they relate to potential alterations in neuro-relevant pathways that lead to neurobehavioral outcomes. There was no activity in *in vitro* HTS assays reporting neuro-related mechanistic data for Red No. 40. Overall, there was insufficient evidence to support an association between exposure to Red No. 40 and neurobehavioral-related key events.

FD&C Yellow No. 5/Tartrazine

Only limited neuro-relevant mechanistic information is available for Yellow No. 5, most of which is focused on measures of oxidative stress and/or damage to neuronal cells which by itself is not indicative of an adverse neuro-related event. OEHHA (page 168, Section 5.3.5) states that, "Studies of oxidative stress following Yellow No. 5 administration have attributed the brain effects to generation of reactive oxygen species (ROS) by Yellow No. 5 aromatic amine metabolites. Oxidative stress has been reported in other tissues by a number of investigators following Yellow No. 5 administration." However, several studies evaluating oxidative stress in brain tissues were of low quality, the doses administered were high and/or exposure occurred through bolus administration via oral gavage. Further, HTS data demonstrated inactivity for markers of oxidative stress markers in non-neural tissues.

Gentry et al. (submitted) identified inconsistent changes in *in vivo* measures of oxidative stress in the brain, neuronal cell damage, and neurotransmitter levels (e.g., serotonin, dopamine, and GABA), with no corresponding measures in behavioral activity evaluated in most mechanistic studies. For *in vitro* data, only a single serotonin receptor binding (loss of signal) HTS assay was active among otherwise inactive endpoints measured in *in vitro* neuro-relevant mechanistic assays (Chappell et al., 2020). Together, these results do not support biological plausibility between exposure to Yellow No. 5 and neurobehavioral effects. There was insufficient evidence supporting an association between exposure to Yellow No. 5 and neurobehavioral-related key events.

FD&C Yellow No. 6/Sunset Yellow

Only limited neuro-relevant mechanistic information is available for Yellow No. 6. OEHHA (page 168, Section 5.3.5) states, "Yellow No. 6 has been shown to inhibit human cholinesterase and pseudocholinesterase *in vitro* and rat cholinesterase *in vivo*, with a potency lower than some

organophosphate pesticides." OEHHA (page 168, Section 5.3.5) goes on to state, "Studies with sulfanilic acid, a metabolite of Yellow No. 6, [suggest that] it is the active agent for Yellow No. 6 effects on cholinergic systems, as well as for effects on behavior[,] and identify this neurotransmitter system as a potential mechanistic pathway for Yellow No. 6 neurotoxicity." Both OEHHA conclusions are based on one poor-quality *in vivo* study (Osman et al., 2004) and an *in vitro* study (Osman et al., 2002).

None of the *in vivo* studies in which Yellow No. 6 was evaluated met all key quality criteria as noted in Gentry et al. (submitted). However, considering the *in vivo* studies identified in consideration of quality of issues, together with the analysis of *in vitro* mechanistic data described by Chappell et al. (2020), there was limited neuro-relevant mechanistic data for Yellow No. 6. A single study reported decreased acetylcholinesterase activity for Yellow No. 6 and its metabolite sulphanilic acid in dietary exposed rats, while *in vitro* activity reported by Chappell et al., 2020 was limited to serotonergic signaling. Taken together, there was insufficient evidence supporting an association between exposure to Yellow No. 6 and neurobehavioral-related key events.

Conclusions and Recommendations

It is clear from a review of the draft report that OEHHA began with a preliminary conclusion that consumption of synthetic color additives contributes to behavioral effects for children and identified tenuous links to support the agency's preferred outcome. Many of the studies, in our view, do not follow any standard guidelines, nor have they been validated or reproduced for their ascribed effects. Many of the *in vitro* studies are of limited relevance for neuro-related endpoints. Even where limited relevant data is available for some colors, OEHHA nevertheless draws conclusions despite insufficient evidence.

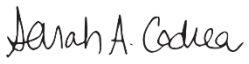
Study reliability (i.e., quality of methods and reporting) was not considered in the OEHHA draft report. Instead, the OEHHA draft report used a broad search strategy to identify studies associated with each color and did not seem to use filters with a clear objective for sorting the animal and/or *in vitro* information for neurobehavioral relevant endpoints. Studies were included in the assessment that other expert bodies excluded based on low quality or relevance. There are also several instances where additional relevant literature was not identified by OEHHA or included in the literature search.

OEHHA concludes that there is no clear evidence of causality but suggests enough "noise" exists to warrant risk management. OEHHA also draws conclusions where contradictions exist. If an increase in effect is shown in one model where another model shows a decrease, the OEHHA draft report indicates a hazard, where most risk assessors would note a lack of consistent response. OEHHA then takes this baseless hazard identification to point to a risk management need. However, a lack of consistent or clear evidence should place the focus on additional research before any short-term risk management measures are even contemplated. Parents who want to avoid serving their children food or providing medications containing FD&C colors can already do so by referring to the ingredient lists on the labels of **all** processed foods or medicines in California, where these colors are clearly labeled as required by the U.S. FDA. There is also no evidence of any public health benefit from additional risk management measures taken elsewhere. In Europe and the U.K., there is no documented evidence that interventions, including a warning label for added azo colors, have impacted neurobehavioral effects in children, including any decrease in the prevalence of ADHD.

The color and packaged goods industries follow good manufacturing practices and have received affirmation of the safety of these colors for the uses allowed by FDA and other global authorities, and at levels typically well below established ADIs. We encourage OEHHA to revisit IACM's public comments provided in February 2019 that provided extensive detail as to how studies conducted by both IACM and the FDA recently confirmed this point. Consumer packaged goods companies already work with their suppliers on opportunities to provide consumers with information on the colors contained in products and will continue to seek pathways to provide consumers with helpful information on how colors are used in their products utilizing existing regulatory structures and communication channels.

Importantly, IACM continues to maintain extensive data on the safety of synthetic colors and highlight new information as it becomes available. This includes sponsorship of research, either in response to regulatory requests, or when the need is identified by industry. We encourage OEHHA to consider our comments in the finalization of its report.

Sincerely,



Sarah Codrea
Executive Director

References

Basak DK, Aylak F, Karadayi N, Gultekin F. 2014. Effects of maternally exposed food coloring additives on laryngeal histology in rats. Journal of Environmental Pathology, Toxicology, and Oncology 33:123-130.

Basak K, Basak PY, Doguc DK, Aylak F, Oguztuzun S, Bozer BM, Gultekin F. 2017. Does maternal exposure to artificial food coloring additives increase oxidative stress in the skin of rats? Human and Experimental Toxicology, 36:1023-1030

Chappell GA, Britt JK, Borghoff S.J. 2020. Systematic assessment of mechanistic data for FDA-certified food colors and neurodevelopmental processes. Food Chem Toxicol:111310

Ceyhan BM, Gultekin F, Doguc DK, Kulac E. 2013. Effects of maternally exposed coloring food additives on receptor expressions related to learning and memory in rats. Food Chem Toxicol 56:145-148.

Dalal A, Poddar M.K. 2009. Short-term erythrosine B-induced inhibition of the brain regional serotonergic activity suppresses motor activity (exploratory behavior) of young adult mammals. Pharmacology, Biochemistry, and Behavior, 92: 574-582.

Dalal A, Poddar M.K. 2010. Involvement of high plasma corticosterone status and activation of brain regional serotonin metabolism in long-term erythrosine-induced rearing motor hyperactivity in young adult male rats. Toxicology Mechanisms and Methods, 20: 287-297.

Doguc DK, Ceyhan BM, Ozturk M, Gultekin F. 2013. Effects of maternally exposed colouring food additives on cognitive performance in rats. Toxicology and Industrial Health, 29: 616-623.

Doguc DK, Aylak F, Ilhan I, Kulac E, Gultekin F. 2015. Are there any remarkable effects of prenatal exposure to food colourings on neurobehaviour and learning process in rat offspring? Nutritional Neuroscience, 18:12-21.

Doguc DK, Deniz F, Ilhan I, Ergonul E, Gultekin F. 2019. Prenatal exposure to artificial food colorings alters nmda receptor subunit concentrations in rat hippocampus. Nutritional Neuroscience,1-11. Epub ahead of print.

EFSA AFC (European Food Safety Authority and Additives, Flavourings, Processing Aids and Contact Material). 2008. Assessment of the results of the study by McCann et al. (2007) on the effect of some colours and sodium benzoate on children's behavior. EFSA Journal. 660:1-54. https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/j.efsa.2008.660.

EFSA. 2009a. Scientific Opinion on the re-evaluation of Allura Red AC (E 129) as a food additive. EFSA Journal 7(11).

EFSA. 2009b. Scientific Opinion on the re-evaluation of Sunset Yellow FCF (E 110) as a food additive. EFSA Journal 7(11).

EFSA. 2009c. Scientific Opinion on the re-evaluation Tartrazine. EFSA Journal 7(11).

EFSA. 2010. Scientific Opinion on the re-evaluation of Brilliant Blue FCF (E 133) as a food additive. EFSA Journal 8(11).

EFSA. 2011. Scientific Opinion on the re-evaluation of Erythrosine (E 127) as a food additive. EFSA Journal 9(1).

EFSA. 2014. Scientific Opinion on the re-evaluation of Indigo Carmine (E 132) as a food additive. EFSA Journal 12(7).

Erickson JC, 3rd, Lauron F. 1960. Vasopressor effect of indigo carmine. A preliminary report. Guthrie Clin Bull 30:30-35.

Erickson ZT, Falkenberg EA, Metz GA. 2014. Lifespan psychomotor behaviour profiles of multigenerational prenatal stress and artificial food dye effects in rats. PLoS One 9:e92132

FDA (United States Food and Drug Administration). 2011. Center for Food Safety and Applied Nutrition, Food Advisory Committee Meeting (March 30-31, 2011), Transcript: March 31, 2011 Food Advisory Committee Meeting, Page 325. https://wayback.archive-it.org/org-1137/20170406211705/https://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/FoodAdvisoryCommittee/UCM255119.pdf (Accessed 12 November 2020)

Gao Y, Li C, Shen J, Yin H, An X, Jin H. 2011. Effect of food azo dye tartrazine on learning and memory functions in mice and rats, and the possible mechanisms involved. Journal of Food Science, 76: T125-T129.

Gentry R, Greene T, Chappell G, Borghoff S, Yang C, Rathman J, Ribeiro JV, Hobocienski B, Mostrag A, Rodricks J, Clewell H. (submitted). Integration of evidence to evaluate the potential for neurobehavioral effects following exposure to USFDA- approved food colors. Food and Chemical Toxicology.

Goldenring JR, Wool RS, Shaywitz BA, Batter DK, Cohen DJ, Young JG, et al. 1980. Effects of continuous gastric infusion of food dyes on developing rat pups. Life Sci 27:1897-1904.

Jo YY, Lee MG, Yun SY, Lee KC. 2013. Profound hypotension after an intradermal injection of indigo carmine for sentinel node mapping. J Breast Cancer 16:127-128.

Judson R, Houck K, Martin M, Richard AM, Knudsen TB, Shah I, et al. 2016. Analysis of the effects of cell stress and cytotoxicity on in vitro assay activity across a diverse chemical and assay space. Toxicol Sci 153:409.

Kantor MA, Trout JR, Lachance PA. 1984. Food dyes produce minimal effects on locomotor activity and Vitamin B-6 levels in postweanling rats. The Journal of Nutrition, 114:1402-1412.

Kavale, K. A., & Forness, S. R. 1983. Hyperactivity and Diet Treatment: A Meta-Analysis of the Feingold Hypothesis. Journal of Learning Disabilities, 16(6), 324–330. https://doi.org/10.1177/002221948301600604

Lee KH, Baek DJ, Jeon SY. 2015. Repetitive severe hypotension induced by indigo carmine. J Anesth 29:156-156.

Lok, K.Y.W., Chan, R.S.M., Lee, V.W.Y., Leung, P.W., Leung, C., Leung, J. and Woo, J. 2013. Food additives and behavior in 8- to 9- year-old children in Hong Kong: A randomized, double-blind, placebo-controlled trial. Journal of Developmental & Behavioral Pediatrics. 34(9):642-650.

Mailman, R.B., 1987. Mechanisms of CNS injury in behavioral dysfunction. Neurotoxicol Teratol 9, 417-426.

Mailman RB, Ferris RM, Tang FL, Vogel RA, Kilts CD, Lipton MA, et al. 1980. Erythrosine (Red No. 3) and its nonspecific biochemical actions: What relation to behavioral changes? Science (New York, NY) 207:535-537.

Mailman, R.B., Lewis, M.H., 1987. Neurotoxicants and central catecholamine systems. Neurotoxicology 8, 123-139.

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

Nigg, J.T., Lewis, K., Edinger, T. and Falk, M. 2012. Meta-analysis of attention-deficit/hyperactivity disorder or attention-deficit/hyperactivity disorder symptoms, restriction diet, and synthetic food colors. J Am Acad Child Adolesc Psychiatry. 51:86-97. Doi: 10.1016/j.jaac.2011.10.015.

Noorafshan A, Hashemi M, Karbalay-Doust S, Karimi F. 2018. High dose allura red, rather than the ADI dose, induces structural and behavioral changes in the medial prefrontal cortex of rats and taurine can protect it. Acta Histochemica, 120: 586-594.

OEHHA. 2020. Health Effects Assessment: Potential Neurobehavioral Effects of Synthetic Food Dyes in Children. California Environmental Protection Agency, Office of Environmental Health Hazard Assessment, Children's Environmental Health Center. Public Review Draft.

Osman MY, Sharaf IA, el-Rehim WMA, el-Sharkawi AM. 2002. Synthetic organic hard capsule colouring agents: *In vitro* effect on human true and pseudo-cholinesterases. Br J Biomed Sci 59:212-217.

Osman MY, Sharaf IA, Osman HMY, El-Khouly ZA, Ahmed EI. 2004. Synthetic organic food colouring agents and their degraded products: Effects on human and rat cholinesterases. Br J Biomed Sci 61:128-132.

Pelsser, L.M., Frankena, K., Toorman, J. and Pereira, R.R. 2017. Diet and ADHD, Reviewing the evidence: A systematice review of meta-analyses of double-blind placebo controlled trials evaluating the efficacy of diet interventions on the behavior of children with ADHD, 2014, PLOS One, January 25, 2017. Doi:10.137/journal.pone.0169277 http://www.readcube.com/articles/10.1371/journal.pone.0169277

Rafati A, Nourzei N, Karbalay-Doust S, Noorafshan A. 2017. Using vitamin e to prevent the impairment in behavioral test, cell loss and dendrite changes in medial prefrontal cortex induced by tartrazine in rats. Acta Histochemica, 119: 172-180.

Reisen CA, Rothblat LA. 1986. Effect of certified artificial food coloring on learning and activity level in rats. Neurobehavioral Toxicology and Teratology 8:317-320.

Schab, D.W. and Trinh, N.H. 2004. Do artificial food colors promote hyperactivity in children with hyperactive syndromes? A meta-analysis of double-blind placebo-controlled trials. J Dev Behav Pediatr. 25(6):423-34. https://www.ncbi.nlm.nih.gov/pubmed/15613992

Shaywitz BA, Goldenring JR, Wool RS. 1979. Effects of chronic administration of food colorings on activity levels and cognitive performance in developing rat pups treated with 6-hydroxydopamine. Neurobehav Toxicol 1:41-47.

Sobotka TJ, Brodie RE, Spaid SL. 1977. Tartrazine and the developing nervous system of rats. Journal of Toxicology and Environmental Health, 2: 1211-1220.

Sonuga-Barke, E.J.S., Brandeis, D., Cortese, S., Daley, D., Ferrin, M., Holtmann, M., Stevenson, J., Danckaerts, M., van der Oord, S., Döpfner, M., Dittman, R.W., Simonoff, E., Zuddas, A., Banaschewski, T., Buitelaar, J., Coghill, D., Hollis, C., Konofal, E., Lecendreux, M., Wong, I.C.K., Seargent, J. and on behalf of the European ADHD Guidelines Group. 2013. Nonpharmacological interventions for ADHD: Systematic Review and meta-analysis of randomized control trials of dietary and psychological treatments. Am J Psychiatry. 170:275-289.

Stevenson, J., Buitelaar, J., Cortese, S., Ferrin, M., Konofal, E., Lecendreux, M., Simonoff, E., Wong, I.C.K., Sonuga-Barke, E. and on behalf of the European ADHD Guidelines Group. 2014. Research Review: The role of diet in the treatment of attention-deficit/hyperactivity disorder – an appraisal of the evidence on efficacy and recommendations on the design of future studies. Journal of Child Psychology and Psychiatry. 55:5:416-427. Doi: 10.1111/jcpp12215.

Tanaka T. 1994. Reproductive and neurobehavioral effects of Allura red ac administered to mice in the diet. Toxicology, 92: 169-177.

Tanaka T. 1996. Reproductive and neurobehavioral effects of sunset yellow FCF administered to mice in the diet. Toxicology and Industrial Health, 12(1): 69-79.

Tanaka T. 2001. Reproductive and neurobehavioural toxicity study of erythrosine administered to mice in the diet. Food and Chemical Toxicology, 39(5): 447-454.

Tanaka T. 2006. Reproductive and neurobehavioural toxicity study of tartrazine administered to mice in the diet. Food and Chemical Toxicology, 44 (2): 179-187.

Tanaka T, Takahashi O, Oishi S, Ogata A. 2008. Effects of tartrazine on exploratory behavior in a three-generation toxicity study in mice. Reproductive Toxicology, 26: 156-163.

Tanaka T, Takahashi O, Inomata A, Ogata A, Nakae D. 2012. Reproductive and neurobehavioral effects of brilliant blue FCF in mice. Birth Defects Research. Part B Developmental Reproductive Toxicology, 95: 395-409.

Vorhees C, Butcher R, Brunner R, Wootten V, Sobotka T. 1983a. A developmental toxicity and psychotoxicity evaluation of FD&C red dye #3 (erythrosine) in rats. Archives of Toxicology, 53: 253-264.

Vorhees C, Butcher R, Brunner R, Wootten V, Sobotka T. 1983b. Developmental toxicity and psychotoxicity of FD&C red dye no. 40 (Allura red ac) in rats. Toxicology, 28: 207-217.

Wang J, Jackson DG, Dahl G. 2013. The food dye FD&C Blue No. 1 is a selective inhibitor of the ATP release channel panx1. J Gen Physiol 141:649-656.

WHO/FAO (World Health Organization/Food and Agricultural Organization). 2017a. Eighty-fourth report of the Joint FAO/WHO Expert Committee on Food Additives. Evaluation of certain food additives.

WHO/FAO. 2017b. Eighty-second report of the Joint FAO/WHO Expert Committee on Food Additives. Evaluation of certain food additives.

WHO/FAO. 2019. Eighty-sixth report of the Joint FAO/WHO Expert Committee on Food Additives. Evaluation of certain food additives.

Yang J, Liu R, Lu F, Xu F, Zheng J, Li Z, et al. 2019. Fast green FCF attenuates lipopolysaccharide-induced depressive-like behavior and downregulates tlr4/myd88/nf-κb signal pathway in the mouse hippocampus. Front Pharmacol 10:501.

October 16, 2020

Appendix A

Review of Animal Toxicity Data in OEHHA's Draft Report: Potential Neurobehavioral Effects of Synthetic Food Dyes in Children

The following are the reports and publications discussed in this critical review conducted by Ramboll:

- OEHHA (Office of Environmental Health Hazard Assessment) Draft Report: Sections 3.1, 3.2, 3.3, 3.4;
- Gentry et al. (submitted); and
- Selected quality scoring from the ToxStrategies report "Assessment of Available Neuro-Related Mechanistic Data and Behavioral Activity for Seven FDA-Approved Synthetic Colors", Appendix B of IACM comment submission.

General Comments

While a study quality assessment was conducted for the human studies, the animal toxicity studies that were considered in the OEHHA (2020) Draft Report were not evaluated by OEHHA using any type of published methods for study quality or reliability. In contrast, the single dye and mixture animal toxicity studies included in Ramboll's critical review (Gentry et al., submitted) were assessed for study quality and reliability using the ARRIVE guidelines (Kilkenny et al. 2010, 2012) and the single dye studies were also assessed using Klimisch Guidelines (Klimisch et al. 1997). Further, many of these studies were assessed independently using the Science in Risk Assessment and Policy (SciRAP) tool (SciRAP 2020) and the Toxicological Data Reliability Assessment Tool (ToxRTool) (Schneider et al. 2009) as noted in Appendix B.

