August 23, 2019

VIA ELECTRONIC SUBMISSION

Heidi King
Deputy Administrator
National Highway Traffic Safety Administration
1200 New Jersey Avenue, SE
Washington, DC 20590

Andrew R. Wheeler
Administrator
Environmental Protection Agency
1200 Pennsylvania Ave., NW
Washington, DC 20460

Attn: Docket No. NHTSA-2018-0067
Docket No. NHTSA-2017-0069
Docket No. EPA-HQ-OAR-2018-0283

Re: Supplemental Comments of the Center for Biological Diversity, Environment America, Environmental Defense Fund, Environmental Law & Policy Center, Public Citizen, Inc., Sierra Club, and Union of Concerned Scientists on the Environmental Protection Agency's and National Highway Traffic Safety Administration's Proposed Rule: The Safer Affordable Fuel-Efficient (SAFE) Vehicles Rule for Model Years 2021–2026 Passenger Cars and Light Trucks, 83 Fed. Reg. 42,986 (Aug. 24, 2018)

The Center for Biological Diversity, Environment America, Environmental Defense Fund, Environmental Law & Policy Center, Public Citizen, Inc., Sierra Club, and Union of Concerned Scientists ("Commenters") respectfully submit this supplemental comment on the Environmental Protection Agency's ("EPA") and the National Highway Traffic Safety Administration's ("NHTSA") Proposed Rule: The Safer Affordable Fuel-Efficient ("SAFE") Vehicles Rule for Model Years 2021-2026 Passenger Cars and Light Trucks, 83 Fed. Reg. 42,986 (Aug. 24, 2018) ("Proposed Rule" or "Proposal"). As detailed in the attachment, NHTSA's recent peer review of the "sales" and "scrappage" models used in Proposed Rule underscores that the analysis underlying the agencies' proposal is fundamentally flawed and the proposed standards must be withdrawn. The supplemental comment and attached materials must be considered as part of this ongoing rulemaking as they contain material that is "of central relevance to the rulemaking."[1]

[1] 42 U.S.C. § 7607(d)(4)(B)(i); *see also id.* § 7607(d)(7)(A) (providing that such material forms part of the administrative record for judicial review); Proposed Rule, 83 Fed. Reg. 42,986, 43,471 (Aug. 24, 2018) (citing 49 C.F.R. § 553.23 (committing that "[l]ate filed comments will be considered to the extent practicable")).

Commenters respectfully submit the information contained herein into the SAFE rulemaking docket. Please contact Martha Roberts at mroberts@edf.org or 202-572-3243 if you have any questions regarding this comment.

Respectfully submitted,

Center for Biological Diversity
Environment America
Environmental Defense Fund
Environmental Law & Policy Center
Public Citizen
Sierra Club
Union of Concerned Scientists

Attachment to August 23, 2019, Supplemental Comment of the Center for Biological Diversity, et al., regarding the Peer Review of the Sales and Scrappage Models

Background

In the analysis for the Proposed Rule, EPA and NHTSA (collectively, "the agencies") relied upon two brand new models—the sales response model ("sales model") and the scrappage model—which aimed to quantify the impacts of the proposed fuel economy and greenhouse gas (GHG) emissions standards on new vehicle sales and the scrappage of used vehicles.[2] A wide variety of stakeholders and experts severely criticized these models in public comments on the Proposed Rule. Among other issues, commenters noted fundamental flaws in the design of the models, the lack of connection between the two models even though they were allegedly modeling the interrelationship between sales and scrappage, and the resulting "implausible" model outputs.[3] Commenters also pointed out that the models had never been peer reviewed, which is in violation of legal requirements.[4]

These issues pose fundamental problems for the rulemaking given that the agencies relied on the outputs of these models to justify the rollback of the existing fuel economy and GHG standards. As noted by the Association of Global Automakers in its comments, the scrappage model, in particular, was responsible for "the overwhelming majority of the net benefits associated with each of the Alternatives [in the Proposed Rule]."[5] In addition, the scrappage model was

---

[2] The agencies also relied upon a new "safety model," that estimated fleet fatality rates for use in calculating total fatalities per miles driven. 83 Fed. Reg. at 43,135-145. As far as we know, this new model has never been peer-reviewed, and, as discussed more below, commenters roundly criticized it, as well, including with respect to issues similar to those raised regarding the sales and scrappage models. *See, e.g.*, Comments of the Institute for Policy Integrity at New York University School of Law, NHTSA Docket #2018-0067-12213, Appendix ("Comments of Policy Integrity") at 91-99; Comments of the California Air Resources Board, NHTSA Docket #2018-0067-11873, Analysis in Support of Comments ("Comments of CARB") at 258-82; Comments of R.M. Van Auken, Dynamic Research, Inc., NHTSA Docket #2018-0067-11881; Comments of the Center for Biological Diversity, et al., NHTSA Docket #2018-0067-12000, as corrected NHTSA Docket #2018-0067-12368, Appendix A ("NGO Joint Legal Comments") at 190-99 (noting that the agencies are legally required to have all three models, including the safety model, peer reviewed).

[3] *See, e.g.*, Comments of CARB at 188-282; Comments of Policy Integrity at 13-98; Comments of Dr. James Stock, et al., EPA Docket #HQ-OAR-2018-0283-6220; Comments of Dr. David S. Bunch, EPA Docket #HQ-OAR-2018-0283-5842; Comments of Dr. Kenneth Gillingham, EPA Docket #HQ-OAR-2018-0283-5842; Comments of Dr. Mark Jacobsen and Dr. Arthur van Benthem, NHTSA Docket #2018-0067-7788; Comments of Dr. Antonio Bento, NHTSA Docket #2018-0067-11598; NGO Joint Legal Comments at 171-86; Comments of the Association of Global Automakers, Inc., NHTSA Docket #2018-0067-12032, Appendix A ("Comments of the Association of Global Automakers") at A-22 to A-31.

[4] NGO Joint Legal Comments at 190-99; Comments of CARB at 20, 92 n.34, 216, 226, 250; Comments of Policy Integrity at 86.