The studies that initially appeared to be the focus of OEHHA's quantitative risk assessment (QRA) for individual colorants, based on the 2019 symposium presentation, were the studies conducted predominantly in the Tokyo Metropolitan Laboratory of Public Health (Tanaka 1994, 1996, 2001, 2006; Tanaka et al. 2008, 2012) in addition to a few others (Doguc et al. 2013, 2015, 2019; Ceyhan et al. 2013; Başak et al. 2014, 2017). An additional 13 animal studies were considered for quantitative risk assessment in OEHHA's 2020 draft report. OEHHA (2020) states that primarily developmental studies with oral administration and neurobehavioral endpoints were included in the review; however, OEHHA (2020) also included studies when dye exposure was only during adulthood in the animals. These studies do not address the charge. In addition, if non-behavioral endpoints were included in the study, they were also reviewed. Therefore, the studies under consideration were broader than the original OEHHA presentation implied.

In regard to the Tanaka studies, OEHHA noted:

> *It is tempting to compare across the food dyes for developmental neurotoxicity using the studies from the Tokyo lab. However, these data alone are not adequate to conclude that some dyes are more toxic than others by comparing across studies. While the design of the studies is similar there were changes in procedures, equipment and statistical analysis over time (1994-2012). An effort was made by the laboratory to test comparable dose ranges in this set of studies by using multiples of the JECFA ADIs for the low dose. However, different multiples (100X, 10X) were used and the JECFA Red No. 3 ADI is notable for being two orders of magnitude lower than the FDA ADI. In addition, reproductive and developmental toxicity varied across studies and could influence the later behavioral assessments.*

- In reviewing the Tanaka studies, Ramboll (2020) noted that the comparison of the results across all of the Tokyo Metropolitan Laboratory of Public Health studies can be conducted to evaluate the implication for a quantitative risk assessment, as the chemical composition of the colors would suggest similar results might be expected. The results of an integrated review found a lack of evidence of consistent or sustained adverse changes within studies or across studies with individual colors.
- A statistical evaluation of the results for specific endpoints reported for control groups across the individual studies from this laboratory, where possible, indicated statistically significant differences in endpoints in the control groups (absent any treatment) over time. This demonstrated significant variability in the background response to these neurobehavioral assays within this laboratory. Further, combination of control results for comparison to comparable treatment groups resulted in the loss of statistical significance in selected responses, because of the significant variability in control responses over time.
- OEHHA notes that the JECFA ADI for Red No. 3 is two orders of magnitude lower than the FDA ADI, but its actually only one order of magnitude lower.

The strengths and limitations of the 25 studies included in the OEHHA (2020) report are summarized in the following bullets and table, along with counterpoints and recommendations regarding consideration and/or exclusion of each of the 25 studies for risk assessment. Ramboll considered several factors to be imperative when including a study for consideration in the QRA. First, the source and/or purity of the colors being tested in each study should be reported in order to ensure the test substance is of sufficient quality for assessing the potential health effects. Second, in order to assess dose-response, the study had to include multiple dose groups, as well as a negative control group. Third, a study must be of adequate quality and reliability having an ARRIVE score in the high to mid-range (> 50%) and a Klimisch score of 1 (reliable without restrictions) or 2 (reliable with restrictions).

Allura Red/Red No. 40

- Overall, three studies were identified for review (Tanaka 1994; Vorhees et al. 1983b; Noorafshan et al. 2018). Tanaka (1994) was identified in the OEHHA (2019) presentation, while Vorhees et al. (1983b) and Noorafshan et al. (2018) were identified in the OEHHA (2020) report and in the literature searches conducted in Gentry et al. (submitted) to identify neuro-related mechanistic and behavioral *in vivo* data.
- Based on the results of Noorafshan et al. (2018), OEHHA proposed a LOAEL for Allura Red of 7 mg/kg/day, which is the same as the USFDA and JECFA ADI of 7 mg/kg/day, indicating that the ADI may not be protective against neurobehavioral effects. OEHHA concluded that if the results of Noorafshan et al. (2018) were to be used as the basis for setting an ADI the resulting ADI would be 100 to 1000 fold lower than the existing ADI, depending on the method used to derive the point of departure.

- Ramboll concluded that the results of Noorafshan et al. (2018) would not be considered for any quantitative evaluation of neurobehavioral effects in children because the study was conducted in adult rats. In addition, the Radial Arm Maze test conducted by the authors to assess learning and memory included a forced food restriction prior to testing so the animals would lose 15% of body weight. Food-based reward was used to encourage the animals to complete the test. The authors offered no details on the potential effects of the forced food restriction and reduction of body weight on the bioavailability of the test substance, the effect of 15% body weight reduction on the overall outcome compared to normal weight control animals, or the relevancy of the forced food restriction to normal human behavior.
- The study conducted by Vorhees et al. (1983b) would not be considered for quantitative risk assessment because of limited data presented for quantitative risk assessment and specific study quality and reliability criteria were not met. Specifically, the test substance purity, and the housing and feeding conditions of animals were not reported in the study (i.e. ToxRTool score of 3 – not reliable).
- Ramboll considered the results of Tanaka (1994) to be adequate for quantitative risk assessment and concluded that the study did not provide strong or consistent evidence of effects, including following doses higher than JECFA of USFDA NOAELs.

Sunset Yellow/Yellow No. 6

- Ramboll and OEHHA only identified one study for Sunset Yellow (Tanaka 1996).
- OEHHA concluded that while some neurobehavioral effects in offspring were reported for preweaning development and maze learning in Tanaka (1996), conclusions could not be drawn due to the statistical approach and varying group sizes in the study. OEHHA stated that a NOAEL without a LOAEL in the same study is not suitable for risk assessment.
- Ramboll considered the results of Tanaka (1996) to be adequate for quantitative risk assessment and concluded that the study did not provide strong or consistent evidence of effects, including following doses higher than the JECFA or USFDA NOAELS.

Erythrosine/Red No. 3

- Overall, four studies were identified for review (Tanaka 2001; Vorhees et al. 1983a; Dalal and Poddar 2009, 2010). Tanaka (2001) was identified in the OEHHA (2019) presentation, while Vorhees et al. (1983a) and Dalal and Poddar (2009, 2010) were identified in the OEHHA (2020) report and in the literature searches conducted in Gentry et al. (submitted) to identify *in vivo* neuro-related mechanistic and behavioral data.
- OEHHA (2020) concluded that the Tanaka (2001) study had a NOAEL of 28 mg/kg/day which is lower than the USFDA NOAEL of 250 mg/kg/day based on the observation of "distended cecum" in rats, indicating that the ADI should be protective against neurobehavioral effects.
- Gentry et al. (submitted) concluded that the results of Dalal and Poddar (2009, 2010) would not be considered for any quantitative evaluation of neurobehavioral effects in children because both studies were conducted in mature adult rats. In addition, the specifications for the test substance reported in Dalal and Poddar (2009, 2010), Erythrosine B, purchased from Sigma is not comparable to FDA specification for FD&C 3 under 21CFR74.303. Therefore, critical information about purity of the color additive, as well as the extent of critical impurities and reaction by-products in the color additive for which maximum limits have been set by regulation are unknown. For example, erythrosine is required to have a purity of 87%, and regulations specifically restrict the presence of other iodinated compounds like sodium iodide,

(≤0.4%), triiodoresorcinol (≤ 0.2 %); 2(2',4'-Dihydroxy-3', 5'-diiodobenzoyl) benzoic acid (≤ 0.2 %); monoiodofluoresceins (≤ 1.0 %); other lower iodinated fluoresceins, (≤ 9.0 %).

- Gentry et al. (submitted) concluded that the study conducted by Vorhees et al. (1983a) would not be considered for quantitative risk assessment because specific study quality and reliability criterial were not met (i.e. ToxRTool score of 3 – not reliable).
- Gentry et al. (submitted) conducted a statistical evaluation to investigate the potential variability in responses in control groups over time. Comparable control data for comparison were reported across the Tanaka publications and included Tanaka (2001). Overall, the results of the comparison of control groups indicate the model systems used by the Tokyo Metropolitan Laboratory of Public Health (Tanaka 2001) have significant variability over time for the majority of endpoints reported to be statistically signficantly different from controls, hindering the ability to draw any meaningful conclusions from the reported results.
- Ramboll considered the results of Tanaka (2001) to be adequate for quantitative risk assessment and concluded that the study did not provide strong or consistent evidence of effects, including following doses higher than the JECFA or USFDA NOAELs.

Tartrazine/Yellow No. 5

- Overall, 5 studies were identified for review (Tanaka 2006; Tanaka et al. 2008; Sobotka et al 1977; Gao et al. 2011; Rafati et al. 2017). Tanaka (2006) and Tanaka et al. (2008) were identified in the OEHHA (2019) presentation, while Sobotka et al (1977); Gao et al. (2011) and Rafati et al. (2017) were identified in the OEHHA (2020) report and in the literature searches conducted by Gentry et al. (submitted) to identify neuro-related mechanistic and behavioral data.
- OEHHA (2020) concluded that the lowest NOAELs identified of 175 mg/kg/day in mice and 125 mg/kg/day in rats (Gao et al. 2011) were lower than the NOAELs identified by FDA of 500 mg/kg/day in dogs and 1000 mg/kg/day in rats, and 1.5 % in diet estimated at 2250 mg/kg/day in mice.
- Gentry et al. (submitted) concluded that the results of Sobotka et al. (1977) did not provide strong or consistent evidence of effects. The study conducted by Rafati et al. (2017) would not be considered for quantitative risk assessment because the study did not provide strong or consistent evidence of effects. In addition, the advanced age of rats is a confounder in the memory tests conducted as age alone has been shown to affect radial arm maze test performance. Finally, the results of Gao et al. (2011) were not considered because specific study quality and reliability criterial were not met (i.e. ToxRTool score of 3 – not reliable).
- Gentry et al. (submitted) conducted a statistical evaluation to investigate the potential variability in responses in control groups over time. Comparable control data for comparison were reported across the Tanaka publications and included Tanaka (2006, Tanaka et al. 2008). Overall, the results of the comparison of control groups indicate the model systems used by the Tokyo Metropolitan Laboratory of Public Health (Tanaka 2006; Tanaka et al. 2008) have significant variability over time for the majority of endpoints reported to be statistically signficantly different from controls, hindering the ability to draw any meaningful conclusions from the reported results.
- Gentry et al. (submitted) considered the results of Tanaka (2006) and Tanaka et al. (2008) for quantitative risk assessment and concluded that the studies did not provide strong or consistent evidence of effects, including following doses higher than the JECFA or USFDA NOAELs.

Brilliant Blue/Blue No. 1

- Only one study was identified for Brilliant Blue (Tanaka et al. 2012).
- OEHHA (2020) concluded the LOAEL from Tanaka et al. (2012) was 127 mg/kg/day. However, Gentry et al. (submitted) concluded that the results of Tanaka (20121996) provide only limited evidence of isolated changes that are not sustained over time despite chronic exposure to the test color and are limited to one or two endpoints measured from a large number intended to evaluate neurobehavioral domains.
- Gentry et al. (submitted) conducted a statistical evaluation to investigate the potential variability in responses in control groups over time. Comparable control data for comparison were reported across the Tanaka publications and included Tanaka (2012). Overall, the results of the comparison of control groups indicate the model systems used by the Tokyo Metropolitan Laboratory of Public Health (Tanaka 2012) have significant variability over time for the majority of endpoints reported to be statistically signficantly different from controls, hindering the ability to draw any meaningful conclusions from the reported results.
- Ramboll considered the results of Tanaka et al. (2012) to be adequate for quantitative risk assessment and concluded that the study did not provide strong or consistent evidence of effects, even following doses higher than ADI.

AFC Mixtures

- Overall, 11 studies were identified by Gentry et al. (submitted)for review (Doguc et al. 2013, 2015, 2019; Basak et al. 2014, 2017; Ceyhan et al. 2013; Kantor et al. 1984; Shaywitz et al. 1979; Goldenring et al. 1980; Reisen and Rothblat 1986; Erickson et al. 2014). Doguc et al. (2013, 2015, 2019); (Basak et al. 2014, 2017) and Ceyhan et al. (2013) were identified in the OEHHA (2019) presentation while the remaining citations were cited in the OEHHA (2020) report.
- OEHHA did not consider mixture studies in risk characterization.
- Ramboll concluded that the mixture studies included significant uncertainty in the potential application of the study results in a quantitative risk assessment. Ultimately, because individual colors were not assessed separately, there is no way of determining whether the effects were from one colorant or a combination of two or more colorants or any other potential confounder or impurities, especially if the purity of test substance was not reported.

Overall Conclusions

Based on the results of the animal studies reviewed, OEHHA (2020) concluded:

> "a number of animal studies of single synthetic food dyes and a dosing regimen that included *in utero*, postnatal and juvenile exposures found evidence of effects on behavior in the offspring. A handful of these studies observed effects at doses lower than the NOAELs used by the FDA to derive their ADIs."

In studies conducted in adult animals OEHHA (2020) concluded:

> "Almost all the studies in mature animals that measured behavioral changes and/or changes in the brain found effects of the synthetic food dyes at doses lower than the NOAELs used by the US FDA for the derivation of the ADIs. A number of these studies observe effects on behavior in animals at doses close to or even lower than the existing FDA ADIs."OEHHA noted that when all single dye and mixture studies were considered, effects on activity, and learning and memory were reported in both young and adult animals.

Based on Gentry et al. (submitted), none of the studies evaluated provided strong or consistent evidence of effects.

Table 1. Comparison of Ramboll and OEHHA reviews of Animal Studies Considered in OEHHA (2020) for Quantitative Risk Assessment			
Study citation	**Ramboll Strengths/ Limitations of study**	**OEHHA Strengths/ Limitations of study**	**Ramboll counterpoints to OEHHA arguments**
Allura Red/Red No. 40			
Tanaka 1994	Strengths: -Included three treated groups plus negative controls. -Exposure via diet -Two-generation reproductive/developmental study -Test purity > 85% -Exposure from 5 weeks of age in F0 to 9 weeks of age in F_1 -Analysis included test for functional and behavioral outcomes, exploratory behavior, and learning and memory. -ToxRTool score: 1 (Reliable without restrictions) Limitations: -Study relied upon the mouse as the animal model instead of the rat, the rat being the recommended species in the OECD/ICH test guidelines. -Limited number of animals per dose groups decreasing study quality and statistical power of the results. -Doses in the study were magnitudes higher than the ADIs and higher than expected in the human population. -ARRIVE score was below median (44%)	Strengths: -Activity assessed with an automated test using a doughnut-shaped arena, unlike the Open Field test of the early studies which used a rectangular arena. -Learning & memory were tested with a water maze, unlike the shock-motivated tests used in earlier studies. -Tanaka studies are valuable for their dose-response designs and extensive data reporting Limitations: -The group size and power varied from 7-10 mice per group making the power of these studies inadequate given the sensitivity of many measures. -Pups were pooled across litters for statistical analysis of the preweaning data, so that litter-based statistics were not performed. -Groups were not compared in the maze data analysis for most dyes; statistical analysis was within groups across trials only.	Ramboll agrees with the limitations of the study as noted by OEHHA. Ramboll considered the results of Tanaka 1994 for quantitative risk assessment and concluded that the study did not provide strong or consistent evidence of effects. The results of Tanaka (1994) provide evidence of isolated changes that are not sustained over time despite chronic exposure to the test color, and are limited to one or two endpoints measured from a large number intended to evaluate neurobehavioral domains and are not observed consistently in males and females. Furthermore, the study was limited by low study quality due to the small number of animals per dose group which decreases the statistical power of the results and limits the value of the study results for dose-response.
Vorhees et al. 1983b	Strengths: -Three treatment groups plus control -Exposure via the diet -Study used preferred animal model recommended by OECD for neurobehavioral /developmental toxicity studies. -Exposed two weeks prior to breeding through post-natal day 110. Limitations: -Test substance purity not reported. -Study conducted 37 years ago. ARRIVE score was low range (50%). -ToxRTool score: 3 (Not reliable). -Data are not provided all of the endpoints indicated to be significantly changed that would allow for quantitative evaluation.	Strengths: -Test protocols and statistical analysis state-of-the-art for the period when the study was done. -Interpretation of the data did not account for the extensive general developmental toxicity seen at the doses used. -Postnatal mortality, offspring growth restriction and delayed vaginal opening could be valid endpoints for general developmental toxicity risk assessment. Limitations: -Use of the $p<0.01$ statistical significance criterion is relevant because it requires a larger effect size to reach statistical significance. -Sensitivity of the learning and memory tests hard to determine without data from control group.	Ramboll would not include Vorhees et al. (1983b) in any quantitative evaluation because specific study quality and reliability criterial were not met. Further, all of the data needed for a quantitative evaluation are not provided. Regarding study quality, the test substance purity, and the housing and feeding conditions of animals were not reported in the study. (i.e. ToxRTool score of 3 – not reliable). In considering the available data, the significant dose-related behavioral effects were limited to two endpoints (decrease in running wheel activity, significant increase in postweaning open-field rearing activity (opposite impacts on activity). The remaining endpoints assessed were sporadic, with a lack of dose-dependence reported. The authors acknowledge the inconsistency in endpoints in that one is associated with hypoactivity and the other hyperactivity.

Table 1. Comparison of Ramboll and OEHHA reviews of Animal Studies Considered in OEHHA (2020) for Quantitative Risk Assessment			
Study citation	**Ramboll Strengths/ Limitations of study**	**OEHHA Strengths/ Limitations of study**	**Ramboll counterpoints to OEHHA arguments**
Noorafshan et al. 2018	Strengths: -Behavioral testing including Novel object recognition and eight-arm radial maze test with two phases of adaptation were performed -Study also included histological examination of the brains -ARRIVE score was high (78%) -ToxRTool score was 2 (Reliable with restrictions) Limitations: -Exposure via oral gavage as opposed to dietary -No developmental outcomes were tested, only neurobehavioral in adult male rats. -Only one sex (males) tested -Test substance source was reported but test substance purity was not specified -Only two concentrations plus negative control tested.	Strengths: -The behavioral tasks in this more recent study is well-known for sensitivity to neurotoxins. -The gavage administration may result in higher internal doses and less binding to fiber in the intestines than the diet administration route used in most early dye toxicity studies. -The detailed examination of brain histomorphology helps provide biological plausibility for the behavioral effects.	Ramboll would not include Noorafshan et al. (2018) in any quantitative evaluation neurobehavioral effects in children because the study was conducted in mature adult rats. In addition, the Radial Arm Maze test for learning and memory was conducted under altered feeding regimen to force a body weight reduction of 15%. It is unknown what the effect of the forced-food restriction could have had on the bioavailability of the test substance or the results of the learning and memory test in general. The findings are also questionable due to the lack of information on test substance purity and inconsistency with results from multi-generation studies in animals (Tanaka 1994).
Sunset Yellow/Yellow No. 6			
Tanaka 1996	Strengths: -The study included three treated groups plus negative controls. -Two-generation reproductive/developmental study -Exposure via the diet -Test purity > 85% -Exposure from 5 weeks of age in F0 to 9 weeks of age in F1 -Analysis included test for functional and behavioral outcomes, exploratory behavior, and learning and memory. ToxRTool score: 1 (Reliable without restrictions) Limitations: -Study relied upon the mouse as the animal model instead of the rat, the rat being the recommended species in the OECD/ICH test guidelines. -Limited number of animals per dose groups decreasing study quality and statistical power of the results.	Strengths: -Activity assessed with an automated test using a doughnut-shaped arena, unlike the Open Field test of the early studies which used a rectangular arena. -Learning & memory were tested with a water maze, unlike the shock-motivated tests used in earlier studies. -Tanaka studies are valuable for their dose-response designs and extensive data reporting Limitations: -The group size and power varied from 7-10 mice per group making the power of these studies inadequate given the sensitivity of many measures. -Pups were pooled across litters for statistical analysis of the preweaning data, so that litter-based statistics were not performed. -Groups were not compared in the maze data analysis for most dyes; statistical analysis was within groups across trials only.	Ramboll agrees with the limitations of the study as noted by OEHHA. Ramboll considered the results of Tanaka (1996) for quantitative risk assessment and concluded that the study did not provide strong or consistent evidence of effects. The results of Tanaka (1996) provide only limited evidence of isolated changes that are not sustained over time despite chronic exposure to the test color and are limited to one or two endpoints measured from a large number intended to evaluate neurobehavioral domains. Furthermore, the study was limited by low study quality due to the small number of animals per dose group which decreases the statistical power of the results and limits the value of the study results for dose-response.

Table 1. Comparison of Ramboll and OEHHA reviews of Animal Studies Considered in OEHHA (2020) for Quantitative Risk Assessment			
Study citation	**Ramboll Strengths/ Limitations of study**	**OEHHA Strengths/ Limitations of study**	**Ramboll counterpoints to OEHHA arguments**
	-Doses in the study were magnitudes higher than the ADIs and higher than expected in the human population. -ARRIVE score was below the median (43%)		
Erythrosine/Red No. 3			
Tanaka 2001	<u>Strengths:</u> -The study included three treated groups plus negative controls. -Two-generation reproductive/developmental study -Exposure via the diet -Test purity > 85% -Exposure from 5 weeks of age in F0 to 9 weeks of age in F_1 -Analysis included tests for functional and behavioral outcomes, exploratory behavior, and learning and memory. -ToxRTool score: 1 (Reliable without restrictions) <u>Limitations:</u> -Study relied upon the mouse as the animal model instead of the rat, the rat being the recommended species in the OECD/ICH test guidelines. -Limited number of animals per dose groups decreasing study quality and statistical power of the results. -Doses in the study were magnitudes higher than the ADIs and higher than expected in the human population. -ARRIVE score was mid-range (68%)	<u>Strengths</u>: -Activity assessed with an automated test using a doughnut-shaped arena, unlike the Open Field test of the early studies which used a rectangular arena. -Learning & memory were tested with a water maze, unlike the shock-motivated tests used in earlier studies. -Tanaka studies are valuable for their dose-response designs and extensive data reporting <u>Limitations:</u> -The group size and power varied from 7-10 mice per group making the power of these studies inadequate given the sensitivity of many measures. -Pups were pooled across litters for statistical analysis of the preweaning data, so that litter-based statistics were not performed. -Groups were not compared in the maze data analysis for most dyes; statistical analysis was within groups across trials only.	Ramboll agrees with the limitations of Tanaka (2001) as noted by OEHHA. Ramboll considered the results of Tanaka 2001 for quantitative risk assessment and concluded that the study did not provide strong or consistent evidence of effects. The results and conclusions of Tanaka (2001) are based on changes in a single endpoint for a battery of endpoints to assess exploratory behavior and observed at a single time point, but not at other timepoints, despite continued exposure to the test color. Results from the exploratory behavior test were inconsistent and in some cases the changes could indicate an improvement in exploratory behavior and not an adverse treatment-related effect. Furthermore, the results of the comparison of control groups indicate the model systems used by the Tokyo Metropolitan Laboratory of Public Health (Tanaka 2001) have significant variability over time, hindering the ability to draw any meaningful conclusions from the reported results. Finally, the study was limited by low study quality due to the small number of animals per dose group which decreases the statistical power of the results and limits the value of the study results for dose-response.

Table 1. Comparison of Ramboll and OEHHA reviews of Animal Studies Considered in OEHHA (2020) for Quantitative Risk Assessment			
Study citation	**Ramboll Strengths/ Limitations of study**	**OEHHA Strengths/ Limitations of study**	**Ramboll counterpoints to OEHHA arguments**
Vorhees et al. 1983a	Strengths: -Three treatment groups plus control -Two-generation reproductive/developmental study -Exposure via the diet Study used preferred animal model recommended by OECD for neurobehavioral /developmental toxicity studies. -Test substance purity reported to be 91% -Animals were exposed two weeks prior to breeding through post-natal day 110. Limitations: -Authors did not report adequate information regarding doses administered, number of animals exposed and appropriateness of study design. -Results cannot be evaluated due to a lack of data. - ARRIVE score was below median (45%) -ToxRTool score: 3 (Not reliable)	Strengths: -Provided several different activity and learning and memory tests Used an ANOVA analysis across all dye groups. Limitations: -Sensitivity of the analysis is reduced by the use of $p<0.01$ as the threshold for statistical significance. -Sensitivity of the learning and memory tests cannot be evaluated because no data were shown. -Two experiments were conducted by replicability but differed extensively. -While the authors interpret an absence of toxicity due to lack of replicability and linear dose response trends, dye effects were demonstrated.	Ramboll would not include Vorhees et al. (1983a) in any quantitative evaluation because specific study quality and reliability criterial were not met including specific data related to doses administered, number of animals exposed and specification on study design. (i.e. ToxRTool score of 3 – not reliable). The authors report that although tests of behavioral development and adult performance showed several statistically significant changes, the non-dose-dependent nature of these effects, combined with their lack of replicability across experiments, renders these findings unconvincing as evidence. While OEHHA concluded the lowest dose tested was the LOAEL (a dose higher than the FDA ADI), this is inconsistent with the authors conclusions for behavioral effects.
Dalal and Poddar 2009	Strengths: -Four treatment groups plus control -Test substance purity reported to be 90% -ToxRTool score of 1 (Reliable without restrictions) Limitations: -Single exposure study in adult rats. -Only one sex (males) tested. -Exposure via oral gavage. -No developmental outcomes were tested, only neurobehavioral in mature adult male rats. -Only one sex (males) tested -Test substance source was reported but test substance purity was not specified -ARRIVE score was mid-range (70%)	Strengths: -Well-developed hypothesis concerning mechanism, replication of the main effect on behavior, and three to four doses with graphic illustration of dose response. -The statistical analysis (ANOVA with post hocs) was appropriate. Limitations: -Although the effect is transient, which reduced its toxicological status, a transient effect mirrors the effects seen in children with follow-up after challenge. -The statistical approach to the activity apparently uses an ANOVA with post hocs (Scheffe) at each test time point. While a repeated measure design would have been better for the behavioral measures, the analysis is convincing.	Ramboll would not include Dalal and Poddar (2009) in any quantitative evaluation of neurobehavioral effects in children because the study was conducted in young adult male rats only. The study consistent of a single oral gavage dose, with transient changes in neurotransmitter levels reported with no associated change in brain morphology. The findings are also questionable due to the inconsistency with results from multi-generation studies in animals (Tanaka 2001) and the longer term study conducted by Dalal and Poddar (2010). In addition, the specifications for the test substance, Erythrosine B, purchased from Sigma is not comparable to FDA specification for FD&C 3 under 21CFR74.303. Therefore, critical information about purity of the color additive, as well as the extent of critical impurities and reaction by-products in the color additive for which maximum limits have been set by regulation are unknown.

Table 1. Comparison of Ramboll and OEHHA reviews of Animal Studies Considered in OEHHA (2020) for Quantitative Risk Assessment			
Study citation	**Ramboll Strengths/ Limitations of study**	**OEHHA Strengths/ Limitations of study**	**Ramboll counterpoints to OEHHA arguments**
Dalal and Poddar 2010	Strengths: -Four treatment groups plus control -Test substance purity reported to be 90% -ToxRTool score of 1 (Reliable without restrictions) Limitations: -Only one sex (males) tested. -Exposure via a single oral gavage dose -No developmental outcomes were tested, only neurobehavioral in adult male rats. -Only one sex (males) tested -Test substance source was reported but test substance purity was not specified -ARRIVE score was mid-range (70%)	Strengths: -Well-developed hypothesis concerning mechanism, replication of the main effect on behavior, and three to four doses with graphic illustration of dose response. -The statistical analysis (ANOVA with post hocs) was appropriate. Limitations: -Although the effect is transient, which reduced its toxicological status, a transient effect mirrors the effects seen in children with follow-up after challenge. -The statistical approach to the activity apparently uses an ANOVA with post hocs (Scheffe) at each test time point. While a repeated measure design would have been better for the behavioral measures, the analysis is convincing.	Ramboll would not include Dalal and Poddar (2010) in any quantitative evaluation neurobehavioral effects in children because the study was conducted in young adult male rats only. Further the changes were a decrease in activity in contract to concerns in children regarding hyperactivity. In addition, the specifications for the test substance, Erythrosine B, purchased from Sigma is not comparable to FDA specification for FD&C 3 under 21CFR74.303. Therefore, critical information about purity of the color additive, as well as the extent of critical impurities and reaction by-products in the color additive for which maximum limits have been set by regulation are unknown.
Tartrazine/Yellow No. 5			
Tanaka 2006	Strengths: -Three treated groups plus negative controls. -Two-generation reproductive/developmental study Exposure from 5 weeks of age in F0 to 9 weeks of age in F_1 -Exposure via the diet -Analysis included test for functional and behavioral outcomes, exploratory behavior, and learning and memory. -Test purity > 85% --ToxRTool score: 1 (Reliable without restrictions) Limitations: -Study relied upon the mouse as the animal model instead of the rat, the rat being the recommended species in the OECD/ICH test guidelines. -Limited number of animals per dose groups decreasing study quality and statistical power of the results.	Strengths: Activity assessed with an automated test using a doughnut-shaped arena, unlike the Open Field test of the early studies which used a rectangular arena. Learning & memory were tested with a water maze, unlike the shock-motivated tests used in earlier studies. Tanaka studies are valuable for their dose-response designs and extensive data reporting Limitations: -The group size and power varied from 7-10 mice per group making the power of these studies inadequate given the sensitivity of many measures. -Pups were pooled across litters for statistical analysis of the preweaning data, so that litter-based statistics were not performed. -Groups were not compared in the maze data analysis for most dyes; statistical analysis was within groups across trials only.	Ramboll agrees with the limitations of Tanaka (2006) as noted by OEHHA. Ramboll considered the results of Tanaka 2006 for quantitative risk assessment and concluded that the study did not provide strong or consistent evidence of effects. The changes reported in Tanaka (2006) are limited to single endpoints in a single sex at a single time point in the study; and, in most cases, the results from behavior tests were changes often in a direction of accelerated achievement of coordination indicating better performance in treated versus control animals. Further, while dietary doses of tartrazine were slightly different between the Tanaka (2006) and Tanaka et al. (2008) studies, comparable changes in endpoints were not reported in the F_1 animals across the two studies. Furthermore, the results of the comparison of control groups indicate the model systems used by the Tokyo Metropolitan Laboratory of Public Health (Tanaka 2006) have significant variability over time, hindering the ability to draw any meaningful

Table 1. Comparison of Ramboll and OEHHA reviews of Animal Studies Considered in OEHHA (2020) for Quantitative Risk Assessment			
Study citation	**Ramboll Strengths/ Limitations of study**	**OEHHA Strengths/ Limitations of study**	**Ramboll counterpoints to OEHHA arguments**
	-Doses magnitudes higher than the ADIs and higher than expected in the human population. -Behavioral changes reported in animals were, in some cases, indicative of accelerated achievement of coordination and better performance -Study conclusions were based on changes in a single endpoint at a single time point, but not at other timepoints, despite continued exposure to the test color. -ARRIVE score was below median value (46%).	-Nonparametric statistics were used because preliminary tests failed to support homogeneity of variance. -Rank-based trend tests (Jonkheere) were used to evaluate dose-response in many studies but were often reported without pairwise comparisons of the individual dose groups with controls.	conclusions from the reported results. Finally, the study was limited by low study quality due to the small number of animals per dose group which decreases the statistical power of the results and limits the value of the study results for dose-response.
Tanaka et al. 2008	<u>Strengths:</u> -The study included three treated groups plus negative controls. -Three-generation reproductive/developmental study -Exposure via the diet -Test purity > 85% -Exposure from 5 weeks of age in F0 to 9 weeks of age in F2 -Analysis included test for functional and behavioral outcomes, exploratory behavior, and learning and memory. -ToxRTool score: 1 (Reliable without restrictions) <u>Limitations:</u> -Study relied upon the mouse as the animal model instead of the rat, the rat being the recommended species in the OECD/ICH test guidelines. -Limited number of animals per dose groups decreasing study quality and statistical power of the results. -Doses magnitudes higher than the ADIs and higher than expected in the human population. -Behavioral changes reported in animals were, in some cases, indicative of accelerated achievement of coordination and better performance -Study conclusions were based on changes in a single endpoint at a single time point, but not at other timepoints, despite continued exposure to the test color. -ARRIVE score was mid-range 59%	<u>Strengths</u>: -Activity assessed with an automated test using a doughnut-shaped arena, unlike the Open Field test of the early studies which used a rectangular arena. -Learning & memory were tested with a water maze, unlike the shock-motivated tests used in earlier studies. -Tanaka studies are valuable for their dose-response designs and extensive data reporting <u>Limitations:</u> -The group size and power varied from 7-10 mice per group making the power of these studies inadequate given the sensitivity of many measures. -Pups were pooled across litters for statistical analysis of the preweaning data, so that litter-based statistics were not performed. -Groups were not compared in the maze data analysis for most dyes; statistical analysis was within groups across trials only.	Ramboll agrees with the limitations of Tanaka et al. (2008) as noted by OEHHA and concluded that the study did not provide strong or consistent evidence of effects. The changes reported in Tanaka et al. (2008) are limited to single endpoints in a single sex at a single time point in the study; and, in most cases, the results from behavior tests were changes often in a direction of accelerated achievement of coordination indicating better performance in treated versus control animals. While dietary doses of tartrazine were slightly different between the Tanaka (2006) and Tanaka et al. (2008) studies, comparable changes in endpoints were not reported in the F_1 animals across the two studies. Furthermore, the results of the comparison of control groups indicate the model systems used by the Tokyo Metropolitan Laboratory of Public Health (Tanaka et al. 2008) have significant variability over time, hindering the ability to draw any meaningful conclusions from the reported results. Finally, the study was limited by low study quality due to the small number of animals per dose group which decreases the statistical power of the results and limits the value of the study results for dose-response.
Sobotka et al. 1977		<u>Strengths:</u>	Ramboll agrees with the limitations of Sobotka et al. (1977) as noted by OEHHA. While OEHHA

Table 1. Comparison of Ramboll and OEHHA reviews of Animal Studies Considered in OEHHA (2020) for Quantitative Risk Assessment			
Study citation	**Ramboll Strengths/ Limitations of study**	**OEHHA Strengths/ Limitations of study**	**Ramboll counterpoints to OEHHA arguments**
	Strengths: -Study was conducted in the OECD preferred rodent species (rats) -Exposure from early pregnancy through 3 months post-natal. -Exposure via the diet -Test purity 93% -ToxRTool score of 2 (Reliable with restrictions) Limitations: -Only two concentrations tested. -Lack of litter-based design -Lack of appropriate statistical analysis -Pregnancy outcome data not presented -Study conducted over 40 years ago and the results of behavioral tests conducted are limited by the age of the study. -No treatment-related effects in measures of motor activity and learning ability. -Since the behavioral changes noted were either transient or not consistent, the toxicological relevance of these findings could not be evaluated. -ARRIVE score was below median value (46%). -ToxRTool score: 2 (Reliable with restrictions)	-One of first experiments to look systematically at food dye neurodevelopmental toxicity -Doses were at or below doses known at the time to be toxic. -General developmental toxicity was seen in terms of offspring growth, thymus weights and red blood cells. Limitations: -Power of the statistical tests weakened by the use of individual t-tests without an initial ANOVA. -Statistical analysis not litter-based. -No data on behavioral tests in controls so impossible to judge their validity and statistical power.	(2020) identified the highest dose tested as the LOAEL, this is inconsistent with the authors conclusions that with the exception of a small transient improvement in development of neuromotor clinging ability in female neonates (which was defined as speculative), the exposure to tartrazine exerted little apparent effect on the functional development of the CNS. Based on the data provided, Ramboll concluded that the study did not provide strong or consistent evidence of effects, even following doses higher than ADI. In addition, limitations of the study included the lack of appropriate statistical analysis and lack of pregnancy outcome data. The behavioral changes reported were transient or not consistent. Further, the NOAEL identified by OEHHA (2020) is higher than the USFDA ADI, even with standard safety factors (e.g. 100).
Gao et al. 2011	Strengths: -Both rats and mice were tested -Three treated groups plus negative controls. -Test purity 85% Limitations: -Exposure via oral gavage -Introduce uncertainty into the analysis by including a "wash-out" period before recording behavioral, molecular, and morphometric measurements on different PNDs, impeding the interpretation of the results in the context of meaningful toxicological outcomes. -ARRIVE score was mid-range (61%) -ToxRTool score was 3 (Not reliable)	Strengths: -The study used both male and female subjects -The demonstration of changes relevant to an oxidative stress hypothesis was based on similar experiments with dyes in other tissues and is consistent with their findings. Limitations: -The study did not include male vs. female as a factor in the statistical analysis. -The sample as a whole was balanced for sex but the composition of the individual groups was not stated.	Ramboll agrees with the OEHHA (2020) limitations and would not include Gao et al. (2011) in any quantitative evaluation because specific study quality and reliability criterial were not met including specific data related to doses administered, number of animals exposed and specification on study design. (i.e. ToxRTool score of 3 – not reliable). In addition, the "wash-out" period before recording behavioral, molecular, and morphometric measurements on different PNDs, impeding the interpretation of the results in the context of meaningful toxicological outcomes.

Table 1. Comparison of Ramboll and OEHHA reviews of Animal Studies Considered in OEHHA (2020) for Quantitative Risk Assessment			
Study citation	**Ramboll Strengths/ Limitations of study**	**OEHHA Strengths/ Limitations of study**	**Ramboll counterpoints to OEHHA arguments**
Rafati et al. 2017	Strengths: - Study was conducted in the OECD preferred rodent species (rats) - Behavioral testing including Novel object recognition and eight-arm radial maze test with two phases of adaptation were performed -ToxRTool Score 2 (Reliable with restrictions) Limitations: -Test substance source was reported but not purity -Two treated groups plus a negative control -Exposure did not begin until 7 weeks of age -Exposure via oral gavage -Introduce uncertainty into the analysis by including a "wash-out" period before recording behavioral, molecular, and morphometric measurements on different PNDs, impeding the interpretation of the results in the context of meaningful toxicological outcomes. -ARRIVE score was mid-range (61%)	Strengths: -The behavioral tasks in this more recent study is well-known for sensitivity to neurotoxins. -The gavage administration may result in higher internal doses and less binding to fiber in the intestines than the diet administration route used in most early dye toxicity studies. -The detailed examination of brain histomorphology helps provide biological plausibility for the behavioral effects. Limitations: -The text and statistics presentation suggest that both dose groups were sometimes combined for comparison to controls; therefore, conclusions about the individual dose groups cannot be reached.	Ramboll agrees with the limitations of Rafati et al. (2017) as noted by OEHHA (2020) and concluded that the study did not provide strong or consistent evidence of effects. Although decreases in learning and memory parameters evaluated by a *novel* object recognition test and a radial arm maze test were reported in the study, the advanced age of rats is a confounder in these memory tests as age alone has been shown to affect radial arm maze test performance
Brilliant Blue/Blue No. 1			
Tanaka et al. 2012	Strengths: -The study included three treated groups plus negative controls. -Two-generation reproductive/developmental study -Test purity > 85% -Exposure via the diet -Exposure from 5 weeks of age in F0 to 9 weeks of age in F_1 -Analysis included test for functional and behavioral outcomes, exploratory behavior, and learning and memory. -ToxRTool score: 1 (Reliable without restrictions) Limitations: -Study relied upon the mouse as the animal model instead of the rat, the rat being the recommended species in the OECD/ICH test guidelines. -Limited number of animals per dose groups decreasing study quality and statistical power of the results.	Strengths: -Activity assessed with an automated test using a doughnut-shaped arena, unlike the Open Field test of the early studies which used a rectangular arena. -Learning & memory were tested with a water maze, unlike the shock-motivated tests used in earlier studies. -Tanaka studies are valuable for their dose-response designs and extensive data reporting Limitations: -The group size and power varied from 7-10 mice per group making the power of these studies inadequate given the sensitivity of many measures. -Pups were pooled across litters for statistical analysis of the preweaning data, so that litter-based statistics were not performed.	Ramboll agrees with the limitations of Tanaka et al. (2012) as noted by OEHHA and concluded that the study did not provide strong or consistent evidence of effects. The results of Tanaka (2012) provide only limited evidence of isolated changes that are not sustained over time despite chronic exposure to the test color and are limited to one or two endpoints measured from a large number intended to evaluate neurobehavioral domains. Furthermore, the results of the comparison of control groups indicate the model systems used by the Tokyo Metropolitan Laboratory of Public Health (Tanaka et al. 2012) have significant variability over time, hindering the ability to draw any meaningful conclusions from the reported results. Finally, the study was limited by low study quality due to the small number of animals per dose group which decreases the statistical power of the results and

Table 1. Comparison of Ramboll and OEHHA reviews of Animal Studies Considered in OEHHA (2020) for Quantitative Risk Assessment			
Study citation	**Ramboll Strengths/ Limitations of study**	**OEHHA Strengths/ Limitations of study**	**Ramboll counterpoints to OEHHA arguments**
	-Doses magnitudes higher than the ADIs and higher than expected in the human population. -Behavioral changes reported in animals were, in some cases, indicative of accelerated achievement of coordination and better performance -Study conclusions were based on changes in a single endpoint at a single time point, but not at other timepoints, despite continued exposure to the test color. -ARRIVE score was mid-range (73%)	-Groups were not compared in the maze data analysis for most dyes; statistical analysis was within groups across trials only.	limits the value of the study results for dose-response.
AFC Mixtures			
Doguc et al. 2013	<u>Strengths:</u> -Study performed in rats, the preferred animal model for neurobehavioral/developmental studies according to OECD. ARRIVE score was high (76%) <u>Limitations</u>: -Too few animals per dose group were utilized -Lack of test substance purity information -Some colors included in the mixture are not approved for use in the United States -Individual colors were not assessed separately; therefore, there is no way of determining whether the effects were from one colorant or a combination of two or more colorants. -Although exposure may be to combinations of these colors, the exposures in these studies are much higher than the ADIs.	<u>Strengths:</u> -The state-of-the-art version of the Morris maze with extensive measures helps support the lack of effect on learning and memory. -The finding of behavioral and tissue marker effects of in utero only exposure detected long after discontinuation of treatment speaks to an interference with developmental processes. <u>Limitations:</u> -The use of both parametric and nonparametric statistics and separate vs. pooled male and female behavioral data makes interpretation more difficult, but generally the identification of dye effects was supported.	Ramboll considered the results of Doguc et al. (2013) for the quantitative risk assessment and concluded there was significant uncertainty in the potential application of the study in a quantitative risk assessment because individual colors were not assessed separately; therefore, there is no way of determining whether the effects were from one colorant or a combination of two or more colorants or any other potential confounder or impurities, especially if the purity of test substance was not reported Therefore, this study should not be considered as part of a quantitative risk assessment.