[5] Comments of the Association of Global Automakers at A-24. Global Automakers also noted that its technical modeling indicated that the results of the scrappage model were "not consistent with reality," and advised that the scrappage model should "therefore be removed from the Volpe [CAFE] model at this time for purposes of the final rule." *Id.* at A-25. *See also, e.g.*, Comments of Dr. David S. Bunch, EPA Docket #HQ-OAR-2018-0283-5842, at 10 (noting that "the Existing standards cost $14.3B *less* than the Rollback with the scrappage model turned off") (emphasis original); NGO Joint Legal Comments at 184-86.

responsible for virtually all of the projections of avoided fatalities the agencies attributed to the rollback, which was also central to their justification.[6]

Last month, NHTSA updated the CAFE Model[7] Peer Review that it had previously conducted in July 2018, prior to the addition of the sales and scrappage models. This update added the findings of a peer review NHTSA recently conducted of the sales and scrappage models.[8] We highlight below several of the significant substantive findings from the peer review of the sales and scrappage models. We have earlier noted the agencies' wholesale failure to abide by procedural requirements requiring peer review,[9] and we elsewhere further document the agencies' continuing violations of peer review requirements.[10] In particular, we underscore that our comments here do not in any way remediate these errors.

The Peer Review Update Raises Major Substantive Concerns with the New Models

Substantively, the peer review strongly reinforces the commenters' criticisms of the major flaws with the scrappage and sales models, flaws which severely undermine the validity of their projections and reliance on those projections in this rulemaking. As stated in the NHTSA summary of the peer review update, the reviewers' analysis raises "fundamental issues" regarding the models' "specification and implementation."[11] The reviewers make clear that the new model components are only helpful in better understanding the effects of the standards if their development is done correctly. Otherwise, their results would be so untrustworthy as to be worse than not attempting to model the effects of scrappage and sales, which had been the agencies' previous approach given the high degree of uncertainty in predicting the effects of standards on sales and scrappage and the lack of appropriate models.[12] (This approach was

[6] Comments of the Association of Global Automakers at A-24 (noting that when the scrappage model is disabled (or turned "off"), "the non-rebound fatality costs and non-fatal crash costs are higher in Preferred Alternative as compared to the augural standards," demonstrating "the importance of the [scrappage] module on driving the results of the cost/benefit analysis"). In addition, we note that it is highly unlikely that any of the projected fatalities that the agencies attribute to the current standards are statistically significant. *See, e.g.*, Comments of Environmental Defense Fund, NHTSA Docket #2018-0067-12137, at 2-3 (noting that the agencies conceded in the Proposed Rule that their fatality projections due to possible mass reduction of vehicles are not statistically significant and also that, leaving those non-statistically significant projected fatalities aside, the fleet fatality rate is lower under the current standards than under the rollback); Comments of Dr. David S. Bunch, EPA Docket #HQ-OAR-2018-0283-5842, at 11 and 65-67 (finding that the uncertainty in the predicted values from the dynamic scrappage model is so large that the observed differences between the predicted scrap rates under the current standards and the rollback are not statistically meaningful – calling into question the statistical significance of all the costs and benefits that result from these differences).

[7] The CAFE Model (also called the Volpe model) is the overall model that NHTSA has used to evaluate the proposed changes to the fuel economy and GHG standards; the sales, scrappage, and safety models are all sub-models included within the CAFE Model.

[8] NHTSA, *CAFE Model Peer Review* (July 2019 (Revised)), NHTSA Docket #2018-0067-0055 ("Revised CAFE Model Peer Review").

[9] NGO Joint Legal Comments at 190-99.

[10] See forthcoming supplemental comment regarding on-going violations of applicable peer review requirements.

[11] Revised CAFE Model Peer Review at B-3 and B-9 (noting "analogous … issues" with the scrappage model).

[12] NGO Joint Legal Comments at 171-82.

particularly appropriate in light of the fact that other factors, such as the overall economy, are the predominant forces affecting sales and scrappage.[13])

At their core, the peer reviewers' critiques highlighted below can be summarized as follows:

1. The design and development of the models was severely flawed. NHTSA did not use economic theory or best practices in designing the models. Consequently, the models may mirror historical data, but they cannot reliably predict the future effects of fuel economy and GHG standards on sales and scrappage. In addition, the type of model used was inappropriate for the question being investigated, and the models do not include a number of variables understood to influence sales and scrappage. As such, NHTSA's peer reviewers called the models "inherently problematic," "potentially misleading," and "not credible by modern academic standards."[14]
2. The predictions of the models were "implausible." The models predicted that when the price of both new and used vehicles went up, consumers would own more of them—even though in the sales and scrappage modeling, nothing else about the vehicles had changed. In addition, NHTSA assumed that the number of miles that vehicles are driven is entirely determined by the number of vehicles on the road—more vehicles, more driving—without taking into account how much consumers need or want to drive. These "implausible" model projections of more vehicles and more driving under the current standards led to more emissions and more traffic accidents. Avoiding those accidents was the primary benefit cited to justify the rollback.
3. Some of the models' results indicated the models were not in fact doing what NHTSA described them as doing, which raised fundamental questions about their validity.
4. In a number of instances, it was not possible to understand key elements of the design and testing of the models because NHTSA had not provided the necessary information.

Each of these four categories of critiques is summarized in more detail below.

1. The design and development of the models was flawed.

The reviewers explain in detail how the development of the sales and scrappage models failed to follow economic best practice and theory—shortcomings that undermine the models' predictive abilities and necessitate major overhauls of the models if they are to be relied upon. First, the models were not based on economic theory; rather, NHTSA inappropriately focused on taking historical data and developing a model that "fit" that data—i.e., that could predict that same historical data again—but failed to use economic theory and best practices to develop a model that could provide insight into causation, or how variables (like new vehicle prices) are affecting

[13] *Id.*; *see also* EPA, NHTSA, and CARB, *Draft Technical Assessment Report: Midterm Evaluation of Light-Duty Vehicle Greenhouse Gas Emission Standards and Corporate Average Fuel Economy Standards for Model Years 2022-2025*, EPA-420-D-16-900 (July 2016) ("Draft TAR") at 6-1 ("It is difficult, if not impossible, to separate the effects of the standards on vehicle sales and other characteristics from the impacts of macroeconomic or other forces on the auto market.")