Table 1. Comparison of Ramboll and OEHHA reviews of Animal Studies Considered in OEHHA (2020) for Quantitative Risk Assessment			
Study citation	**Ramboll Strengths/ Limitations of study**	**OEHHA Strengths/ Limitations of study**	**Ramboll counterpoints to OEHHA arguments**
Doguc et al. 2019	Strengths: -Study performed in rats, the preferred animal model for neurobehavioral/developmental studies according to OECD. ARRIVE score was high (86%) Limitations: -Individual colors were not assessed separately; therefore, there is no way of determining whether the effects were from one colorant or a combination of two or more colorants. -Although exposure may be to combinations of these colors, the exposures in these studies are much higher than the ADIs.	Strengths: -The state-of-the-art version of the Morris maze with extensive measures helps support the lack of effect on learning and memory. -The finding of behavioral and tissue marker effects of in utero only exposure detected long after discontinuation of treatment speaks to an interference with developmental processes. Limitations: -The use of both parametric and nonparametric statistics and separate vs. pooled male and female behavioral data makes interpretation more difficult, but generally the identification of dye effects was supported.	Ramboll considered the results of Doguc et al. (2019) for the quantitative risk assessment and concluded there was significant uncertainty in the potential application of the study in a quantitative risk assessment because individual colors were not assessed separately; therefore, there is no way of determining whether the effects were from one colorant or a combination of two or more colorants or any other potential confounder or impurities, especially if the purity of test substance was not reported Therefore, this study should not be considered as part of a quantitative risk assessment.
Doguc et al. 2015	Strengths: -Study performed in rats, the preferred animal model for neurobehavioral/developmental studies according to OECD. ARRIVE score was high (77%) Limitations: -Individual colors were not assessed separately; therefore, there is no way of determining whether the effects were from one colorant or a combination of two or more colorants. -Although exposure may be to combinations of these colors, the exposures in these studies are much higher than the ADIs.	Strengths: -The state-of-the-art version of the Morris maze with extensive measures helps support the lack of effect on learning and memory. -The finding of behavioral and tissue marker effects of in utero only exposure detected long after discontinuation of treatment speaks to an interference with developmental processes. Limitations: -The use of both parametric and nonparametric statistics and separate vs. pooled male and female behavioral data makes interpretation more difficult, but generally the identification of dye effects was supported.	Ramboll considered the results of Doguc et al. (2015) for the quantitative risk assessment and concluded there was significant uncertainty in the potential application of the study in a quantitative risk assessment because individual colors were not assessed separately; therefore, there is no way of determining whether the effects were from one colorant or a combination of two or more colorants or any other potential confounder or impurities, especially if the purity of test substance was not reported Therefore, this study should not be considered as part of a quantitative risk assessment.
Basak et al. 2014	Strengths: -Study performed in rats, the preferred animal model for neurobehavioral/developmental studies according to OECD. Limitations: -Individual colors were not assessed separately; therefore, there is no way of determining whether the effects were	Strengths: -The finding of behavioral and tissue marker effects of in utero only exposure detected long after discontinuation of treatment speaks to an interference with developmental processes. Limitations: -The use of both parametric and nonparametric statistics and separate vs. pooled male and female	Ramboll considered the results of Basak et al. (2014) for the quantitative risk assessment and concluded there was significant uncertainty in the potential application of the study in a quantitative risk assessment because individual colors were not assessed separately; therefore, there is no way of determining whether the effects were from one colorant or a combination of two or more colorants or any other potential confounder or impurities, especially if the purity

Table 1. Comparison of Ramboll and OEHHA reviews of Animal Studies Considered in OEHHA (2020) for Quantitative Risk Assessment			
Study citation	**Ramboll Strengths/ Limitations of study**	**OEHHA Strengths/ Limitations of study**	**Ramboll counterpoints to OEHHA arguments**
	from one colorant or a combination of two or more colorants. -The results of this study indicate maternal exposure to AFCs may affect the larynx in rats; however, these effects are not relevant to neurobehavioral risk assessment. ARRIVE score was mid-range 71%	behavioral data makes interpretation more difficult, but generally the identification of dye effects was supported.	of test substance was not reported Therefore, this study should not be considered as part of a quantitative risk assessment.
Basak et al. 2017	Strengths: -Study performed in rats, the preferred animal model for neurobehavioral/developmental studies according to OECD. -ARRIVE score was high (76%) Limitations: -Individual colors were not assessed separately; therefore, there is no way of determining whether the effects were from one colorant or a combination of two or more colorants. -The results of this study indicate maternal exposure to AFCs may play a role in both neoplastic and nonneoplastic skin diseases; however, these effects are not relevant to neurobehavioral risk assessment.	Strengths: -The finding of behavioral and tissue marker effects of in utero only exposure detected long after discontinuation of treatment speaks to an interference with developmental processes. Limitations: -The use of both parametric and nonparametric statistics and separate vs. pooled male and female behavioral data makes interpretation more difficult, but generally the identification of dye effects was supported.	Ramboll considered the results of Basak et al. (2017) for the quantitative risk assessment and concluded there was significant uncertainty in the potential application of the study in a quantitative risk assessment because individual colors were not assessed separately; therefore, there is no way of determining whether the effects were from one colorant or a combination of two or more colorants or any other potential confounder or impurities, especially if the purity of test substance was not reported Therefore, this study should not be considered as part of a quantitative risk assessment.
Ceyhan et al. 2013	Strengths: -Study performed in rats, the preferred animal model for neurobehavioral studies according to OECD. -ARRIVE score was high (77%) Limitations: -Individual colors were not assessed separately; therefore, there is no way of determining whether the effects were from one colorant or a combination of two or more colorants. -The observed changes in western blot optical densities are confounded by significant background noise and are not corroborated with pathology/immune-histochemistry or impaired behavior on a single model of learning and memory that was tested. -Reproducibility and clinical relevance of fractional decrease of abundantly expressed protein is unclear.	Strengths: -Study used contemporary techniques (Western blots) to examine expression of glutamate and acetylcholine receptor proteins and looked at one specific cortical area (hippocampus). -Behavioral and tissue marker effects of in utero only exposure detected long after discontinuation of treatment speaks to an interference with developmental processes. Limitations: -More research would be needed to define a mechanism pathway from the tissue assays.	Ramboll considered the results of Ceyhan et al. (2013) for the quantitative risk assessment and concluded there was significant uncertainty in the potential application of the study in a quantitative risk assessment because individual colors were not assessed separately; therefore, there is no way of determining whether the effects were from one colorant or a combination of two or more colorants or any other potential confounder or impurities, especially if the purity of test substance was not reported. In addition, the results of the study indicate exposure to AFCs may lead to alterations in expressions of NMDARs and nAChRs in adulthood and there may be differences in males and females; however, the toxicological meaning of these changes is not completely understood. Therefore, this study should not be considered as part of a quantitative risk assessment.

Table 1. Comparison of Ramboll and OEHHA reviews of Animal Studies Considered in OEHHA (2020) for Quantitative Risk Assessment			
Study citation	**Ramboll Strengths/ Limitations of study**	**OEHHA Strengths/ Limitations of study**	**Ramboll counterpoints to OEHHA arguments**
Kantor et al. 1984	Strengths: -Study performed in rats, the preferred animal model for neurobehavioral studies according to OECD. -Four dose levels were administered plus negative control Test substance administered via the diet Limitations: -Only male rats were included in the study - Exposure did not start until 33 days of age -The only neurological endpoint evaluated was locomotor activity -Study was conducted 36 years ago -Individual colors were not assessed separately; therefore, there is no way of determining whether the effects were from one colorant or a combination of two or more colorants.	Strengths: -Shaywitz et al. 1979, Goldenring et al. 1980, Kantor et al. 1984 and Reisen and Rothblat 1986 were the first dye experiments conducted with automated recording of activity. Limitations: -The decreases in 24-hour activity are difficult to compare to the early DNT studies with the same mixture at lower doses and shorter monitoring periods.	In Kantor et al. (1984), because the dyes were administered as a mixture it is impossible to evaluate the results of the study on an individual dye basis or establish any potential dose-response relationship for individual dyes. Significant uncertainty arises because only one sex was tested, only locomotor activity in adult animals was evaluated, and individuals colors were not evalated. Consequently, this study should not be considered as part of a quantitative risk assessment.
Shaywitz et al. 1979	Strengths: -Study performed in rats, the preferred animal model for neurobehavioral studies according to OECD. -Test substance was administered in three concentrations plus a negative control -Exposure began shortly after birth and continued through puberty simulating infant and childhood exposure. Limitations: -Exposure was via oral gavage -Study was conducted 41 years ago, which may suggest assays are not representative of the current state of the science. -Individual colors were not assessed separately; therefore, there is no way of determining whether the effects were from one colorant or a combination of two or more colorants.	Strengths: -Shaywitz et al. 1979, Goldenring et al. 1980, Kantor et al. 1984 and Reisen and Rothblat 1986 were the first dye experiments conducted with automated recording of activity. Limitations: -The decreases in 24-hour activity are difficult to compare to the early DNT studies with the same mixture at lower doses and shorter monitoring periods.	In Shaywitz et al. (1979), because the dyes were administered as a mixture, it is not possible to evaluate the results of the study on an individual dye basis or establish any potential dose-response relationship for individual dyes. Significant uncertainty would arise because exposure was via oral gavage and not diet as expected in the human population and individual colors were not tested. Therefore, this study should not be considered as part of a quantitative risk assessment.

Table 1. Comparison of Ramboll and OEHHA reviews of Animal Studies Considered in OEHHA (2020) for Quantitative Risk Assessment			
Study citation	**Ramboll Strengths/ Limitations of study**	**OEHHA Strengths/ Limitations of study**	**Ramboll counterpoints to OEHHA arguments**
Goldenring et al. 1980	Strengths: -Study performed in rats, the preferred animal model for neurobehavioral studies according to OECD. -Exposure began shortly after birth and continued through puberty simulating infant and childhood exposure. Limitations: -Limited number of animals per dose groups decreasing study quality and statistical power of the results. -Only one concentration of test substance administered -Exposure was via a gastric cannula -Study was conducted 36 years ago -Individual colors were not assessed separately; therefore, there is no way of determining whether the effects were from one colorant or a combination of two or more colorants.	Strengths: -Shaywitz et al. 1979, Goldenring et al. 1980, Kantor et al. 1984 and Reisen and Rothblat 1986 were the first dye experiments conducted with automated recording of activity. Limitations: -The decreases in 24-hour activity are difficult to compare to the early DNT studies with the same mixture at lower doses and shorter monitoring periods.	In Goldenring et al. (1980), because the dyes were administered as a mixture it is impossible to evaluate the results of the study on an individual dye basis or establish any potential dose-response relationship for individual dyes. The exposure route was by gastric cannula and not through the diet as would be expected in the human population, thus limiting the use of the study results further. Significant uncertainty also arises due to the limited number of animals per dose group decreasing the statistical power of the results, only one dose group was tested and individual colors were not evaluated separately. Therefore, this study should not be considered as part of a quantitative risk assessment.
Reisen and Rothblat 1986	Strengths: -Study performed in rats, the preferred animal model for neurobehavioral studies according to OECD. -Test substance was administered in three concentrations plus a negative control -Exposure began shortly after birth and continued through puberty simulating infant and childhood exposure. Limitations: -Exposure was via oral gavage -Study was conducted 34 years ago -Individual colors were not assessed separately; therefore, there is no way of determining whether the effects were from one colorant or a combination of two or more colorants.	Strengths: -Shaywitz et al. 1979, Goldenring et al. 1980, Kantor et al. 1984 and Reisen and Rothblat 1986 were the first dye experiments conducted with automated recording of activity. Limitations: -The decreases in 24-hour activity are difficult to compare to the early DNT studies with the same mixture at lower doses and shorter monitoring periods.	In Reisen and Rothblat (1986), because the dyes were administered as a mixture, it is impossible to evaluate the results of the study on an individual dye basis or establish any potential dose-response relationship for individual dyes. Significant uncertainty would arise because exposure was via oral gavage and not diet as expected in the human population and individual colors were not tested. Therefore, this study should not be considered as part of a quantitative risk assessment.

Table 1. Comparison of Ramboll and OEHHA reviews of Animal Studies Considered in OEHHA (2020) for Quantitative Risk Assessment			
Study citation	**Ramboll Strengths/ Limitations of study**	**OEHHA Strengths/ Limitations of study**	**Ramboll counterpoints to OEHHA arguments**
Erickson et al. 2014	Strengths: -Study performed in rats, the preferred animal model for neurobehavioral studies according to OECD. -Exposure was via drinking water -Animals were evaluated for effects on motor activity and anxiety-like behaviors from adolescence to 13 months of age. Limitations: -Exposure during limited juvenile development period only -Individual colors were not assessed separately; therefore, there is no way of determining whether the effects were from one colorant or a combination of two or more colorants. -Study was part of a multigeneration study of prenatal stress, so that some of the young rats exposed to food dyes were offspring of the fourth generation of stressed dams and some were offspring of unstressed controls.	Strengths: -Though the dye variable was added on to a larger study of developmental stress, statistical analysis did not detect dye-stress interactions that might limit generalization of the findings. -The experimental protocols and statistical analysis were state-of-the-art in this recent study. Limitations: -All the dyes were given at the same dose in a mixture, but the actual doses of each were close to the FDA ADI. -Only males were tested and there were 2/litter in a group size of eight that included offspring from the maternal stress and maternal no-stress line.	In Erickson et al. (2014), because the dyes were administered as a mixture, it is impossible to evaluate the results of the study on an individual dye basis or establish any potential dose-response relationship for individual dyes. Significant uncertainty would arise because exposure was only during a limited time during juvenile development and individual colors were not tested. Therefore, this study should not be considered as part of a quantitative risk assessment..

References

Başak DK, Aylak F, Karadayi N, Gultekin F. 2014. Effects of maternally exposed food coloring additives on laryngeal histology in rats. Journal of Environmental Pathology, Toxicology, and Oncology, 33: 123-130.

Başak K, Basak PY, Doguc DK, Aylak F, Oguztuzun S, Bozer BM, Gultekin F. 2017. Does maternal exposure to artificial food coloring additives increase oxidative stress in the skin of rats? Human and Experimental Toxicology, 36: 1023-1030.

Ceyhan BM, Gultekin F, Doguc DK, Kulac E. 2013. Effects of maternally exposed coloring food additives on receptor expressions related to learning and memory in rats. Food and Chemical Toxicology, 56: 145-148.

Dalal A, Poddar MK. 2009. Short-term erythrosine B-induced inhibition of the brain regional serotonergic activity suppresses motor activity (exploratory behavior) of young adult mammals. Pharmacology, Biochemistry, and Behavior, 92: 574-582.

Dalal A, Poddar MK. 2010. Involvement of high plasma corticosterone status and activation of brain regional serotonin metabolism in long-term erythrosine-induced rearing motor hyperactivity in young adult male rats. Toxicology Mechanisms and Methods, 20: 287-297.

Doguc DK, Ceyhan BM, Ozturk M, Gultekin F. 2013. Effects of maternally exposed colouring food additives on cognitive performance in rats. Toxicology and Industrial Health, 29: 616-623.

Doguc DK, Aylak F, Ilhan I, Kulac E, Gultekin F. 2015. Are there any remarkable effects of prenatal exposure to food colourings on neurobehaviour and learning process in rat offspring? Nutritional Neuroscience, 18: 12-21.

Doguc DK, Deniz F, Ilhan I, Ergonul E, Gultekin F. 2019. Prenatal exposure to artificial food colorings alters nmda receptor subunit concentrations in rat hippocampus. Nutritional Neuroscience, 1-11. Epub ahead of print.

Erickson ZT, Falkenberg EA, Metz GA. 2014. Lifespan psychomotor behaviour profiles of multigenerational prenatal stress and artificial food dye effects in rats. PLoS One, 9: e92132.

Gentry R, Greene T, Chappell G, Borghoff S, Yang C, Rathman J, Ribeiro JV, Hobocienski B, Mostrag A, Rodricks J, Clewell H. (submitted). Integration of evidence to evaluate the potential for neurobehavioral effects following exposure to USFDA- approved food colors. Food and Chemical Toxicology.

Gao Y, Li C, Shen J, Yin H, An X, Jin H. 2011. Effect of food azo dye tartrazine on learning and memory functions in mice and rats, and the possible mechanisms involved. Journal of Food Science, 76: T125-T129.

Goldenring JR, Wool RS, Shaywitz BA, Batter DK, Cohen DJ, Young JG, Teicher MH. 1980. Effects of continuous gastric infusion of food dyes on developing rat pups. Life Sciences, 27: 1897-1904.

Kantor MA, Trout JR, Lachance PA. 1984. Food dyes produce minimal effects on locomotor activity and Vitamin B-6 levels in postweanling rats. The Journal of Nutrition, 114: 1402-1412.

Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. 2010. Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research. PLoS Biology, 8(6): e1000412.

Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. 2012. Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research. Osteoarthritis and Cartilage Research Society International, 20: 256-260.

Klimisch HJ, Andreae M, Tillmann U. 1997. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regulatory Toxicology and Pharmacology, 25: 1-5.

Noorafshan A, Hashemi M, Karbalay-Doust S, Karimi F. 2018. High dose allura red, rather than the adi dose, induces structural and behavioral changes in the medial prefrontal cortex of rats and taurine can protect it. Acta Histochemica, 120: 586-594.

OEHHA. 2019. 2019 Symposium on Synthetic Food Dyes. California Office of Environmental Health Hazard Assessment. Available at: https://oehha.ca.gov/risk-assessment/general-info/2019-symposium-synthetic-food-dyes (Accessed September 30, 2020)

OEHHA. 2020. Health Effects Assessment: Potential Neurobehavioral Effects of Synthetic Food Dyes in Children. California Environmental Protection Agency, Office of Environmental Health Hazard Assessment, Children's Environmental Health Center. Public Review Draft.

Rafati A, Nourzei N, Karbalay-Doust S, Noorafshan A. 2017. Using vitamin e to prevent the impairment in behavioral test, cell loss and dendrite changes in medial prefrontal cortex induced by tartrazine in rats. Acta Histochemica, 119: 172-180.

Ramboll. 2020. Assessment of neurobehavioral studies under consideration for a quantitative risk assessment for USFDA-approved food colors. Prepared for American Beverage Association. Ramboll US Corporation.

Reisen CA, Rothblat LA. 1986. Effect of certified artificial food coloring on learning and activity level in rats. Neurobehavioral Toxicology and Teratology, 8: 317-320 (As cited in OEHHA 2020).

Schneider K, Schwarz M, Burkholder I, Kopp-Schneider A, Edler L, Kinsner-Ovaskainen A, Hartung T, Hoffmann S. 2009. "ToxRTool", a new tool to assess the reliability of toxicological data. Toxicology Letters, 189: 138-144.

SciRAP. 2020. Science in Risk Assessment and Policy. Available at: http://www.scirap.org/ (Accessed September 30, 2020)

Shaywitz BA, Goldenring JR, Wool RS. 1979. Effects of chronic administration of food colorings on activity levels and cognitive performance in developing rat pups treated with 6-hydroxydopamine. Neurobehavioral Toxicology, 1: 41-47 (As cited in OEHHA 2020).

Sobotka TJ, Brodie RE, Spaid SL. 1977. Tartrazine and the developing nervous system of rats. Journal of Toxicology and Environmental Health, 2: 1211-1220.

Tanaka T. 1994. Reproductive and neurobehavioral effects of allura red ac administered to mice in the diet. Toxicology, 92: 169-177.

Tanaka T. 1996. Reproductive and neurobehavioral effects of sunset yellow FCF administered to mice in the diet. Toxicology and Industrial Health, 12(1): 69-79.

Tanaka T. 2001. Reproductive and neurobehavioural toxicity study of erythrosine administered to mice in the diet. Food and Chemical Toxicology, 39(5): 447-454.

Tanaka T. 2006. Reproductive and neurobehavioural toxicity study of tartrazine administered to mice in the diet. Food and Chemical Toxicology, 44 (2): 179-187.

Tanaka T, Takahashi O, Oishi S, Ogata A. 2008. Effects of tartrazine on exploratory behavior in a three-generation toxicity study in mice. Reproductive Toxicology, 26: 156-163.

Tanaka T, Takahashi O, Inomata A, Ogata A, Nakae D. 2012. Reproductive and neurobehavioral effects of brilliant blue FCF in mice. Birth Defects Research. Part B Developmental Reproductive Toxicology, 95: 395-409.

Vorhees C, Butcher R, Brunner R, Wootten V, Sobotka T. 1983a. A developmental toxicity and psychotoxicity evaluation of fd and c red dye #3 (erythrosine) in rats. Archives of Toxicology, 53: 253-264.

Vorhees C, Butcher R, Brunner R, Wootten V, Sobotka T. 1983b. Developmental toxicity and psychotoxicity of fd and c red dye no. 40 (allura red ac) in rats. Toxicology, 28: 207-217.

ToxStrategies

Appendix B

Review of Mechanistic Data in OEHHA's Draft Report "Health Effects Assessment: Potential Neurobehavioral Effects of Synthetic Food Dyes in Children"

- ToxStrategies, Inc ("TS") is a multidisciplinary scientific consulting firm. TS scientists work with businesses, law firms, and government agencies to provide technical services in computational analyses & modeling, environmental science, exposure assessment, product safety, risk assessment, and toxicology in order to address the scientific, technical, and regulatory challenges. TS scientists routinely evaluate the potential health risks associated with exposures to a wide variety of consumer products, food ingredients and additives, pharmaceuticals, medical devices, pesticides, industrial chemicals, and environmental contaminants.
- TS conducted a critical review of the OEHHA Draft Report 'Health Effects Assessment: Potential Neurobehavioral Effects of Synthetic Food Dyes in Children' ("Draft Report") relative to the mechanistic evidence and corresponding sections.[1/] As part of this critical review, investigations noted elsewhere have been considered and incorporated for contextual information.[2/3/] ToxStrategies, Inc. offers the below general and specific comments.

[1/] OEHHA draft report on "Health Effects Assessment: Potential Neurobehavioral Effects of Synthetic Food Dyes in Children" Released August 28, 2020. Relevant sections on mechanistic evidence: 4.1, 4.3, 5.3, 5.4, 5.5, 7.2.2, 7.5, 8.2, 8.3.

[2/] Chappell GA, Britt JK, Borghoff SJ. 2020. Systematic assessment of mechanistic data for FDA-certified food colors and neurodevelopmental processes. Food Chem Toxicol:111310.

[3/] Gentry, R., T. Greene, G. Chappell, S. Borghoff, C. Yang, J. Rathman, J. Vinnie Ribeiro, B. Hobocienski, A. Mostrag, J. Rodricks, H. Clewell. *Submitted*. Integration of Evidence to Evaluate the Potential for Neurobehavioral Effects Following Exposure to USFDA-Approved Food Colors.

General Comments

Study reliability (i.e., quality of methods and reporting) was not considered in the OEHHA assessment. While study reliability was not used as a means to include or exclude studies in the mechanistic investigation reported elsewhere,[3]/ study reliability was evaluated for each study using the SciRAP tool (among other tools specific to quality of reporting) and discussed in the report to provide important contextual considerations related to the available data.

The OEHHA assessment used a broad search strategy to identify studies associated with each color and did not seem to use filters with a clear objective for sorting the animal and/or *in vitro* information relevant to neurobehavioral endpoints. As such, they identified and included studies in their evaluation that others excluded or, in several instances, did not identify in the literature search.[2]/[3]/

For *in vitro* studies, the significance of the extent of protein binding on the evaluation of neuro-relevant mechanistic data, especially *in vitro* evaluations, is unclear (e.g., Red No. 3). As discussed below, non-specific protein binding of Red No. 3 may interfere with *in vitro* measurements, particularly in model systems with high concentration of tissue. Without confirmation of binding of Red No. 3 in the *in vitro* model system, there is less confidence in the data reported from the assay

Brilliant Blue (FD&C Blue No. 1)

- No mechanistic *in vivo* studies were identified by either TS or OEHHA.
- Only three *in vitro* studies from the literature were discussed in the OEHHA assessment (Section 4.1.5, page 143), one of which (Lau et al., 2006) was also included in the assessment presented in Chappell et al. (2020). While both reviews noted the neurite outgrowth inhibition effect of Brilliant Blue in mouse neuroblastoma cells, the OEHHA assessment did not mention the cytotoxicity results reported by Lau et al. (2006), in which approximately 35% cell death was observed at a concentration over 1000x lower than the IC_{50} for neurite outgrowth reduction. A study by Wang et al. (2013) that identified alterations to purinergic signaling was excluded from the Chappell et al. (2020) assessment based on the fact that purinergic receptors (specifically P2X7) are involved in Huntington's disease (which is inherited), spinal cord injury, "and other neurodegenerative diseases involving neuroinflammation" (as specified in Wang et al., 2013), as opposed to neurobehavioral effects or disorders related to children. Another study (Chen et al., 2016) was not identified in the search conducted by Chappell et al. (2020) because studies on protein binding without measures of neuro-relevant mechanistic events were not included in the objective of that analysis.

- For high-throughput screening (HTS) assays, Chappell et al. (2020) applied filters for assay-specific data quality and cytotoxic interference in the determination of assay endpoint activity.[2]/ Blue No. 1 was active in one of the 11 neuro-relevant assays in which it was tested (Chappell et al., 2020). This assay measured thyroid hormone antagonism. The OEHHA assessment, which did not apply filters for data quality issues in the determination of activity, considered Blue No. 1 to be active in assays for a serotonergic receptor, thyroid peroxidase (TPO) inhibition, thyroid hormone alterations, and androgenic effects. Conflicting results in the OEHHA assessment exist between the summary of food dye activities in Table 4.1 (pages 152–154), which shows that Blue No. 1 was not tested in dopaminergic and opioid signaling, compared to the statement, "All dyes were active in assays targeting dopaminergic and opioid receptor subtypes" within Section 4.3.3.4 ("Summary of HTS evaluation," pages 158–160). Further details and comments on the methods used to select HTS assays that OEHHA considered as within scope of neurobehavioral mechanisms, and determination of activity for each color, can be found in the section "HTS Data Review" further below.

In summary, only very limited neuro-relevant mechanistic information is available for Blue No. 1. OEHHA states (Section 5.3.5, page 169), "Both Blue No. 1 and Green No. 3 inhibit purinergic receptors"; however, the relevance of this finding to potential neurobehavioral outcomes is unclear because purinergic receptors are potentially related to hereditary neurodegenerative diseases involving neuroinflammation (e.g., Huntington's disease) as described for Blue No. 1 by Wang et al., 2013), or have been shown to have an antinociceptive (i.e., beneficial) effect (related to exposure to Green No. 3, as described in Yang et al., 2019). The relevance of purinergic signaling to neurobehavioral effects in children is not explained by OEHHA. It is also stated that "Blue No. 1 inhibited neurite outgrowth in cultured neuroblastoma cells." As pointed out by Chappell et al. (2020), the concentration at which neurite outgrowth was inhibited was significantly cytotoxic to the cells, questioning the relevance of this response and its consideration in the evidence base of neuro-relevant mechanistic data. Gentry et al. (*submitted*) did not identify any *in vivo* studies in which Blue No. 1 was evaluated that met all key quality criteria identified in the mechanistic assessment. Considering the totality of the *in vitro* and *in vivo* mechanistic evidence,[2/3]/ there was insufficient evidence to support an association between exposure to Blue No. 1 and mechanistic changes that are potentially related to neurobehavioral outcomes.

Indigo Carmine (FD&C Blue No. 2)

- No *in vivo* neuro-related mechanistic studies were identified by either TS or OEHHA.

- As noted in OEHHA's report, "We did not locate contemporary toxicology studies using gavage, *in vitro* mechanism studies, or investigation of dye protein binding for Blue No. 2."
- Five studies were identified by OEHHA (Section 4.1.6, pages 144–145) based on the use of Blue No. 2 for color-based visualization in clinical diagnostics, colonoscopy, and brain tumor surgery (Erickson and Lauron, 1960; Jo et al., 2013; Lee et al., 2015; Choi et al., 2011; Kawaguchi et al., 2007). These studies administered Blue No. 2 through non-oral routes; not considered in other's assessments because the inclusion criterion for the route of administration was limited to oral exposures only.[3]/ The route of administration used in these studies for clinical diagnostics that OEHHA included in their assessment (i.e., intravenous, subcutaneous, and intramuscular) would not account for the full toxicokinetic profile of oral exposures, ignoring absorption and distribution considerations. Also, relatively high systemic exposure to Blue No. 2 achieved through these non-oral routes (that would be atypical for the oral route) may have been responsible for the observations noted.
- Seven studies identified by OEHHA (Section 4.1.6, page 144) provide toxicology information from oral exposure routes (Hollingsworth, 1982; Kobylewski and Jacobson, 2012; Borzelleca and Hogan, 1985; Borzelleca et al., 1985; Butterworth et al., 1975; Gaunt et al., 1969; Hansen et al., 1966a). These studies were either not identified within searches conducted by others[2]/[3]/ – for example, Hollingsworth, 1982, not peer-revied or publicly available; Kobylewski and Jacobson, 2012, review article cited in OEHHA for treatment effect on mammary tumor incidence; Butterworth et al., 1975, general toxicity, cited but not detailed in OEHHA report – or were identified but excluded from the assessment because they did not provide neuro-related mechanistic data (i.e., Borzelleca and Hogan, 1985; Borzelleca et al., 1985; Gaunt et al., 1969; Hansen et al., 1966a – all short term or chronic toxicity/carcinogenicity studies without mention of neuro-related effects). These studies are all cited in the OEHHA assessment as demonstrating general toxicity (e.g., decreased pup and dam body weights at the end of lactation, Hollingsworth, 1982; growth inhibition in male rats fed 1% Blue No. 2 in the diet for two years, Hansen et al., 1966a) (Section .3.1.6, page 259) or tumorigenicity (Hollingsworth, 1982; Kobylewski and Jacobson, 2012) (Section 4.1.6, page 144). While OEHHA considers an "indication of possible Blue No. 2 neurotoxic effects is the production of brain tumors" (Section 4.1.6, page 144) based on the significant increase in gliomas in rats fed Blue No. 2 in the diet (2% of diet), without relevant discussion or data linking mechanisms of brain tumorigenicity to mechanisms of neurobehavioral changes, the relevance of the tumor data to neurobehavioral effects in children is unclear. The Hollingsworth, 1982 reference is a color additive petition that is not publicly available and, thus, was not reviewed.

- Two *in vitro* studies (Shinoda et al., 1999, inhibition of human aldehyde reductase; and Kuno and Mizutani, 2005, study of a drug-metabolizing cytochrome P450) from the literature were discussed in the OEHHA assessment (Section 4.1.6, page 145) but were not identified by Chappell et al. (2020), because they were outside the scope of the assessment. These studies did not report data considered specific to neuro-relevant pathways. The rationale for including these studies and/or a relationship between inhibition in human aldehyde reductase and neurobehavioral effects is not given in the OEHHA report.
- For ToxCast HTS assays, Chappell et al. (2020) applied filters for assay-specific data quality (i.e., "flags" for data quality issues, as assigned within ToxCast and detailed below in the section: High Throughput Screening Data Review) and cytotoxic interference in the determination of assay endpoint activity.[2/] Blue No. 2 was not active in any of the eight neuro-relevant assays (as mapped within Chappell et al., 2020) in which it was tested. The OEHHA assessment, which did not apply filters for either data quality issues or cytotoxic interference in the determination of activity, considered Blue No. 2 active for interaction with the aryl hydrocarbon receptor, and shows multiple active "hits" across other neuro-relevant categories in summary Table 4.1 (pages 152–154). However, these active assays/categories were not discussed further in the assessment. Conflicting results in the OEHHA assessment exist between the summary of food dye activities in Table 4.1, which shows that Blue No. 2 was not tested in dopaminergic and opioid signaling, while the statement, "All dyes were active in assays targeting dopaminergic and opioid receptor subtypes" was included within Section 4.3.3.4 ("Summary of HTS evaluation" pages 158–160). Further details and comments on the methods used to select HTS assays that OEHHA considered to be neuro-relevant, and determination of activity for each color, can be found in the section "HTS Data Review" further below.

In summary, no neuro-relevant mechanistic data were identified for Blue No. 2. OEHHA (Section 5.3.5, page 169) points to studies that identify cardiac "side effects" that occur under unique exposure conditions, which have led to hypotheses that a relationship exists between these measured changes and possible serotonin-based mechanisms and histamine release into circulation (Erickson and Lauron, 1960; Jo et al., 2013; Lee et al., 2015). The studies that suggest this information were based on non-oral routes of exposure (intravenous, subcutaneous or intramuscular) for using Blue No. 2 for clinical diagnostics. There was no evidence to suggest that the same effects would occur following oral exposure to Blue No. 2. Overall, no *in vivo* Blue No. 2 studies provided mechanistic data relevant to neurobehavioral outcomes,[3/] while the limited HTS assays mapped to neuro-relevant mechanistic data were inactive (Chappell et al., 2020).[2/]

Fast Green (FD&C Green No. 3)

- No *in vivo* neuro-related mechanistic studies were identified by TS.[3]/
- No *in vivo* neuro-related mechanistic data were identified by OEHHA.
    - OEHHA identified one single short-term toxicology study in rats (125 mg/kg/day by gavage) (Ashour and Abdelaziz, 2009) (Section 4.1.5, page 142); however, this study did not provide mechanistic data that mapped to neurobehavioral adverse outcomes and therefore, was not included in other assessments of the evidence.[3]/ Specifically, Ashour and Abdelaziz (2009) only investigated "standard assays conducted on blood samples," (OEHHA report Section 4.1.5, page 142) including a clinical chemistry panel. The OEHHA assessment references the decreased serum glucose, cholesterol and triglycerides, "among other differences" without any hypothesized or known association with neurobehavioral effects. OEHHA also denotes that confidence in the study is limited by study design issues. Overall, it is not clear why this study is included in the OEHHA assessment.
- Two *in vitro* studies from the literature (van Hooft, 2002, study on hippocampal synaptic function; Yang et al., 2019, study on purinergic receptors) were discussed in the OEHHA assessment (Section 4.1.5, page 143), but were not included in Chappell et al. (2020).[2]/ The study on synaptic function used Green No. 3 as a dye to identify proteins in neuronal cell cultures. The study on the inhibition of purinergic receptors evaluated potential neurotherapeutic value of Green No. 3, i.e., the opposite of an adverse effect, and is considered not relevant to an evaluation of adverse neurobehavioral effects.
- For ToxCast HTS assays, Chappell et al. (2020) applied filters for assay-specific data quality and cytotoxic interference in the determination of assay endpoint activity.[2]/ Green No. 3 was identified as active in two of the eleven neuro-relevant assays (as mapped within Chappell et al., 2020) in which it was tested. These two assays evaluated thyroid hormone antagonism. The OEHHA assessment, which did not apply filters for either data quality issues or cytotoxic interference in the determination of activity, considered Green No. 3 active in assays across all of the neuro-relevant categories identified in their assessment in which the color was tested, as presented in Summary Table 4.1 (pages 152–154). Within the summary presented in the OEHHA assessment, activity for thyroperoxidase (TPO) inhibition was specifically highlighted. Conflicting results in the OEHHA HTS assessment exist between the summary of food-dye activities in Table 4.1, which shows that Green No. 3 was not tested in dopaminergic and opioid signaling, compared to the statement, "All dyes were active in assays targeting dopaminergic and opioid receptor subtypes" within Section 4.3.3.4 ("Summary of HTS evaluation," pages 158–160). Further details and comments on the methods used to select HTS assays that OEHHA considered neuro-relevant, and determination of activity for each color, can be found in section "HTS Data Review" further below.

In summary, only limited neuro-relevant mechanistic information was identified by OEHHA for Green No. 3. OEHHA states (Section 5.3.5, page 169), "Both Blue No. 1 and Green No. 3 inhibit purinergic receptors"; however, the relevance of this finding to potential neurobehavioral outcomes is unclear because inhibition of purinergic receptors has potential neurotherapeutic value and would be considered not relevant for evaluating adverse neurobehavioral effects. The rationale for such a link is not explained further in the OEHHA assessment. Overall, TS did not identify any *in vivo* Green No. 3 studies that provided mechanistic data relevant to neurobehavioral outcomes and only two out of the 11 HTS assays mapped to neuro-relevant mechanistic data were identified as active (Chappell et al., 2020).

Erythrosine (FD&C Red No. 3)

- Summary of mechanistic evidence identified by TS and OEHHA:
    - Fourteen *in vivo* studies were identified by TS and/or OEHHA; two studies provided behavioral data only (Tanaka, 2001; Vorhees et al., 1983a).
    - Four studies that provided mechanistic data were reviewed by both TS and OEHHA (Dalal and Poddar, 2009, 2010; Gardner et al., 1987; Jennings et al., 1990).
    - Six studies that provided mechanistic data were reviewed by TS but not by OEHHA (Bernstein et al., 1975; Butterworth et al., 1976a; Butterworth et al., 1976b; Capen and Martin, 1989; Hansen et al., 1973b; Hiasa et al., 1988)
    - Two studies were identified and reviewed by OEHHA but not by TS (Kurebayashi et al., 1988; Capen, 1998). These studies were excluded by TS, according to the inclusion criterion that any mechanistic data should map to neuro-relevant pathways.
- Eight *in vitro* studies included in the OEHHA assessment (Section 4.1.1, pages 134–135) were either not identified in the review by Chappell et al. (2020),[2]/ or were identified but excluded from the assessment because they did not meet the inclusion criteria. These studies evaluated effects of Red No. 3 on synaptic dopamine uptake inhibition, ATPase, protein binding and protein-protein interaction inhibition, and protein aggregation inhibition (Ganesan et al., 2011; Ganesan and Buchwald, 2013; Lafferman and Silbergeld, 1979; Lee et al., 2016; Morris et al., 1982; Shimizu et al., 2013; Silbergeld et al., 1982; Wong and Kwon, 2011).
- Two *in vivo* mechanistic studies reviewed by both TS and OEHHA report neurotransmitter changes in rat brain after single or multi-day exposure to Red No. 3 (Dalal and Poddar, 2009, 2010). These studies focused on repeat measures of serotonin up to nine hours following administration of Red No. 3; single administration (Dalal and Poddar, 2009) or daily administration for 15 or 30 days (Dalal and Poddar, 2010). The time course of peak behavioral

and neurotransmitter changes coincided with peak Red No. 3 levels in circulation after dosing, suggesting the potential for Red No. 3 to alter neurotransmitter levels in the brain (specifically dopamine and serotonin). The measured neurotransmitter levels returned to baseline levels within 7 hours of dosing. Although the treatment effect was only observed at relatively high doses of Red No. 3 (10, 100 or 200 mg/kg bw/d, >100x ADI), TS considered these studies to have higher reliability (evaluated using the SciRAP tool)[3]/ than many of the other studies evaluated in OEHHA's assessment of Red No. 3. The OEHHA assessment also notes the transient nature of the changes - Section 3.3.5, page 102, it is noted that the changes in neurotransmitter levels were transient, reducing its toxicological status.