[14] Revised CAFE Model Peer Review at B-54, B-57, B-66.

other variables (like sales or scrappage volumes).[15] As such, it is not clear that the models can predict what will happen to sales or scrappage in the future with any reliability, particularly when a new policy is introduced.

Relatedly, a major red-flag that a model may not be properly assessing causation is when key variables (meaning variables that economic theory would suggest are important to the "response" being modeled) are not included, or if when they are included, they lead to illogical results. This is precisely the case with NHTSA's models. As an example, one reviewer noted that, "fuel-economy variables did not improve the explanatory power of the [sales] model," and even though the agencies explored several variants of such variables, "it is concerning that none of these variables improved the time-series model statistically."[16] The model also omitted two other key variables—interest rates for car loans and used car prices—and the omission of the latter was "particularly concerning since the linkage between consumer demand for new versus used vehicles is a key theme of the PRIA [Preliminary Regulatory Impact Analysis] and the preamble's case for less stringent standards."[17]

Further, one reviewer found that "The econometric estimates used [in the sales model] are not credible by modern academic standards,"[18] and also called NHTSA's choice of model for scrappage "inherently problematic."[19] In addition, one peer reviewer described NHTSA's approach in both the sales and the scrappage models as "literally the textbook example for simultaneity bias in nearly every econometrics textbook."[20] Simultaneity bias means that the model does not account for the possibility that the outcome being predicted by the model (for example, new vehicles sales) could itself affect one of the causal variables (for example, new vehicle prices), and therefore causation is running in both directions. In other words, while new car prices would be assumed to impact new car sales, the reverse is also true—for example, if sales are low, car dealers might lower prices. These changes happen simultaneously, and the failure to control for this "reverse causation" element means that the relationship between price increases and sales is likely misestimated.[21] The reviewers doubted that this concern could be

---

[15] *See, e.g.*, Revised CAFE Model Peer Review at B-58 (James Sallee) (in assessing the model, NHTSA refers "only to its time series properties and goodness of fit... The goal of this regression [for the sales model], however, is to identify the causal effect of prices on sales, not to achieve forecast accuracy. … Perfect prediction in sample is not evidence of unbiased (consistent) causal identification.").

[16] *Id.* at B-34 (John Graham) (noting further that, "This omission leaves the sales-response model vulnerable to the allegation that it overstates the adverse effect of fuel-economy regulation on new vehicle demand…").

[17] *Id.* at B-33 (Graham).

[18] *Id.* at B-54 (Sallee).

[19] *Id.* at B-66 (Sallee). *See also, e.g.*, Bento, et al., *Flawed analyses of U.S. auto fuel economy standards*, at 1120, Science (Dec. 7, 2018), NHTSA Docket #2018-0067-12326 (the "newly developed [scrappage] model departs substantially from state-of-the-art vehicle scrappage models," and "does not account for changes in used vehicle prices that result from interactions between new and used car markets"); Comments of Policy Integrity at 72-78 and 87-91.

[20] Revised CAFE Model Peer Review at B-63 (Sallee) (regarding the scrappage model); *see also id.* at B-57 (Sallee) (with respect to the sales model, similarly notes that "This is literally the textbook example of simultaneity bias presented in most econometrics texts.").

[21] *Id.* at B-35 (Graham); *see also id.* at B-19, B-20 and B-22 (Alicia Birky); B-54, B-57 and B-63 (Sallee). *See also, e.g.,* Comments of Dr. James Stock, et al., EPA Docket #HQ-OAR-2018-0283-6220, at 2 (NHTSA's sales model "is inappropriate for estimating the demand for new vehicles due to the fact that supply and demand are simultaneously determined" and further noting that correcting this and other errors with the sales model "reduces the estimated

addressed within NHTSA's current modeling framework.[22] They also questioned whether current modeling techniques and economic research were even capable of addressing the complexities presented in reliably quantifying the effects of the standards on sales and scrappage.[23] These issues and others relating to the design of the models are discussed in Section 1, below.

2. The projections made by the models are implausible.

In addition to criticizing NHTSA's design and development of the models, the reviewers noted the implausibility of the models' outputs. The reviewers found the models' projection that consumers would choose to own more vehicles when vehicles are more expensive "unexpected and unlikely"[24] and "at odds with economic theory."[25] The models' prediction is especially strange because the sales model did not account for the one pathway by which sales could increase under stricter standards, whereby consumer demand for more efficient vehicles outpaces the cost of efficiency technology.

The reviewers further criticized NHTSA's decision to have fleet size determine total vehicle miles traveled (VMT), without any consideration of how much people actually want or need to drive. This approach meant that the projected larger fleet under the current standards automatically led to a proportionally higher amount of VMT, which the reviewers found "unlikely,"[26] and for which NHTSA gave no justification or rationale. It is important to note that this additional VMT under the current standards significantly skewed the cost-benefit analysis presented in the Proposed Rule, because it resulted in greater fuel consumption and GHG and other air pollutant emissions under the current standards than would have otherwise been projected by the models. It was also responsible for virtually all of the fatalities that the agencies attributed to the current standards, because the agencies assumed that traffic fatalities would increase proportionately with VMT. These issues and others relating to the models' outputs are discussed in Section 2, below.

3. In some instances the models' results indicated that they were not doing what NHTSA said they were doing, which raised fundamental questions about their validity.

---

effect on light duty sales of a given price increase by approximately 87%, compared to the estimates in the NPRM/PRIA").

[22] Revised CAFE Model Peer Review at B-19, B-20, and B-22 (Birky); B-35 (Graham); and B-54 (Sallee).

[23] *Id.* at B-22 (Birky) ("No national-level transportation demand models (that this reviewer is aware of) tackle the issue with this level of complexity.") and B-54 (Sallee) (stating he was "not aware of any credible estimates of the causal effect of an aggregate (i.e., market wide) cost (or price) shock in the new vehicle market on new vehicle sales").