- Two *in vivo* mechanistic studies that included measures of thyroid effects were reviewed by both TS[3]/ and OEHHA (Gardner et al., 1987; Jennings et al., 1990). Although the OEHHA assessment states (Section 5.3.5, page 168), "…Red No. 3 has been shown to affect thyroid hormones in both rodents and humans", the same two studies provide limited evidence to support the emphasis on these changes as relevant in this assessment.
    - The rodent study (Jennings et al., 1990) was conducted at high dietary concentrations of Red No. 3 (0.5%, 1%, and 4%), with no information provided for test-article purity. The finding in this study did show changes in thyroid hormone measures (i.e., increased T4, T3, free T4 index), however, basal levels of thyroid-stimulating hormone (TSH) where not changed unless thyrotropin-releasing hormone (TRH) was administered. These changes were also only identified as statistically significant at the highest dietary exposure concentration where there was a significant decrease in body weight. It was stated in the OEHHA assessment (Section 4.1.1, page 135), "The mechanism of this effect has been considered inhibition of the enzyme iodothyronine deiodinase (Jennings et al. 1990), although iodotyrosine deiodinase has also recently been implicated (Shimizu et al., 2013)." The data used to consider inhibition of iodothyronine deiodinase was by an indirect measure noting a decrease in T3 production and conversion of T4 to T3 in T4 perfused rat liver with 2 days of ip administration to these rats at a high dose of 50 mg/kg Red No. 3. Shimizu et al. (2013) reported Red No. 3 inhibition of iodotyrosine deiodinase *in vitro*. This study was not discussed in any detail within the OEHHA assessment. In the review of this study,[3]/ study quality issues were identified, including a lack of positive controls, and no validation that the concentration of Red No. 3 was not interfering in the analysis based on non-specific binding. Also, this inhibition of iodotyrosine deiodinase could not be bridged to any adverse outcomes associated with exposure to Red No. 3 based on a series of *in vivo* studies conducted in rats and as reviewed elsewhere

(summarized in Table S8-1 in Supplement 8 in Gentry et al., *submitted*).[3]/

- A human study (Gardner et al., 1987) is reviewed in the OEHHA assessment (Section 4.1.1, page 135), and the use of this study by JECFA as the basis for the ADI for Red No. 3 is also discussed (Section 7.2.2, Page 251). Thorough review of this study is summarized elsewhere (see Table S8-1 in Supplement 8 in Gentry et al., *submitted*).[3]/ Overall, no significant change was observed in T4, T3, or rT3, and T3-charcoal uptake, and the greatest increase in TSH (within normal range of the hormone) was observed only following a challenge with thyroid-releasing hormone (TRH). Neither with or without TRH were there any significant changes in serum levels of thyroid hormones (T3 or T4).
- Six additional *in vivo* mechanistic studies conducted in experimental animals as reviewed by TS in Gentry et al. (*submitted*)[3]/ yet not included in the OEHHA assessment showed an overall equivocal response in various measures of thyroid pathway effects following oral exposure to Red No. 3, with effects most often observed at high dose levels (4% in diet) (Bernstein et al., 1975; Butterworth et al., 1976a, 1976b; Capen and Martin, 1989; Hansen et al., 1973b; Hiasa et al., 1988; detailed in Table S8-1 in Supplement 8 in Gentry et al. (submitted)[3]/).
- OEHHA (Section 4.1.1, page 135) referenced two *in vivo* studies on thyroid effects that they considered mechanistic, both of which TS excluded because they either did not specifically map to neuro-relevant pathways or they did not address one of the seven FD&C synthetic colors.
    - Effects of Red No. 3 on thyroid tumorigenesis in rats (Capen, 1998)
    - Effect of Rose Bengal on serum T3 and T4 levels in mice exposed to (Kurebayashi et al., 1988)
        - The OEHHA assessment cites this study in the following context only: "Red No. 3 [has] been shown to affect thyroid hormones in both rodents and humans (Gardner et al. 1987; Kurebayashi et al. 1988)" (Section 4.1.1, page 135). The Kurebayashi et al. study included exposure of mice to Rose Bengal, Food Red No. 105, without any mention of erythrosine or Red No. 3.
- Six *in vitro* studies from the literature were discussed in both assessments by OEHHA and Chappell et al. (2020). These studies reported effects on neurotransmitter release, uptake, and levels in neuronal cells, as well as non-specific protein binding and the influence of the latter on *in vitro* studies of alterations to neurotransmitters. Overall, both studies concluded that FD&C Red No. 3 is capable of affecting neurotransmitters in the brain, while attributes related to the test system and protein binding should be considered in assay design and interpretation.

- The OEHHA assessment did not include two studies identified and reviewed by Chappell et al. (2020), which investigated the potential effects of Red No. 3 on central catecholamine systems and central nervous system injury and reported a lack of treatment effect (Mailman 1987; Mailman and Lewis, 1987).
- For ToxCast HTS assays, Chappell et al. (2020) applied filters for assay-specific data quality and cytotoxic interference in the determination of assay endpoint activity.[2]/ Red No. 3 was active in four of the 15 neuro-relevant assays (as mapped within Chappell et al., 2020). These four assays were active for loss of gene expression related to the production, transport, or degradation of the neurotransmitters dopamine and serotonin. The OEHHA assessment, which did not apply filters for either data quality issues or cytotoxic interference in the determination of activity and therefore brings into question the applicability of this assessment, considered Red No. 3 active in assays for all neuro-relevant categories in which it was tested (noted in the Summary Table 4.1, pages 152–154). Although the OEHHA assessment specifically highlights activity related to the androgen, estrogen, and thyroid receptors, TPO inhibition, and monoamine oxidase, only the loss of signal for monoamine oxidase by Red No. 3 was reported in Chappell et al. (2020) assessments after data quality considerations were integrated.

In summary, Red No. 3 has a fuller neuro-relevant mechanistic evidence base compared to all other synthetic colors/dyes being evaluated. Nonetheless, much of the available evidence must be evaluated with care.

OEHHA provides reference to many of these studies without conducting due diligence on study quality. OEHHA summarizes the available evidence on Red No. 3 in Section 5.3.5, page 169: "Some studies of Red No. 3 have reported changes in neurotransmitter uptake by brain tissues, inhibition of enzymes (acetylcholinesterase, Na+/K+ ATPase), and photooxidation of enzymes. Together this *in vitro* work suggests that Red No. 3 can have many biological targets relevant to brain function and is consistent with contemporary work on Red No. 3 protein binding. As well, Red No. 3 has been shown to affect thyroid hormones in both rodents and humans. The absorption of Red No. 3 appears to be low based on very limited pharmacokinetic data. However, there is some absorption and metabolism. Deiodinated metabolites have been measured, and the time of peak Red No. 3 circulating levels corresponds to the time of peak Red No. 3 effects on behavior (activity) and impacts on neurotransmitters measured in *in vivo* studies." Despite OEHHA's summary, and although change in neurotransmitter function remains a key event that could – in combination with other key events – potentially lead to neurobehavioral adverse outcomes, many of the studies (especially the *in vitro* studies) did not consider artifacts at the time these studies were conducted that would otherwise potentially erroneously produce the assay results noted. These artifacts include non-specific protein binding characteristics of Red No. 3.

Additionally, as described elsewhere (Chappell et al., 2020), the changes measured *in vitro* with Red No. 3 are most likely due to methodological artifacts related to the amount of tissue used in incubation systems, more fully described by Mailman et al. (1980). Mailman et al. (1980) found that synaptosomal protein concentration present in the incubation medium significantly influenced the inhibitory effect of Red No. 3 on synaptosomal dopamine uptake in rat brain preparations. The authors attributed this effect to non-specific interactions with neural membranes. This would be a concern with all of the *in vitro* studies cited by OEHHA in Section 4.1.1, starting on Page 134, in which the model systems were not validated. Notably, none of the critical studies conducted by Mailman (1987) and Mailman and Lewis (1987), as cited by Chappell et al., 2020, were even considered by OEHHA (no references to these studies were made). The Mailman studies investigated potential factors contributing to such effects upon *in vitro* exposure to Red No. 3. Overall, there is alignment in results from *in vivo* and *in vitro* studies, indicating that Red No. 3 can potentially disrupt neurotransmitter function at high dose levels (>100x ADI) and that there is a lack of activity for Red No. 3 for thyroid pathway perturbations that may be related to neurobehavioral outcomes.[3]/ Considering the totality of the evidence, there is insufficient evidence supporting an association between exposure to Red No. 3 and neurobehavioral-related key events.

Allura Red (FD&C Red No. 40)

- Three *in vivo* studies were identified by TS in Gentry et al. (*submitted*)[3]/ and/or OEHHA:
    - Two studies provided behavioral data only (Tanaka, 1994; Vorhees et al., 1983b), and one study provided mechanistic data and was reviewed by both TS and OEHHA (Noorafshan et al., 2018). *[Note, however, this study was not identified in the section of the OEHHA document in which mechanistic data are described (Section 4.5.2).]*
        - OEHHA (Section 5.3.4, page 166) noted that "The brain assays demonstrated alterations in cell number, volume, and cell shape in the medial frontal cortex in dye-treated animals compared to controls." However, it was not clear what "dye" the authors were referring to since this section discussed Red No. 3, Red No. 40, Yellow No. 5 and Yellow No. 6. For Red No. 40, the only study that provided mechanistic data, as reported in Gentry et al. (*as submitted*, see Table S8-2 in Supplement 8), was Noorafshan et al., 2018. This study observed changes in stereological assessment of the brain at the high dose group of 70 mg/kg administered by oral gavage, in which the test article purity was not reported. OEHHA (Section 7.3.1.3, page 257) described the induction of morphological changes in the medial pre-frontal cortex of the brain following oral administration of Allura Red to rats at 70 mg/kg bw/d (10x the ADI), with no effect at 7 mg/kg bw/d (the ADI) (Noorafshan et al., 2018). The test article

purity was not reported in this study; as such, it is possible that the effect at the high dose could be attributed to impurities in the test article. OEHHA states that, in this study (Noorafshan et al., 2018), "sample size was small, particularly for the histomorphology/stereology (6 animals per dose group)" (OEHHA, Section 7.5.2, page 273). This would suggest a concern for confidence in the study findings; however, OEHHA goes on to state: "However, there is mechanistic support for oxidative damage from Red No. 40 from other studies and the anti-oxidant taurine reportedly reversed the effects of Red No. 40. Additionally, the changes in the medial prefrontal cortex can be directly related to the cognitive performance of the animals, as this part of the rodent brain is involved in spatial memory, decision-making and attention (Noorafshan et al, 2018)." The specific "other studies" that OEHHA was referring to were not referenced. No other studies were identified in others' investigations to support these findings.[2/3]/

- Four *in vitro* studies from the literature were discussed in the OEHHA assessment (Section 4.1.3, page 140), three of which were not identified by Chappell et al. (2020) as they were considered out-of-scope (e.g., evaluation of potential estrogenic effects, inhibition of carbonic anhydrase) (Axon et al., 2012; Esmaeili et al., 2016; Khodarahmi et al., 2015). One was excluded due to the focus of the study being on the evaluation of a mixture of colors (Park et al., 2009).

- For ToxCast HTS assays, Chappell et al. (2020) applied filters for assay-specific data quality and cytotoxic interference in the determination of assay endpoint activity.[2]/ Red No. 40 was not active in any of the 27 neuro-relevant assays tested (as mapped within Chappell et al., 2020). In contrast, the OEHHA assessment, which did not apply filters for either data quality issues or cytotoxic interference, considered Red No. 40 active for dopaminergic, serotonergic, muscarinic, and nicotinic cholinergic receptors, TPO inhibition, and androgenic effects. While the OEHHA assessment shows active "hits" for estrogenic and glucocorticoid effects in summary Table 4.1 (pages 152–154), these activities were not discussed further. Further details and comments on the methods used to select HTS assays that OEHHA considered neuro-relevant, and determination of activity for each color, can be found in the section "HTS Data Review" further below.

In summary, very limited neuro-relevant mechanistic information is available for Red No. 40 to evaluate mechanistic changes as they relate to potential alterations in neuro-relevant pathways that lead to neurobehavioral outcomes.

In the relevant TS assessments,[2/3]/ only one *in vivo* study was identified that provided neuro-relevant mechanistic data for Red No. 40;, however it did not meet acceptable study criteria (i.e., the study design included only a high bolus dose and there was no information

on test substance purity). Also, there were no activity in *in vitro* HTS assays reporting neuro-related mechanistic data for Red No. 40. Overall, there was insufficient evidence to support an association between exposure to Red No. 40 and neurobehavioral-related key events.

Tartrazine (FD&C Yellow No. 5)

- Twenty-two studies were identified by ToxStrategies (TS) and/or OEHHA:
  - Three studies provided behavioral data only (Sobotka et al., 1977; Tanaka, 2006; Tanaka et al., 2008).
  - Five mechanistic studies were reviewed by both TS and OEHHA (Bhatt et al., 2018; Mohamed et al., 2015; Gao et al., 2011; El-Sakhawy et al., 2019; Rafati et al., 2017)
  - Two mechanistic studies were reviewed by TS but not OEHHA (Albasher et al., 2020; Alsalman et al., 2019)
  - Twelve studies were reviewed by OEHHA but not TS (Abd-Elhakim et al., 2018; Abd-Elhakim et al., 2019; Al-Seeni et al., 2018; El-Desoky et al., 2017; Elbanna et al., 2017; Erdemli et al., 2017; Himri et al., 2011; Khayyat et al., 2017; Lahmass et al., 2017; Lahmass et al., 2018; Mehidi et a., 2017; Velioglu et al., 2019)

- Five *in vivo* mechanistic studies reviewed by both TS and OEHHA report changes in measures of oxidative stress in brain, neuronal cell damage, and neurotransmitter levels (e.g., serotonin, dopamine, and GABA). The majority of these studies, however, did not provide critical study information (see Gentry et al. (*submitted*), Table S8-3 in Supplement 8)[3]/.
  - Gao et al. (2011) as referenced in OEHHA Section 4.1.2, page 137–139 was not described clearly by OEHHA but was compared to the findings reported by Mohamed et al. (2015), based on findings in measures of oxidative damage in rat brains following a similar orally administered dose of 500 mg/kg. A study was also conducted in mice up to 700 mg/kg/day but was also not discussed in this section because no mechanistic data were available. In referring to the Mohamed et al. (2015) study, it was stated, "As in the Gao et al. (2011) study, generation of ROS [reactive oxygen species] by Yellow No. 5 metabolites was suggested as the mechanism of the effects." However, Gao et al. (2011) was identified as reporting measures of oxidative stress in the brains of rats, including increased ROS, only at the highest bolus gavage dose levels (250 and 500 mg/kg bw/d, or 25-50x ADI).[3]/ Gao et al. (2011) and Mohamed et al. (2015) reported purity of the test substance; however, the high dose levels administered by bolus oral gavage may be of concern based on

the significance of these finding under a exposure scenario that is of questionable relevance to humans.

- Bhatt et al. (2018), as referenced in OEHHA Section 4.1.2, page 137, measured markers of oxidative stress in the brain of administered rats. As stated by OEHHA, "The investigators attribute the brain effects to generation of reactive oxygen species (ROS) by Yellow No. 5 aromatic amine metabolites." While it is known that Yellow No. 5 is metabolized in the gut to aromatic amine sulphanilic acid, Bhatt et al. do not provide direct evidence of the presence or activity of sulphanilic acid. Although the study was conducted at the EFSA ADI (7.5 mg/kg/day), this was the only dose level evaluated using a test article in which purity was not provided; also, methodological quality (study reliability) was considered to be low according to SciRAP.[3]/ "Without corresponding measures of behavioral activity, the relationship between changes in measures of oxidative stress in the brain and potential adverse neurobehavioral outcomes cannot be established. Additionally, this study did not report purity of the test substance, resulting in a lack of confidence in reported findings" (see Gentry et al. (*submitted*) and Table S8-3 in Supplement 8).[3]/
- El-Sakhawy et al. (2019) describes lesions in brain following oral administration of Yellow No. 5 to rats, which is briefly mentioned by OEHHA on page 138. Although changes in brain lesions were identified at oral gavage doses ranging from 7.5 to 100 mg/kg/day in rats, there was low confidence in the reporting based on a lack of statistical analysis and test-substance purity information as noted by others.[3]/
- Rafati et al. (2017) was not described in the OEHHA document in Section 4 where mechanistic data are described. However, it was mentioned in Section 7.3.1.2, on page 257, where it was indicated that brain histomorphometric endpoints (dendritic spine length and brain effects) were used to identify the LOAEL of 5 mg/kg/day. The study was described and reviewed further in Section 7.5.3, page 275, indicating that adult male rats were administered, via oral gavage, dose levels of 5 and 50 mg/kg/day for 7 weeks. As stated elsewhere,[3]/ "Although these changes were accompanied by corresponding decreases in learning and memory parameters evaluated by a novel object recognition test and a radial arm maze test, the advanced age of rats is a confounder in these memory tests as age alone has been shown to affect radial arm maze test performance (Shukitt-Hale et al., 2004)." Thus, interpretation of findings was complicated by significant study limitations, including lack of reporting of test substance purity information, use of rats with advanced age, and administration via oral gavage versus dietary exposure.

- Two additional *in vivo* studies showing inconsistent evidence of oxidative stress measures in brain tissue were reviewed by TS but were not mentioned in the OEHHA assessment (Alsalman et al., 2019; Albasher et al., 2020). Changes in markers of oxidative stress, as well as cerebral and medullary neuronal damage, were reported by Albasher et al. in pre-weaned mouse pups from dams receiving oral gavage administrations of 2.5 and 5 mg/kg bw/day (within the ADI). In contrast, Alsalman et al. reported no change in markers of oxidative stress in brain tissue of rats administered a daily bolus dose of 700 mg/kg bw/d tartrazine (70x ADI), but increased brain lesions and expression of cell cycle and proliferation genes occurred. Both studies failed to report test-article purity and received very low study reliability scores for both reporting and methodological quality according to the SciRAP quality assessment tool.[3]/
- Eight *in vivo* studies identified by OEHHA (Section 4.1.2, Page 138) but not by TS in Gentry et al. (*submitted*) provided evidence of oxidative stress in tissues other than brain (Abd-Elhakim et al. 2018; Abd-Elhakim et al. 2019; Elbanna et al. 2017; Erdemli et al. 2017; Al-Seeni et al. 2018; Velioglu et al. 2019; Khayyat et al., 2017; El-Desoky et al., 2017). These studies were considered out-of-scope because they did not focus on neuro-related mechanistic events in the brain and would not be relevant to this assessment nor were they identified by keywords/search terms used by others.[3]/
- Three *in vivo* studies reviewed by OEHHA (Section 4.1.2, Page 138) did not provide mechanistic data relevant to neuro-related pathways (Himri et al., 2011; Lahmass et al., 2017; Lahmass et al., 2018), and were, thus, not identified in the literature search conducted by TS. The treatment effects cited by OEHHA for these studies are not neuro-relevant: elevated plasma glucose, cholesterol, and creatinine in rats (Himri et al., 2011; no mention of treatment effects in this study related to brain tissue nor behavior); elevated plasma glucose in rats (Lahmass et al., 2017 and 2018). It is not clear why these studies are included in the OEHHA assessment.
- Five *in vitro* studies from the literature were discussed in the OEHHA assessment (Section 4.1.2, page 137). None were reviewed by Chappell et al. (2020) for tartrazine. Four of these studies were considered out-of-scope and not relevant (e.g., evaluation of potential estrogenic effects, inhibition of carboxyl esterase) (Axon et al., 2012; Meyer et al., 2014; Rock and Patisaul, 2018; Sondergaard et al., 1977). One was excluded due to the focus of the evaluation on mixtures of colors (Park et al., 2009).
- For ToxCast HTS assays, Chappell et al. (2020) applied filters for assay-specific data quality and cytotoxic interference in the determination of assay endpoint activity.[2]/ Yellow No. 5 was active in one of the 21 neuro-relevant assays tested (as mapped within Chappell et al., 2020). The OEHHA assessment, which did not apply filters for either data quality issues or cytotoxic interference in the determination of activity, considered Yellow

> No. 5 active for aryl hydrocarbon receptor downregulation and showed multiple active "hits" across other neuro-relevant categories as summarized in Table 4.1 (pages 152–154). These results are not further discussed. The inclusion of the aryl hydrocarbon receptor assay(s) is related to "AOP 150" according to Table 4.1 (page 153). Assuming this refers to AOPWiki AOP #150, which is "Aryl hydrocarbon receptor activation leading to early life stage mortality, via reduced VEGF," it is not clear how this AOP is related to neurobehavioral outcomes. Further details and comments on the methods used to select HTS assays that OEHHA considered neuro-relevant, and determination of activity for each color, can be found in section "HTS Data Review" further below.

In summary, only limited neuro-relevant mechanistic information is available for Yellow No. 5, most of which is focused on measures of oxidative stress and/or damage to neuronal cells. OEHHA (Section 5.3.5, page 168) states that, "Studies of oxidative stress following Yellow No. 5 administration have attributed the brain effects to generation of reactive oxygen species (ROS) by Yellow No. 5 aromatic amine metabolites. Oxidative stress has been reported in other tissues by a number of investigators following Yellow No. 5 administration." However, several studies evaluating oxidative stress in brain tissues were of low quality, the doses administered were high and/or exposure occurred through bolus administration via oral gavage.[3]/ Further, HTS data demonstrated inactivity for markers of oxidative stress markers in non-neural tissues.

Gentry et al. (*submitted*) identified inconsistent changes in *in vivo* measures of oxidative stress in the brain, neuronal cell damage, and neurotransmitter levels (e.g., serotonin, dopamine, and GABA),[3]/ with no corresponding measures in behavioral activity evaluated in most mechanistic studies. For *in vitro* data, only a single serotonin receptor binding (loss of signal) HTS assay was active among otherwise inactive endpoints measured in *in vitro* neuro-relevant mechanistic assays (Chappell et al., 2020).[2]/ Together, these results do not support biological plausibility between exposure to Yellow No. 5 and neurobehavioral effects. Overall, considering the totality of the evidence, there was insufficient evidence supporting an association between exposure to Yellow No. 5 and neurobehavioral-related key events.