[24] Revised CAFE Model Peer Review at B-49 (Howard Gruenspecht).

[25] *Id.* at B-60 (Sallee). *See also, e.g.*, Bento, et al., *Flawed analyses of U.S. auto fuel economy standards*, at 1120, Science (Dec. 7, 2018) ("…the 2018 proposal argues that the rollback in standards will shrink the overall fleet by 6 million vehicles in the year 2029, compared with the current standards. This is inconsistent with basic economic principles."); Comments of Policy Integrity at 62-71.

[26] Revised CAFE Model Peer Review at B-22 (Birky) and B-61 (Sallee) (both calling this outcome "unlikely" and explaining why, as discussed more below); *see also id.* at B-39 (Graham) ("…I might have predicted that any overall change to VMT would be effectively zero, since the regulatory alternatives don't have much obvious impact on the average household's demand for travel"). *See also, e.g.*, Comments by Institute for Policy Integrity at 79-86.

The reviewers identified places where the models produced results that were not consistent with NHTSA's basic descriptions of how the models should operate—for example, where the sales model projected identical new vehicle sales with different new vehicle costs. These issues are discussed in Section 3, below.

4. In a number of instances it was not possible to understand what was happening with the models because NHTSA had not provided the necessary information.

The reviewers also pointed out several instances in which they were unable to understand exactly how NHTSA developed and validated the models, either due to lack of a description of those steps or missing results. These issues are discussed in Section 4, below.

<u>These Flaws Require a New Proposed Rule and New Comment Period</u>

The peer review of the sales and scrappage models is, put bluntly, damning. The reviewers' findings make clear that fundamental changes are needed if the sales and scrappage models are to be included in the analysis and justification for the agencies' final rulemaking. To do otherwise would be patently arbitrary and capricious.

As the results of the sales and scrappage models were central to the agencies' justification for the Proposed Rule, any substantive change—either to revert to the previous modeling approach where sales and scrappage did not change with regulatory alternatives or to redesign the models from the ground up, based on economic theory and best econometric and modeling practice—would undermine the entire analytical basis for the Proposed Rule and inevitably lead to a different proposal. As a result, the agencies must withdraw the fatally flawed Proposed Rule and reinitiate the rulemaking process. Moreover, if the agencies intend to rely upon quantified predictions from sales and scrappage models and attempt to fundamentally redesign those models to address the significant concerns that have been raised, they must: (1) provide a clear explanation of the changes made, as well as the specification and validation of any models the agencies use, including an explanation of how economic theory informed the choices made in model design and development; and (2) subject any new models to a subsequent peer review, as required by law,[27] including public participation. However, given the uncertainty surrounding the effects of the standards on sales and scrappage—uncertainty that both agencies cited just three years ago[28]—and given the peer reviewers' comments regarding the limits of current modeling and economic research in attempting to quantify the complexities of this question, it is possible that the agencies may have to accept that it is currently impossible to develop point estimates of these effects sufficiently reliably and credibly to rely on them in a policy making context.

[27] Memorandum for the Heads of Executive Departments and Agencies, from Russell T. Vought, Acting Director, Office of Management and Budget, regarding Improving Implementation of the Information Quality Act, April 24, 2019, Implementation Update 1.4 ("When influential information that has been peer reviewed changes significantly (e.g., as a result of peer reviewer comments, additional agency analysis, or further consideration), the agency should conduct a second peer review.").
[28] Draft TAR at 6-1 to 6-19.

Finally, regardless of whether the agencies choose to revert back to the previous modeling approach or to instead redesign the sales and scrappage models, they must propose a new rule based on the revised analysis and reopen the public comment period so that interested parties can review and comment upon any new models and the revised proposed rule. In light of the central role of the sales and scrappage models in the justification for the Proposed Rule, these steps are essential before the agencies can issue a final rule.

Specific Significant Issues Noted in the Peer Review Update

1. **The sales and scrappage models are not supported by economic theory, and the serious flaws in their design undermine their ability to do what the agencies rely on them to do—namely, to predict the effects of changes in the standards on sales and scrappage. The following are some of the more significant criticisms of the reviewers on these points, with summary descriptions followed by specific quotations and excerpts from the peer reviewers.**

    a. There are major flaws in the design of the models because economic theory and best econometric practices did not inform the design and development choices, and because the complexity of the dynamics being modeled is not captured or accounted for by the models. As a result, the ability of the models to reliably and credibly predict the effect of the standards on sales and scrappage is highly questionable.

        i. The sales and scrappage models are not based on economic theory. NHTSA's approach inappropriately focused on taking historical data and developing a model that "fit" that data—i.e., that could predict that same historical data again. But the proper goal is to identify causal effects so that the model can predict what will happen to sales and scrappage in the future, not the past, under different possible fuel economy or greenhouse gas standards. There is clear danger of fundamental error in modeling to simply fit the model to the data and then find possible explanations for the relationships that are found (as NHTSA did); instead, the model must be based upon theory and then developed to try to understand the influence the explanatory (causal) variables have on the outcomes.[29] Merely taking data and designing a model to fit the data does not tell you anything about causal relationships or what will happen outside that data set, so the model does not have any true predictive value. For these reasons, it is also critically important to test a model using data outside of the data sample used to develop the model (which NHTSA did not do)—otherwise the model's ability to predict has not been validated.

[29] As an example, historically, the United States' gross domestic product (GDP) has increased over time, and the number of mailboxes in the United States has also increased over time. So one could estimate the relationship between these numbers and might find that GDP growth can be explained by increases in the number of mailboxes, and argue that a "mailbox" model is accurate because it fits the historical data of the number of mailboxes and the GDP used to develop the model. But there is no theory that would support the idea that more mailboxes cause more GDP or that future policies that remove mailboxes would reduce GDP.

- "In assessing the model, the PRIA refers only to its time series properties and goodness of fit... The goal of this regression [for the sales model], however, is to identify the causal effect of prices on sales, not to achieve forecast accuracy. The critical concern should be whether the coefficient is consistently estimated. Perfect prediction in sample is not evidence of unbiased (consistent) causal identification." (Sallee, B-58) Similarly, for the scrappage model, "nearly all specification decisions are described as driven entirely by goodness of fit statistics." (Sallee, B-63).
- "In many cases, the most important impact of new vehicle prices are [sic] in three year lags, and contemporaneous prices are often economically and statistically insignificant. The PRIA argues that the largest effects at three years is logical given the prominence of three year leases. This is plausible, but there are also lots of five year leases, and customers who buy their vehicles tend to put them back on the market later than three years on average. Thus, it begs the question of why all the specifications include only 3[-year] lags. No information is given about what happens at higher lags. In one or two places, it is asserted that 3[-year] lags is 'optimal' but what this means is not explained." (Sallee, B-64)

ii. Moreover, the flaws in the design of NHTSA's models are apparent, as many variables that theory indicates should be important to predicting vehicle sales and/or scrappage (e.g., fuel efficiency, interest rates, and maintenance and repair costs) are omitted from the models; and when NHTSA perform tests that included some of these omitted variables, their inclusion led to illogical results.

- "A paradox of the national time-series modeling is that inclusion of fuel-economy variables did not improve the explanatory power of the model. … While it is encouraging that DOT/EPA analysts explored several variants of fuel-economy variables, it is concerning that none of these variables improved the time-series model statistically. … This omission leaves the sales-response model vulnerable to the allegation that it overstates the adverse effect of fuel-economy regulation on new vehicle demand, since it incorporates only gross technology costs and ignores consumer interest in fuel economy." (Graham, B-34)
- "A weakness in the model is that it does not include important variables concerning consumer access to credit such as average interest rates on car loans. … It also does not address movements in used car prices, a surprising omission given that used cars are a prominent potential substitute for new cars. Both of these variables (interest rates on car loans and used car prices) have been shown to be significant in recent national time-series modeling," and "the sales-response model would be more credible if these two variables were included and if their estimated coefficients exhibited the theoretically expected behavior. The omission of used vehicle prices is particularly concerning since the linkage between consumer demand for new versus used vehicles is a key theme of the PRIA and the preamble's case for less stringent standards." (Graham, B-33)

iii. The design of the models does not sufficiently address the complexity of the questions they are trying to answer. In particular, a number of the variables in the models influence each other, or two variables are both influenced by a third variable that is not in the model—X influences Y but Y also influences X; or Z influences both X and Y, and Z is not in the model. When such relationships between variables are not controlled for, it can appear that a variable has a causal effect on the "response" being modeled (in this case, on vehicle sales or vehicle scrappage) that it does not actually have. In economics terms, this failure to control for relationships between variables is called simultaneity bias. For example, new car prices will affect the volume of new car sales, but the volume of new car sales will also affect car prices.

- NHTSA's approach in both the sales and the scrappage models was "literally the textbook example for simultaneity bias in nearly every econometrics textbook. There is just no reason to believe that this regression delivers unbiased (consistent) estimates of the causal relationship. New vehicle price variation in the time series [used by NHTSA] reflects lots of things-shifts in demand, changes in vehicle attributes, changing composition of vehicles across classes, etc." (Sallee, B-63 and B-57)
- "…I think the national time series model is vulnerable to the criticism that average vehicle transactions prices and average volumes of new vehicle sales are determined simultaneously in the market. … Transactions prices surely do have a negative causative effect on vehicles sales, but this causative relationship could be mis-estimated in the national time series model due to a failure to control for the reverse causation… It is doubtful that [this] concern can be addressed convincingly within the national time-series modeling framework" that NHTSA used. (Graham, B-35)
- "Regarding the simultaneity of average vehicle transaction price and sales: Sales prices of individual models or vehicle body styles and sales volumes are definitely jointly determined, with manufacturers and dealers adjusting price incentives as volumes fluctuate. This does create difficulties that can only be accounted for with complex modeling approaches." (Birky, B-20)

iv. In addition, the models fail to account for the fact that the consumer choices being modeled are made by the same individual and those choices are interrelated. For example, decisions about whether to purchase a new vehicle, which vehicle to purchase, and how much to drive that vehicle are joint decisions made by the same consumer, and they need to be modeled as such to produce reliable results. There are also decisions in these models made by different individuals—for example, the individuals buying new vehicles are likely not the same as those scrapping old vehicles—but the models are not designed to reflect the reality that the consumers making these different decisions are different, and will respond differently to the same variables (like the cost of a vehicle). This also diminishes the reliability of the models' results. Moreover, it does not appear that current

modeling techniques are capable of addressing these complexities, which raises important questions about NHTSA's ability to credibly quantify these effects.

- "VMT demand, the decision to purchase a new vehicle, which vehicle to purchase, and whether to use the purchase to replace an existing vehicle, are joint consumer decisions made at the household level. Therefore, the feedbacks of interest likely are better addressed in a household choice model that includes a market for used vehicles. That said, the decision to *scrap* a vehicle (remove it from the national in-use fleet) and the decision to purchase a new vehicle often are not made by the same household. No national-level transportation demand models (that this reviewer is aware of) tackle the issue with this level of complexity." (Birky, B-22)
- "…it seems that the analysts are tackling issues that are outside the original intent of the model and that current needs may be better met with alternative modeling methodologies and structures. In particular, the PRIA clearly states that the goals of the model changes are to address manufacturer and consumer behavior, yet the model components and system are not choice models." (Birky, B-19)
- "The central parameter (how new vehicle sales will change when new vehicle prices are increased) is difficult to estimate reliably," and the reviewer was "not aware of any credible estimates of the causal effect of an aggregate (i.e., market wide) cost (or price) shock in the new vehicle market on new vehicle sales." (Sallee, B-54)

v. The type of model used for the scrappage model—a "reduced form" model—was not an appropriate model choice for the question being examined. NHTSA attempts to model the effect of a change in new vehicle prices on the rate of scrappage of used vehicles. But in reality, used car prices—not new car prices—would influence scrappage decisions, because the individual making a choice about scrapping a car and potentially replacing that car is looking at used car prices (as well as alternatives to owning a vehicle). NHTSA uses the increase in new car prices as a rough proxy for an increase in used car prices. But the effect of new car price increases on used car prices is a separate causal relationship that needs to be modeled—simply using new car prices, as NHTSA has done here, misses an important step in the process and significantly diminishes the reliability and accuracy of the model.