Sunset Yellow (FD&C Yellow No. 6)

- One mechanistic *in vivo* study identified by both TS and OEHHA (Osman et al., 2004) examined cholinesterase (ChE) enzyme activity following dietary exposure to Yellow No. 6 (or its metabolite sulphanilic acid) at only one dietary exposure concentration estimated to be ~4 g/kg bw/d (which is the ADI for the parent compound) (detailed in Table S8-4 in Supplement 8 in Gentry et al. (submitted)[3]/).
- There were statistically significant changes in ChE activity following exposure to both the parent compound and its metabolite. OEHHA noted

(Section 4.1.4, page 142) that the data suggested a higher potency for inhibiting ChE for sulphanilic acid compared to the parent compound. Potency is difficult to evaluate *in vivo*, especially with data collected at only one dose level.

- The OEHHA assessment (Section 4.1.4, page 142) stated, "This finding with sulfanilic acid suggests it is the active agent for Yellow No. 6 effects on cholinergic systems, as well as for effects on behavior (Goldenring et al., 1982) and identif[ies] this neurotransmitter system as a potential mechanistic pathway for Yellow No. 6 neurotoxicity." However, Osman et al. (2004) is of low quality/reliability based on the SciRAP assessment and the lack of information on test-article purity, along with being conducted at only one exposure concentration,[3]/ thereby precluding any dose-response evaluation.
- Although the *in vivo* study described by Osman et al. (2004) was of poor quality, it supports previous *in vitro* findings reported by the same investigators (Osman et al., 2002) in which Yellow No. 6 inhibited human cholinesterase and pseudocholinesterase activity, discussed by OEHHA (Section 4.1.4, page 142). However, OEHHA also stated that "In a second experiment, the IC50 for sulfanilic acid inhibition of cholinesterase and pseudocholinesterase was also demonstrated *in vitro* with a lower potency than Yellow No. 6." It is not clear what study OEHHA was referring to since the Osman et al. (2002) study did not evaluate sulfanilic acid. Although TS did not review or conduct a quality assessment on the *in vitro* study (Osman et al., 2004), it was noted that there was no positive control used in this assay, along with a lack of confirmation of test substance solubility and/or toxicity in the model system evaluated (Osman et al., 2002).
- Four *in vitro* studies from the literature were discussed in the OEHHA assessment (OEHHA Section 4.1.3 and 4.1.4, pages 140–142), none of which were identified by Chappell et al. (2020).[2]/ One of these studies reported potential estrogenic effects (Axon et al., 2012) and was considered out-of-scope by Chappell et al. (2020), and another study was excluded as it was evaluating effects from a mixture of colors (Park et al., 2009). Two others identified by OEHHA (Section 4.1.4, page 142), but not by TS, evaluated cholinergic effects, one of which is discussed above (Osman et al., 2002). The other study reported by Goldenring et al. (1982) was designed to evaluate behavioral changes of rat pups following chronic injection (ip administration) of one dose level of sulfanilic acid, a metabolite of Yellow No. 6, with and without pre-treatment of 6-hydroxydopamine. As such, this study did not meet the inclusion criteria identified for evaluating neuro-relevant mechanistic data following oral administration to Yellow No. 6.
- For ToxCast HTS assays, Chappell et al. (2020) applied filters for assay-specific data quality and cytotoxic interference in the determination of assay endpoint activity.[2]/ Yellow No. 6 was inactive in all of the 23 neuro-relevant assays tested (as mapped within Chappell et al., 2020). The OEHHA

> assessment, which did not apply filters for either data quality issues or cytotoxic interferences in the determination of activity, considered Yellow No. 6 active for cholinergic effects and shows multiple active "hits" across other neuro-relevant categories in summary Table 4.1 (pages 152–154), without further explanation. Further details and comments on the methods used to select HTS assays that OEHHA considered neuro-relevant, and determination of activity for each color, can be found in the section "HTS Data Review" further below.

In summary, only limited neuro-relevant mechanistic information is available for Yellow No. 6. OEHHA (Section 5.3.5, page 168) states, "Yellow No. 6 has been shown to inhibit human cholinesterase and pseudocholinesterase *in vitro* and rat cholinesterase *in vivo*, with a potency lower than some organophosphate pesticides." OEHHA (Section 5.3.5, page 168) goes on to state, "Studies with sulfanilic acid, a metabolite of Yellow No. 6, [suggest that] it is the active agent for Yellow No. 6 effects on cholinergic systems, as well as for effects on behavior[,] and identify this neurotransmitter system as a potential mechanistic pathway for Yellow No. 6 neurotoxicity." Both of these OEHHA conclusions are based on one poor-quality *in vivo* study (Osman et al., 2004), and an *in vitro* study (Osman et al., 2002, discussed above).

TS did not identify any *in vivo* studies in which Yellow No. 6 was evaluated that met all key quality criteria as noted in Gentry et al. (submitted).[3]/ However, considering the *in vivo* studies identified in consideration of quality of issues, together with the analysis of *in vitro* mechanistic data described by Chappell et al. (2020), there was limited neuro-relevant mechanistic data for Yellow No. 6. A single study reported decreased acetylcholinesterase activity for Yellow No. 6 and its metabolite sulphanilic acid in dietary exposed rats, while *in vitro* activity reported by Chappell et al., 2020 was limited to serotonergic signaling. Overall, considering the totality of the evidence, there was insufficient evidence supporting an association between exposure to Yellow No. 6 and neurobehavioral-related key events.

High-Throughput Screening (HTS) Data Review

Overall Conclusions

OEHHA (Section 4.3.3.4, page 158): "ToxCast activity for the food dyes ranged widely making it difficult to make strong correlations between what was observed, and adverse effects or mechanisms that have been reported in the literature. The lack of substantial correlations can be due to several factors. For one, the assays used in ToxCast do not represent the entire spectrum of biological processes that might be relevant to human health, including neurobehavioral effects. Therefore, there are gaps in biological coverage of the available assays."

- We agree that HTS assay data do not provide complete biological coverage across neuro-relevant pathways. In fact, there is work in progress to address this issue that

would use HTS assays as only the initial screening tier (Behl et al., 2019). Neurobehavioral testing would need to be conducted at the organism level, which is currently proposed and being tested in zebrafish.

OEHHA Section 5.4, page 174: "Based on the subset of assays we evaluated here, the ToxCast assay results provide limited support for *in vivo* neurotoxicity observations for the food dyes. It should be noted that the assays explored here are intended to provide initial information about the capacity to associate *in vitro* work with the ability for a food dye to promote a biological response. However, these assays are limited for predicting long term or indirect adverse effects in complex biological systems, in part, due to the complexity of the *in vivo* pathway interactions leading to neurotoxicity (including neurobehavioral effects) and DNT [developmental neurotoxicity] compared to the current limited spectrum and range of the ToxCast assays. Evaluation of these chemicals in future iterations may offer more refined results and validate that these gene markers play a critical role in chemically-induced mechanisms of neurotoxicity."

- While we agree with the notion of more HTS testing, any response noted in HTS assays would ultimately only help to formulate hypotheses that would require further testing to determine whether support for a particular adverse neurobehavioral outcome pathway exists. HTS assays represent a good screening and prioritization tool, but cannot be used to 'validate' any particular marker as critical to chemically-induced mechanisms of neurotoxicity. Additionally, at this time, these individual assays can only be reliably used when information on assay interference such as cytotoxicity and chemical purity and stability is available and incorporated into the assessment. This information would increase confidence in activity calls and would enable integration into a reliable evidence base that provide support (or not) for an association with a particular neurobehavioral outcome or with particular events within a neurobehavioral adverse outcome pathway. Chappell et al. (2020) concluded that "the results of our assessment of available *in vitro* mechanistic data collected from assays that measure signals related to MIEs or KEs involved in neurodevelopmental processes indicate that the seven FDA-approved food colors (when batch certified) have limited or no activity for such signals. While available information on FD&C colors and genes or enzymes that may have a role in mechanisms of neurodevelopmental alterations may be limited, FD&C Red No. 3 was the only color (of the seven assessed) that showed activity associated with neurodevelopmental pathways. Additional follow-up assays, especially with test articles that pass analytical QC criteria, would provide clarity and increased confidence in these findings. Overall, the FD&C colors do not appear to alter signaling pathways related to neurodevelopmental processes on the molecular or cellular level."

Both OEHHA and Chappell et al. (2020) agree on overall limitations associated with the current set of HTS assays in providing a reliably predictive signal for neuro-relevant biological processes generally, and for the colors under review specifically. The overall activity profile for the various colors differs between the TS and OEHHA assessments however for two main reasons:

- Differences in the set of assays included in the assessments, some that do not appear to be directly relevant to the neuro-endpoint in the case of OEHHA.

- Consideration of data quality issues and/or cytotoxic interference, or lack thereof in the case of OEHHA.

Differences in overall findings further explained

Set of HTS assays used in assessments

- OEHHA included 283 HTS assays in their risk assessment, 182 of which were deemed directly neuro-relevant. The rest of the assays (other than the 182) included in the OEHHA assessment were generally related to inflammation, oxidative stress, or were selected based on the activity of known neurotoxicants (pesticides), regardless of assay target category or type. On the other hand, Chappell et al. (2020) identified 99 neuro-relevant HTS assays.

- However, OEHHA limited their identification of neuro-relevant assays to the NovaScreen (NVS), Attagene (ATG), and Tox21 vendors, whereas Chappell et al. (2020) considered assays from all vendors. OEHHA included other vendors in their selection of assays based on activity of known neurotoxicants (pesticides) and assays related to oxidative stress and inflammation according to Iyer et al. (2019).

- OEHHA included assays with relatively broader (i.e., indirect, unclear, and/or non-specific) potential relationships to neurobehavioral adverse outcomes, such as estrogen receptor (ER), androgen receptor (AR) antagonism/agonism, and inflammation in non-neuronal cells, respectively. Chappell et al. (2020) did not include these assays because inflammatory response in other cell systems is too general to draw conclusions regarding the ability of the color tested to induce inflammation in the brain *in vivo*, and a link between androgenic or estrogenic changes and neurobehavioral outcomes has not been identified to be a key event in any of the pathways identified to date for neurobehavioral adverse outcomes (Chappell et al., 2020).

- In the OEHHA assessment (Section 4.3.3.5, page 158), there is an error in the statement, "Even with the limitations of the *in vitro* data, in contrast to a recent study published by Chappell et al. (2020), our approach resulted in significantly more active assay hits (283 compared to 116 assays)." In fact, a total of 99 HTS assays were mapped to potential mechanisms of neurobehavioral outcomes in Chappell et al. (2020). Across these 99 assays, the "coverage" of the seven colors tested culminated in a total of only 116 assay endpoints. Further, the OEHHA assessment lists 283 assays selected for their assessment based on their designation that these are in some way relevant to neurobehavioral outcomes. For active assays, however, according to OEHHA's Appendix C. Table 1, a total of 350 assay endpoints across the seven colors among these 283 assays were considered active (based solely on "hit-call") but did not account for all the data quality flags noted previously.

- The number of active ”hit-calls” are, consequently, much fewer in Chappell et al. (2020) compared to OEHHA’s assessment due to a lower number of overall assays deemed relevant to neurobehavioral outcomes, as well as integration of data quality issues and cytotoxic interferences, which OEHHA did not account for nor integrate into their assessment. Therefore, OEHHA could state the following to ensure a more appropriate characterization of the difference: “Even with the limitations of the *in vitro* data, in contrast to a recent study published by Chappell et al. (2020), our approach resulted in more assay endpoints included in the assessment (283 for OEHHA compared to 99 for Chappell et al. (2020)) and more corresponding active “hit-calls” (350 for OEHHA and 8 for Chappell et al. (2020). These differences could be explained by the fact that (i) we cast the net much wider to include indirect effects that have questionable associations with neurobehavioral outcomes and (ii) we (OEHHA) did not account for data quality issues or assay interference due to cytotoxicity in determining activity calls in contrast to the approach Chappell et al. (2020) took.”

- A number of the Attagene assays that OEHHA used in their list of 283 neuro-relevant assays were not included in Chappell et al. (2020) because they are listed as “not developed or optimized to detect loss of signal” in the CompTox database, and are only optimized for gain of signal modeling (e.g., ATG_THRA1_TRANS_DN, ATG_GPCR_DRD1_TRANS_dn, among others).

Consideration of cytotoxicity and data quality flags for HTS assay activity

- Cytotoxicity

  - OEHHA (where stated) applied no filter for cytotoxic interference (i.e., the AC_{50} value of activity relative to the cytotoxic concentration), nor data quality flags (assigned by ToxCast, related to issues in data analysis and model fitting), nor chemical analytical quality control (QC) (e.g., purity and identity). Determination of activity or inactivity was based solely on the “hit-call” provided in the ToxCast database. While Judson et al. (2016) is cited by OEHHA as stating that the cytotoxic burst should not be used as a filter, cytotoxic interference is a known and well-established factor that should be considered in the interpretation of *in vitro* data as discussed in the very same paper (Judson et al., 2016). It is worth noting that the U.S. Environmental Protection Agency (EPA), the National Center for Computational Toxicology [now the Center for Computational Toxicology and Exposure] (NCCT/CCTE) – the EPA division that works with these assays – is not using HTS assays at this time for either hazard or risk assessment. Consequently, it is imperative that the cytotoxic interference information is considered in the determination of assay activity; and, more specifically, it provides the necessary context to assign assay data as “unreliable” vs. active or inactive.

  - Outside of cytotoxic burst criteria, the viability assays in the Tox21 program are specific to measuring cell death related to specific individual assays. There is no reason to ignore such assay information that would otherwise help with appropriate

analyses and interpretation of the results. Chappell et al. (2020) prioritized such viability assay data first relative to cytotoxic burst information for any assays for which specific viability assay data were available (such data are only available for Tox21 assays).

- Data quality
    - OEHHA did not consider data quality flags in their assessment (Section 4.3.1, page 146). Chappell et al. (2020) noted more than one data quality flag would render the assay not "active".[2]/ Examples of data quality flags include: "Noisy data," "Hit-call potentially confounded by overfitting," "Only highest conc above baseline, active," among others.
- Chemical (sample) quality
    - OEHHA makes no mention of chemical quality information in the section on HTS assays (Section 4.3). While Chappell et al. (2020) did not exclude data based on sub-optimal or absent chemical quality data, such information was discussed in the broader context of overall interpretation of the data.[2]/
- It should be noted that other groups/programs are applying various criteria when deciding when to assign a classification of "active" to responses observed for HTS assay endpoints, and/or to include the assay data in an assessment.
    - Relative to National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM)/Integrated Chemical Environment (ICE) curated high-throughput screening (cHTS) data:[4]/
        - Data quality is considered by omitting any assay endpoints with a "hit-call" of active from curated data when: the assay is a down-direction assay (i.e., inhibition, antagonism, loss-of-signal, etc.), and the best-fit curve was a gain-loss model; or the best-fit curve was a gain-loss model, and only a single mid-range concentration had activity above the activity cutoff threshold, among other criteria.
        - Sample quality is considered by omitting assay endpoints from the curated ToxCast HTS data within ICE in which chemicals with a chemical QC grade of "caution" were used, among other criteria.
    - Interestingly, and somewhat contradictory to the approach taken for synthetic colors, OEHHA itself excluded HTS assays based on data quality flags and/or chemical QC issues (i.e., lacking QC information or major issues with chemical

[4] https://ice.ntp.niehs.nih.gov/DATASETDESCRIPTION?section=cHTS

analytical QC results) in their Evidence on the Carcinogenicity of Acetaminophen (2019).[5]/

Examples of discrepant activity calls when data quality filters are or are not applied

Thyroid peroxidase (TPO) inhibition

OEHHA assessment Table 4.1, page 154: "Four dyes [Blue No. 1, Green No. 3, Red No. 3, Red No. 40] associated with downregulation; targeting the loss of signal of TPO activity. The assay is associated with AOP 42" [Inhibition of Thyroperoxidase and Subsequent Adverse Neurodevelopmental Outcomes in Mammals][6]

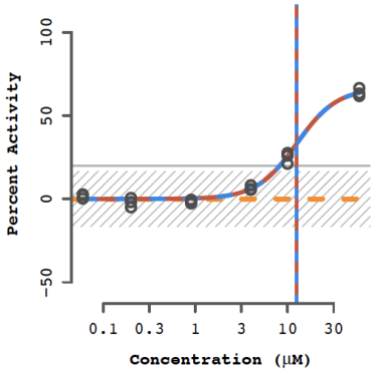
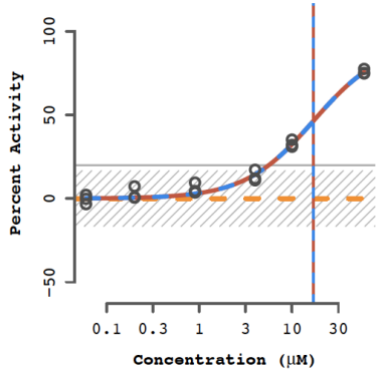
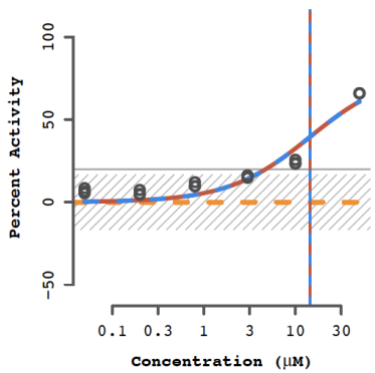
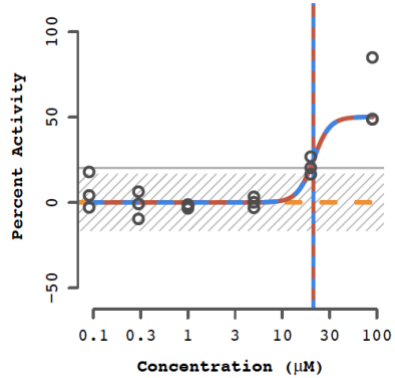
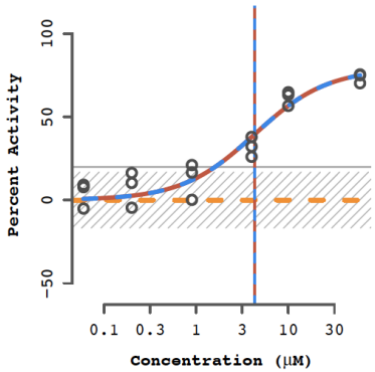
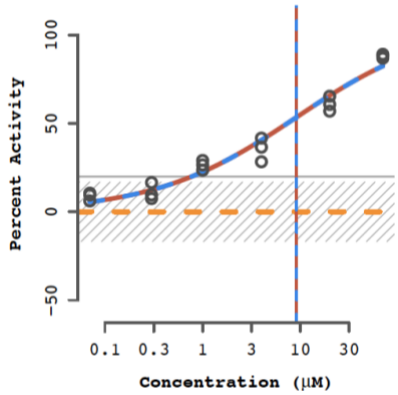
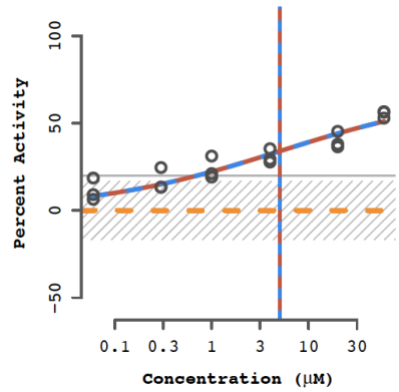
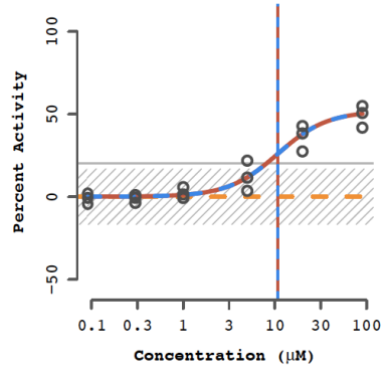
- We do not agree that the TPO assay for these four colors should be considered active. For all four colors, the matched viability assays that are specifically associated with the TPO assay (as reported in a 2016 publication[7] from US EPA scientists within the NCCT/CCTE) are also active, and have AC_{50} values below that of the AC_{50} for the TPO inhibition assays, as shown in Table 1 below. These results indicate that these colors cause a loss-of-signal for TPO enzyme activity only at concentrations at which significant cell death also occurs *in vitro*. Test article purity and identity information was not available for these assays.

[5] https://oehha.ca.gov/media/downloads/crnr/acetaminophenhid092019.pdf

[6] https://aopwiki.org/aops/42

[7] Paul Friedman K, Watt ED, Hornung MW, Hedge JM, Judson RS, Crofton KM, Houck KA, Simmons SO. 2016. Tiered high-throughput screening approach to identify thyroperoxidase inhibitors within the ToxCast Phase I and II chemical libraries. Toxicol Sci 151:160–180.