- "My view is that a reduced form econometric exercise that relies solely on new vehicle prices to determine scrappage is inherently problematic." (Sallee, B-66)
- "What is important is that the [scrappage] model be derived from a consumer choice model that follows economic principles. Such a model would recognize the mechanical relationship between new vehicle sales today and the supply of used vehicles tomorrow, as well as modeling new vehicles as substitutes for used vehicles. Critical also is an explicit representation of the 'outside good'—that is, the choice to not own a car. It is this margin that links

to the overall fleet size, which is the key outcome of the scrap model." (Sallee, B-66)

b. Because of the models' fundamental structural problems, some of which are noted above, NHTSA's attempt to understand and quantify the relationships between the different variables in the models faces additional problems and shortcomings. Specifically, NHTSA's attempt to "estimate" the model—or derive coefficients that indicate the role of a given variable in the overall modeling equation—was flawed. In addition, post hoc adjustments—in layperson terms, artificial constraints placed on the model to prevent it from producing certain results—were needed to make the results fit reasonable bounds.

- "The econometric estimates used are not credible by modern academic standards." (Sallee, B-54)
- Based on certain of the coefficients NHTSA developed and used, "it appears there may be an issue with the specification of the [sales] model. ... This calls into question the results of the other coefficients and indicates possible misspecification." (Birky, B-20)
- "The model produces such implausible survival rates in future cohorts that the modelers chose to add an ad hoc adjustment (the exponential function patch for survival after age 20) to force all vehicles into a (subjectively defined) reasonable scrap pattern. If such an adjustment is required to the regression coefficient outputs, it begs the question of whether the coefficients should be put used [sic] in lieu of a reasonable approximation in the first place." (Sallee, B-64)

2. The results of the sales and scrappage models are "implausible" and "at odds with economic theory." The following are some of the more significant criticisms of the reviewers on this point.

a. Reviewers strongly criticized the lack of connection or integration between the sales and scrappage models. The problems caused by this lack of connectivity are most starkly revealed by the fact that the number of new vehicles added to the fleet (in the sales model) has no effect on the number of vehicles scrapped and removed from the fleet (in the scrappage model). The first[30] and strongest effect of standards for new vehicles on the scrappage of existing vehicles is through any effect of the standards on the sale of new vehicles; if there are fewer new vehicles entering the vehicle fleet (due to lower demand caused by higher new vehicle prices), the total supply of vehicles is effectively reduced, which means there would, all else being equal, be more demand for used vehicles, some of which might then stay in the fleet instead of being scrapped. Because existing vehicles and new vehicles are substitutes for one another, the volume of new vehicle sales is a very important variable to understanding any effect of standards on scrappage rates for existing vehicles. The failure to integrate the sales model with the scrappage model undermines the entire exercise. Specifically, the reviewers recommend switching to a "structural" model that can

[30] A change in fuel economy or emission standards will also eventually affect the used vehicle market (and scrappage) by changing the fuel economy of new vehicles that become existing vehicles.

capture this connection, and, likewise, can capture the impact of used vehicle prices on new vehicle sales.

Reviewers found the lack of connectivity between the sales and scrappage models particularly problematic given the "unexpected" and "unlikely" impact of the scrappage model on total fleet size. In particular, reviewers did not believe there was any sound basis for the model's prediction that an increase in new car prices (and also, consequently, used car prices) would lead to a situation in which consumers wanted to own *more* cars than in the rollback scenario, where new (and used) car prices are lower. This implausible prediction occurred because the models are not integrated, so that total fleet size has no relationship to demand for vehicles and driving. As noted above, under the model, the number of new vehicles added to the fleet has no effect on the number of vehicles scrapped and removed from the fleet.

- "…the model produces outcomes that seem to be at odds with economic theory." (Sallee, B-62)
- The "significant shrinkage in the overall fleet associated with lower new car prices" seems "implausible." (Gruenspecht, B-51)
- "…the consideration of sales responses and scrappage responses as independent processes is problematic, because it fails to use important information regarding the total demand to operate POVs [personally owned vehicles], which has implications for projections of the fleet size." (Gruenspecht, B-48)
- "While some reduction in new LDV [light-duty vehicle] sales under increasingly stringent standards could be reflected in decisions to entirely forego the use of POVs, it [is] difficult to envision that higher new vehicle prices associated with more stringent standards would induce consumers to hold a larger total fleet of POVs. Despite this, the CAFE_ss model run[31] results report a 'many for one' replacement. By 2030, the fleet is nearly 5.9 million vehicles (1.9%) larger in the baseline (B) case with the augural standards than in the preferred alternative (P) case where new care [sic] fuel economy standards and new car prices are lower, a difference that grows to 7.1 million vehicles (2.2%) by 2037. This outcome occurs notwithstanding important costs, including registration fees and required insurance for each vehicle held as discussed in the documents, as well as time-consuming and costly safety and emissions inspection requirements in many jurisdictions that make it extremely awkward and costly to substitute several existing vehicles for a new purchase that is foregone. This unexpected and unlikely result seems directly tied to the use of empirical sales and scrappage models that are independently derived rather than jointly developed within the context of a transportation mode choice model." (Gruenspecht, B-48 to B-49)
- "While substitution between new and existing vehicles in providing services is well established in the literature, the notion that one new LDV would be replaced

[31] The model results that NHTSA used in the rulemaking are from the "CAFE_ss" model runs, which take into account statutory constraints that preclude consideration of some possible real-world compliance strategies. NHTSA also conducts unconstrained model runs ("CAFE" runs), which do not limit real-world compliance strategies. *See* Revised CAFE Model Peer Review at B-43 to B-44; *see also, e.g.*, 83 Fed. Reg. at 43160-61.

with multiple existing ones, as suggested by comparison of the B and P case fleet sizes, seems implausible…" (Gruenspecht, B-52)