Table 1. Assay concentration-response plots for the TPO inhibition and related viability assays for four colors (considered active by OEHHA)

Assay target	FD&C Blue No. 1	FD&C Green No. 3	FD&C Red No. 3	FD&C Red No. 40
TPO inhibition (NCCT_TPO _AUR_dn)	 AC_{50} = 12.59µM	 AC_{50} = 17.08µM	 AC_{50} = 14.48µM	 AC_{50} = 21.36µM
Viability assay (NCCT_ HEK293T _CellTiterGLO)	 AC_{50} = 4.32µM	 AC_{50} = 9.13µM	 AC_{50} = 5.04µM	 AC_{50} = 10.77µM

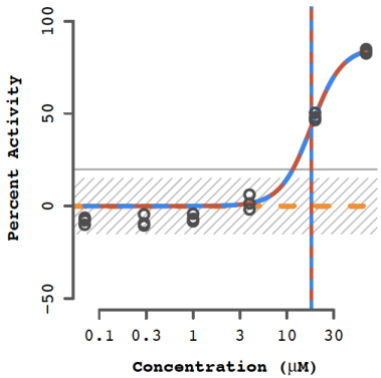
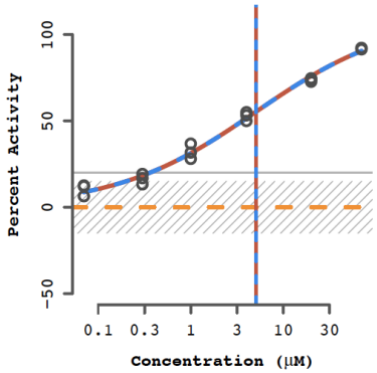
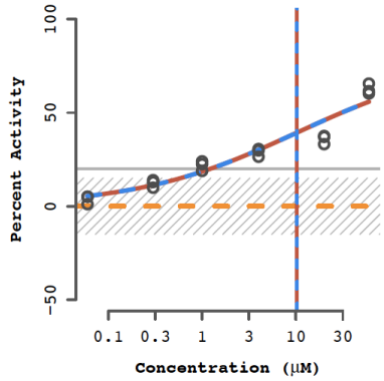
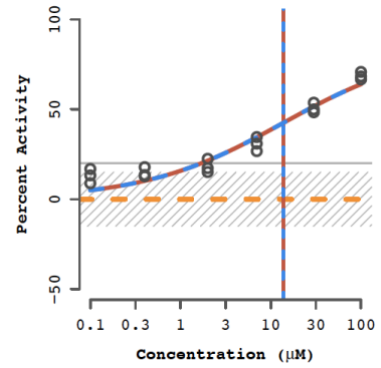
Assay target	FD&C Blue No. 1	FD&C Green No. 3	FD&C Red No. 3	FD&C Red No. 40
Viability assay (NCCT_QuantiLum_inhib_2_dn)	 AC_{50} = 18.23µM	 AC_{50} = 5.08µM	 AC_{50} = 10.26µM	 AC_{50} = 13.79µM

Table 1 Legend: The first row of concentration-response plots represents the assay for TPO inhibition. The bottom two rows of plots represent the concentration response in two cell viability assays specifically related to the main TPO inhibition assay[8]. AC_{50} concentrations are represented by the vertical line on each plot. An AC_{50} value (line) lower in one or both of the viability assays than the AC_{50} value (line) in the TPO inhibition assay indicates that 50% of cells were dead at a lower concentration than the concentration that caused 50% significant loss of TPO enzyme activity. This was the case for the four colors shown in the table. Concentration plots were generated using US EPA's tcpl ("ToxCast pipeline") R package (v2.02) and invitrodb_v3.2 data. Note that the X-axis is log-scaled for all plots, and that the concentration range is not standardized across the colors or assays.

[8] Paul Friedman K, Watt ED, Hornung MW, Hedge JM, Judson RS, Crofton KM, Houck KA, Simmons SO. 2016. Tiered high-throughput screening approach to identify thyroperoxidase inhibitors within the ToxCast Phase I and II chemical libraries. Toxicol Sci 151:160–180.

Thyroid antagonist effects

OEHHA assessment Section 4.3.3.4, page 159: "All the FD&C synthetic food dyes (except for yellow dyes) are active for antagonistic effects with the thyroid hormone receptor."

- We agree that, based upon a single assay, Blue No. 1 and Green No. 3 suggest the potential for thyroid antagonistic activity (as reported in Chappell et al., 2020). However, such a hypothesis would have to be tested further.
- We disagree that the data for Blue No. 2, Red No. 3, and Red No. 40 should be considered active. Data for these colors in this single assay evaluating thyroid receptor (TR) antagonism are either unreliable or inactive because of serious data quality issues (Blue No. 2) or due to significant loss of cell viability in that assay (Red No. 3 and Red No. 40). Test article purity was sufficient and identity was confirmed for Red No. 3 and Red No. 40, while the test article was considered impure by nuclear magnetic resonance (NMR) for Blue No. 2, and no information was available for identity.
    - Blue No. 2: The concentration-response plot clearly shows a lack of activity (Figure 1). The following five flags for poor data quality are reported for this thyroid receptor antagonist assay:
        - AC_{50} less than lowest concentration tested
        - Less than 50% efficacy
        - Gain AC_{50} < lowest conc & loss AC_{50} < mean conc
        - Borderline active
        - Only one conc above baseline, active

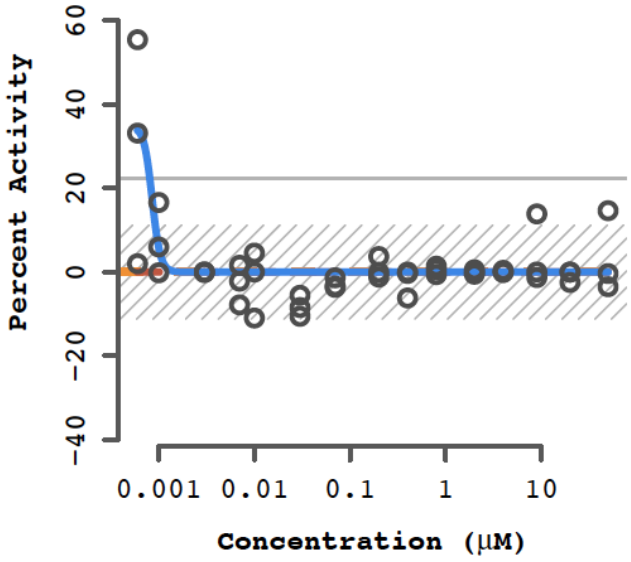
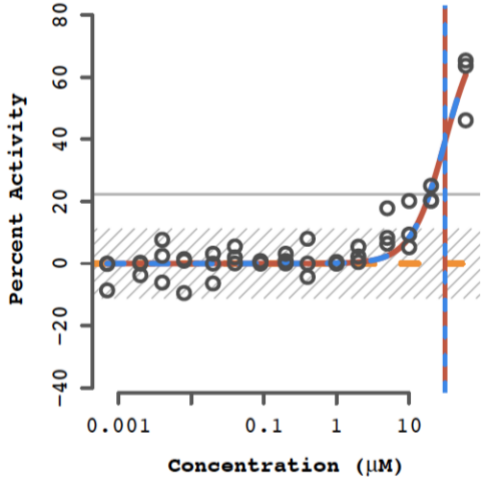
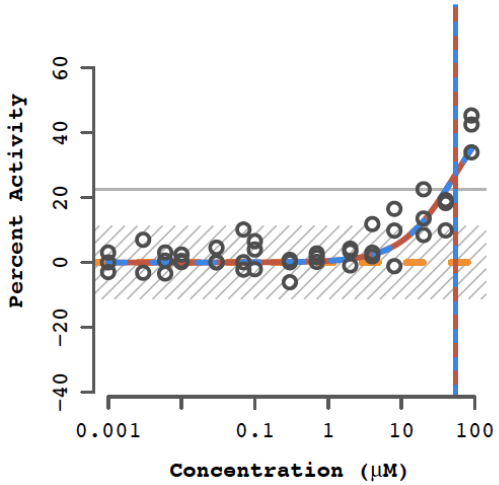


Figure 1. Assay concentration-response plot for TR antagonist assay (TOX21_TR_LUC_GH3_ Antagonist) for Blue No. 2. The plot shows that only a single measurement at lowest concentration tested is active above the baseline cutoff for activity (horizontal grey line). The AC_{50} value for this Blue No. 2 for this assay was estimated to be 0.0000453 µM, which is below the lowest

concentration tested in the assay (0.0006 µM). Concentration plots were generated using the US EPA's tcpl ("ToxCast pipeline") R package (v2.02) and invitrodb_v3.2 data. Note that the X-axis is log-scaled.

Table 2. Assay concentration-response plots for thyroid receptor (TR) antagonist assay and related viability assay for Red No. 3 and Red No. 40

Assay intended target	FD&C Red No. 3	FD&C Red No. 40
TR Antagonism (TOX21_TR_LUC_GH3_ Antagonist)	 AC_{50} = 31.1µM	 AC_{50} = 54.5µM

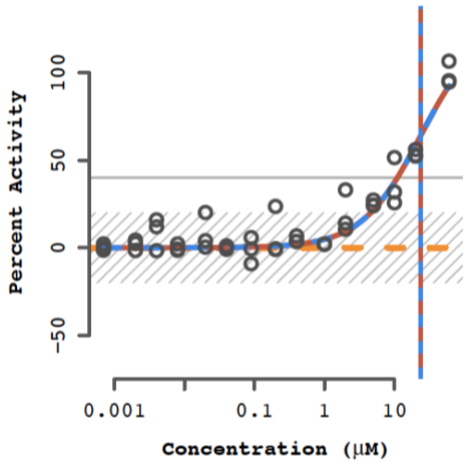
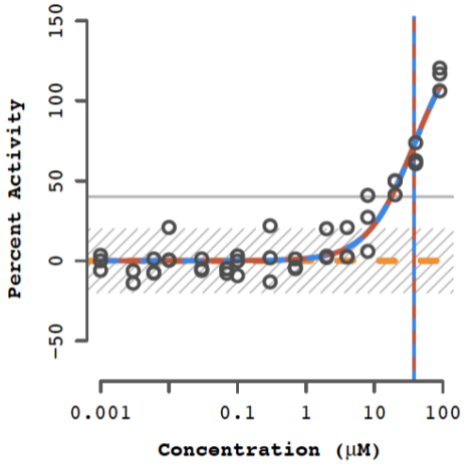
Assay intended target	FD&C Red No. 3	FD&C Red No. 40
TR Antagonism assay viability (TOX21_TR_LUC_GH3_ Antagonist_viability)	 AC_{50} = 18.4µM	 AC_{50} = 37.7µM

Table 2 Legend: The first row of concentration-response plots represents the assay for TR antagonism inhibition. The bottom row of plots represents the concentration response in a cell viability assay specifically related to the main TR antagonism assay. AC_{50} concentrations are represented by the vertical line on each plot. A lower AC_{50} value (line) in the viability assay compared to the AC_{50} value (line) in the TR antagonism assay indicates that 50% of cells were dead at a lower concentration than the concentration that caused 50% significant TR antagonism. This was the case for both colors shown in the table. Concentration plots were generated using the US EPA's tcpl ("ToxCast pipeline") R package (v2.02) and invitrodb_v3.2 data. Note that the X-axis is log-scaled for all plots, and that the concentration range is not standardized across the colors or assays.

Neurotransmitter effects

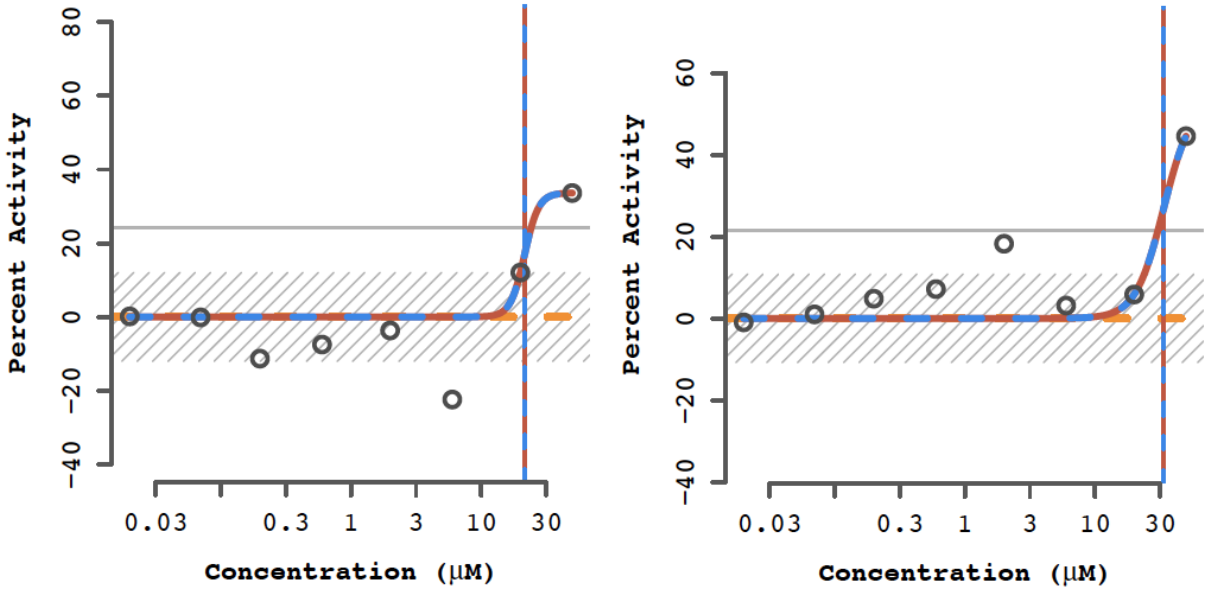
OEHHA assessment states in Section 4.3.3.4, page 159: "All dyes were active in assays targeting dopaminergic […] receptor subtypes."

- In contrast to the quoted statement, according to Table 4.1 on page 152 in OEHHA's report, Blue No. 1, Blue No. 2, and Green No. 3 are listed as "not tested" in assays for "dopaminergic" molecular targets. The assay data in Table 1, Appendix C of the OEHHA assessment corroborate that Blue No. 1, Blue No. 2, and Green No. 3 were not tested in assays for dopamine receptor subtypes.

- Regarding dopaminergic activity in HTS assays, both OEHHA and Chappell et al. (2020) show Red No. 3 to be active. This result, integrated with *in vivo* mechanistic and neurobehavioral findings, is discussed at length within this document above in the section focused on Red No. 3 data. The other colors that were tested in dopamine receptor subtype assays (Red No. 40, Yellow No. 5, and Yellow No. 6) were considered inactive when cytotoxic assay interference and data quality flags were considered. Specifically, the AC_{50} values for all three of these colors in the assay for DRD1 loss-of-signal is well above the respective lower bound estimate for cytotoxicity for each color, according to the ToxCast summary files: Red No. 40, AC_{50} = 25.30μM vs. cytotoxicity lower bound of 6.60μM; Yellow No. 5: AC_{50} = 21.62μM vs. cytotoxicity lower bound of 8.14μM; Yellow No. 6, AC_{50} = 34.40μM vs. cytotoxicity lower bound of 7.94μM. Test article purity and identity information was not available for the dopaminergic signaling HTS assays.

The OEHHA assessment states in Section 4.3.3.4, page 159: "Blue No. 1, Red No. 40, and both yellow dyes were also active for serotonergic receptors."

- Although Yellow No. 5 was active in a *single* assay for serotonergic activity, this loss of signal for the 5-hydroxytryptamine (serotonin) receptor 1A (Htr1a) in rat cortical membranes appears to be species-specific, because other serotonin receptor binding assays for human and guinea pig serotonin receptor genes were inactive for Yellow No. 5. Also, we disagree that the other listed colors were active in serotonergic assays. To illustrate the decision criteria that should be considered in determining activity/inactivity in such assays, see Figure 2 below. Test article purity and identity information was not available for the serotonergic signaling HTS assays.



A. **B.**

Figure 2. **Assay concentration-response plots for Yellow No. 6 in two radioligand binding reporter cell-free assays for human 5-hydroxytryptamine (serotonin) receptor 7, adenylate cyclase-coupled (5HT7) protein (A), and guinea pig 5 hydroxytryptamine (serotonin) receptor 4 (5HT4) protein (B). The data represent loss of signal. The plots show that, for both assays, only a single measurement at the highest concentration tested is active above the baseline cutoff for activity (horizontal grey line). The AC_{50} value for Yellow No. 6 was estimated to be 21.52µM in the 5HT7 assay and 33.57µM in the 5HT4 assay, whereas the lower bound for cytotoxicity for Yellow No. 6, as estimated across a suite of cell viability assays, is 7.94 µM. Concentration plots were generated using the US EPA's tcpl ("ToxCast pipeline") R package (v2.02) and invitrodb_v3.2 data. Note that the X-axis is log-scaled for all plots, and the concentration range is not standardized across the colors or assays.**

Commentary on activity of colors in HTS assays for opioid receptor subcategories

The OEHHA assessment states in Section 4.3.3.4, page 159: "All dyes were active in assays targeting […] opioid receptor subtypes."

- In contrast to the quoted statement, according to Table 4.1 on page 152 in the OEHHA Draft Report, Blue No. 1, Blue No. 2, and Green No. 3 are listed as "not tested" in assays for "opioid" molecular targets. The assay data in Table 1, Appendix C of the OEHHA assessment corroborate that Blue No. 1, Blue No. 2, and Green No. 3 were not tested in assays for opioid receptor subtypes.

- Opioid receptors were not evaluated in Chappell et al. (2020) as these were not identified as having direct relevance to neurobehavioral outcomes according to the resources reviewed within the Chappell et al. (2020) assessment (e.g., literature reviews, AOPWiki, and the Comparative Toxicogenomics Database, among others).[2]/ Nevertheless, in evaluating the available evidence OEHHA considered, we noted the following: only Red No. 3 was active in any of the opioid receptor subtype assays, none of the other colors were active in opioid receptor subtype assays when the same criteria for data quality and cytotoxic assay interference as discussed above (and as used in Chappell et al., 2020) were applied. Red No. 3 was tested in two assays for opioid receptor subtypes and was active for loss of signal for the human opioid receptor, mu 1 (*OPRM1*) in a cell-free radioligand binding assay. Red No. 3 was inactive in an assay for guinea pig opioid receptor, kappa 1 (*Oprk1*). A potential relationship between opioid receptors and mechanisms of neurobehavioral outcomes was not presented by OEHHA. Nonetheless, it would be expected that a potential adverse effect would be related to activation of opioid receptor subtypes, whereas Red No. 3 was only active for *loss* of signal for OPRM1. Test article purity and identity information was not available for the HTS assays related to opioid receptor subtypes.

Commentary on proposed estrogenic and androgenic effects of colors according to HTS data

OEHHA assessment Section 4.3.3.4, page 159: "ToxCast data supports the estrogenic activity observed in literature for Red No. 3 (Dees et al. 1997), but does not support the estrogenic interactions of Yellow No. 5 and No. 6 as reported by Axon, 2012."

OEHHA assessment Section 5.4, page 172: "All the food dyes were active for the androgen assays tested. The dyes, except for Blue No. 2 and the yellow dyes, were active for the receptor-based antagonist assays for the estrogen receptor, potentially indicative of antagonism for this receptor."

- There is no justification or rationale for the inclusion of estrogenic and androgenic assay data provided in the OEHHA assessment, nor discussion regarding the potential effects of ER or AR activity on neurobehavioral outcomes. Estrogenic and androgenic effects were not considered in Chappell et al. (2020), as estrogenic and androgenic effects were not identified as having direct relevance to neurobehavioral outcomes according to the resources reviewed within the Chappell et al. (2020) assessment (e.g., literature reviews, AOPWiki, and the Comparative Toxicogenomics Database, among others) for mode of action and/or underlying mechanisms of neurobehavioral outcomes.

- Nevertheless, upon evaluation of the ER and AR ToxCast Pathway Models, which integrate a battery of HTS assays that represent events across the ER and AR pathways,, inactivity was predicted for agonist, antagonist, and receptor binding for all seven colors, with the exception of a weak AR antagonist activity for Blue No. 1 and Green No. 3. These computational models were developed for use by the

Endocrine Disruptor Screening Program (EDSP), and provide results that can discriminate bioactivity from assay-specific interference and cytotoxicity in the integration of assay data from 18 ER assays[9] or 11 AR assays[10]. Blue No. 1 and Green No. 3 had an area under the curve (AUC) value, which is the model output, of0.107 and 0.209, respectively, for AR antagonism, indicating weak antagonist AR activity (an AUC value of 0.1 corresponds to activity at ∼100 μM by this model).

[9] Browne P, Judson RS, Casey WM, Kleinstreuer NC, Thomas RS. 2015. Screening chemicals for estrogen receptor bioactivity using a computational model. Environ Sci Technol 49:8804–8814.
[10] Kleinstreuer NC, Ceger P, Watt ED, Martin M, Houck K, Browne P, et al. 2017. Development and validation of a computational model for androgen receptor activity. Chem Res Toxicol 30:946–964.