- Not integrating the new and used car markets (and instead modeling new vehicle sales and existing vehicle scrappage separately), "is potentially problematic because any errors in the two analyses could compound, rather than counteract each other, yielding net impacts on the size of the fleet that are at odds with economic theory. This appears to have happened in the PRIA, where less expensive new vehicles are projected to shrink the car market, implying that consumers, faced with cheaper cars, choose to substitute away from cars towards other forms of transportation." (Sallee, B-60)
- "The impact of the change in vehicle stock (both total number and average age) on total VMT should be vetted against expected trends in VMT demand." (Birky, B-26)
- "…it is important that the new and used markets interact within the CAFE model. If the CAFE model wishes to fully incorporate fleet size effects into the cost-benefit analysis, it needs to do so in a way that is internally consistent with economic theory. This will require some theoretical equilibrium bridge between the markets, rather than two parallel reduced form econometric exercises." (Sallee, B-61)
- "The PRIA documents final model results that imply that more expensive new vehicles lead to a larger total vehicle fleet. This is problematic." (Sallee, B-65)
- "What seems most critical is that the new vehicle sales and scrap results be forced into a relationship in a theoretical model…" (Sallee, B-66)

b. Reviewers objected to NHTSA's use of VMT schedules that assumed that total VMT would increase or decrease proportionally with fleet size, which was implausible; reviewers stated that total VMT should track demand for travel, not fleet size.

i. One reviewer noted that he would have expected any change in VMT due to a change in fuel economy or emission standards to be effectively zero, as the regulatory alternatives did not have an obvious impact on an average household's demand for travel.

- "…I might have predicted that any overall change to VMT would be effectively zero, since the regulatory alternatives don't have much obvious impact on the average household's demand for travel. … Leakage in GHG control (or gasoline consumption) that is attributable to shifting the shares of VMT by vehicle ages strikes me as more plausible than leakage in GHG control (or gasoline consumption) that is generated by changes in overall VMT in the country." (Graham, B-39)

ii. Others noted that, at most, there might be a slight increase in VMT where fleet size is larger (which most likely would occur under the rollback, where new vehicle prices were estimated to be lower), but nothing on the scale NHTSA's modeling asserted.

- "…the vehicle-focused method used to calculate total VMT—using historically derived, vintage specific, per-vehicle VMT—neglects important determinants of demand that are central to the issues this update is attempting to address." Further, "…it is unlikely that the total household VMT would increase by the total annual VMT of a new vehicle," as NHTSA's model assumes. (Birky, B-22).
- "Imagine a household with multiple drivers but one car. Suppose they add a second car. It is intuitive to expect that total driving in the household (including both cars) will rise. But, it seems very unlikely that VMT would double. Similarly, as the fleet continues to rise faster than the population (as noted in the PRIA), one would not expect the total VMT to rise at the same proportional rate as the number of registered vehicles, but instead to rise more slowly." (Sallee, B-68)

iii. They also noted that much of the favorable cost-benefit analysis of the rollback (in particular, the lower number of car accidents and fatalities projected under the rollback and the fact that the rollback does not lead to the magnitude of increased fuel consumption and increased emissions relative to the Obama standards as would have been seen under prior analytical approaches) is attributable to the implausible projection that total fleet VMT will shrink as a result of the predicted decrease in fleet size .

- "Much of the final cost-benefit analysis depends on the total VMT in the fleet. … The current model assumes that the fleet VMT schedule is independent of fleet size. This is unlikely." (Sallee, B-61)
- "This exaggeration could very well be substantial. Thus my concern about this issue rivals the central concern about how the new vehicle sales and scrap responses are implemented separately." (Sallee, B-68)
- "This model feature [using set VMT schedules for individual vehicles] causes a significant disconnect in the relationship between the overall fleet size change and aggregate VMT traveled across the B and P cases." (Gruenspecht, B-49).

3. In some instances, the models' projections indicate that the models do not even perform consistently with NHTSA's descriptions of how they operate, further calling into question their validity.

a. For example, according to NHTSA, the sales model is supposed to show the effect of changes in new vehicle prices on new vehicle sales, and the scrappage model is supposed to show the effect of changes in new vehicle prices on the rates of scrappage of existing vehicles. New car price differences started in MY2017, but the sales model did not show changes in sales until MY2022—which is difficult to understand given that the sales model is supposed to show the effect of changes in new vehicle prices on new vehicles sales. In addition, the scrappage model—which is using the same new vehicle price changes (starting in MY2017), showed changes in the rate of the scrappage of vehicles in MY2018. It is difficult to understand how the

standards could have an effect on scrappage of existing vehicles before they had an effect on new vehicle sales. This reinforces the scale of the fundamental flaw of failing to connect the sales model to the scrappage model.

- "Differences in sales between the P and B cases do not begin until MY2022 even though the reported price differences start in MY 2017. Unless I have misread the output files, it would be useful to explain why differences in price levels do not affect sales prior to MY2022 or, if the model code is faulty, to update it to address this problem." (Gruenspecht, B-44)
- "While differences in new LDV prices between the B and P cases do not cause the affect [sic] new LDV sales until 2022, scrappage starts to be affected by new vehicle price differences starting in 2018. As a result, with no change in new vehicle sales, the in-use fleet reported is already 1.18 million vehicles larger in the B case than in the P case. (Gruenspecht, B-52)

b. In addition, under two different runs of the model with different new car prices, the model predicted identical vehicle sales. This outcome calls into question the basic functionality of the sales model, which is designed to predict changes in sales based on changes in new vehicle prices.

- "Another concern arises from comparisons between the CAFE and CAFE_ss[32] versions of the model runs. Although there are differences between the price paths between these two runs, representing different interpretations of limitations on manufacturers' CAFE compliance strategies, the reported sales differences between the B and P cases for MY2022 through MY2032 are identical in the CAFE and CAFE_ss output reports for total LDVs, passenger cars (PCs), and light trucks (LTs) in each year. This outcome suggests that something other than the difference in new LDV prices is driving sales differences across cases representing the B and P policy alternatives. Unless I have misread the model results, it would be useful to understand why the difference in prices between these two cases does not lead to corresponding differences in LDV sales results." (Gruenspecht, B-44)

4. Reviewers noted several instances in which "essential" information about the models was missing, which made it "impossible" to fully evaluate the models, including information on how NHTSA developed and validated the models. Because this information is not available, it is not possible for either the peer reviewers or the public to fully assess the models and to understand what other flaws might be present or what effects such flaws are having on the model's results and the analysis underlying the proposed rules.

- With respect to evaluating the approach NHTSA used for the sales model: "Some essential information is not displayed, which means that it is impossible to fully assess the model. In particular, the dependent variable is not defined. Is this regression

[32] See supra note 32.

estimated in first differences? The right hand side regressors are also not labeled clearly. Are the sales lags differenced as well, or are they in levels? Basically none of the regressors are labeled clearly enough to be sure of how the regression was run based on the PRIA." (Sallee, B-58)

- With respect to evaluating the approach NHTSA used for the scrappage model, one reviewer noted that "There were some modeling choices that I simply could not evaluate with the given information." (Sallee, B-64)
    - "As a minor (but important) point, the main estimating equation does not specify the unit of observation, nor does any table list the number of observations or unit of observation. Tables also do not present standard errors, which makes it difficult to assess many coefficient estimates. Standard errors need to be adjusted for serial correlation, and perhaps two-way clustered to allow correlation in the errors by age." (Sallee, B-64)
    - "More significantly, nearly all of the relationships of interest are polynomials. There are no summary statistics reported, so it is nearly impossible for the reader to judge the economic magnitude of the effects given what is reported (i.e., to assess marginal effects at the mean of the sample.)" (Sallee, B-64)
    - "There are very few alternative specifications shown, with the major difference being the polynomial shape of the age variable. It is simply impossible from the given set of results to judge how robust these estimates are." (Sallee, B-64)
    - "In contrast to the new vehicle sales regression reported in the PRIA's section 8.6, the discussion of the scrappage regressions does not include any discussion of the time series properties of the estimators. It is important to test for non-stationarity, for example." (Sallee, B-64)
- "The model documentation indicates that other vehicle attributes are included in the scrappage model values worksheet but it was not clear (given the scope of this review) how they figure into the model." (Birky, B-24)
- "The independent sales and scrappage functions determine ownership rates, but this result is not reported nor compared to historical trends, so it is not possible to assess how consistent the model is with these trends or with trends in VMT per household or per capita." (Birky, B-22)
- "To fully comment on the model implementation, it would be necessary to see the results of sensitivity analyses over a larger variation in inputs." (Birky, B-26)

Additional Concerns Raised by the Updated Peer Review

Should the agencies move forward and attempt to develop a reasoned, economically sound approach to modeling sales and scrappage effects, the reviewers flagged several additional issues that will need to be addressed. First, the fuel economy improvements in the vehicles under the current standards must be incorporated in some way in the sales model, such as by including in the sales modeling only the net price increase of new vehicles—i.e., the relevant technology costs less consumers' valuation of the fuel savings that will result.[33] Further, the reviewers'

[33] *See, e.g.*, Revised CAFE Model Peer Review at B-34 (Graham), B-55 to B-56 (Sallee), and B-63 (Sallee) ("The PRIA uses estimates of price that do not account for changes in vehicle quality, including fuel economy. This seems to me deeply problematic, as the right conceptual idea is to ask how a change in the desirability of vehicles, taking

comments make clear that simply increasing elasticity of demand for new cars to secure similar scrappage results while eliminating the current model's implausible effects of changes in new vehicle prices on scrappage rates of existing vehicles would not be an acceptable approach. Graham notes that a new car sales elasticity of -1 "does not have a solid grounding in economic evidence,"[34] and based on the relevant literature it is likely "well below" that level.[35] In addition, the reviewers noted that NHTSA's assumption that 100% of the technology costs attributable to the standards will be passed through to consumers (in the form of higher new vehicle prices) is unfounded. The reviewers note that the relevant economic literature "tends to find incomplete pass-through."[36] They note that, "It is likely that some of the burden of additional technology deployment will be borne by producers in the form of lost profits," and that this is especially true for fixed costs, such as research and development, as "[e]conomic theory would predict that only true marginal costs (i.e., costs that scale directly with each new unit sold) would impact strategic pricing."[37] Given this, it is likely that the new sales model "overstates the size of any effects on the new car market."[38]

In addition, we note that in the Proposed Rule, the agencies also employed a new safety model that relied upon future fatality rates (developed by NHTSA) for the fleet. In the public comments on the Proposed Rule, commenters identified several of the same types of problems in that modeling effort that the peer reviewers and public commenters noted with respect to the sales and scrappage models.[39] In particular, the public commenters thought NHTSA's choice of model was inappropriate and also that NHTSA ignored calendar-year effects—both of which significantly skewed the fatalities results in the Proposed Rule. The agencies must also conduct a peer review for the new safety model, especially given the accuracy of the public commenters in identifying the significant issues with the sales and scrappage models and the analogous flaws that those same commenters have noted with the new safety model.[40]

* * *

In sum, the peer reviewers' comments, like the public comments, leave the agencies with only two potential options—(1) revert to the previous modeling approach where sales and scrappage do not change depending on regulatory alternatives, or (2) redesign the sales and scrappage models from the ground up, based on economic theory and best econometric and modeling practice.

The same is true of the new safety model that the agencies relied upon, especially the estimation of future fatality rates for vehicles—a model the agencies still have not had peer reviewed.

---

price and attributes into consideration, changes ownership. The PRIA argues that the ideal specification ignores quality changes, but I do not understand or agree with the arguments made.").

[34] *Id.* at B-33 (Graham).

[35] *Id.* at B-35 (Graham).

[36] *Id.* at B-55 (Sallee).

[37] *Id.*

[38] *Id.*

[39] Comments of Policy Integrity at 91-99; Comments of CARB 258-82; Comments of R.M. Van Auken, Dynamic Research, Inc., NHTSA Docket #2018-0067-11881.

[40] NGO Joint Legal Comments at 190-99 (noting that the agencies are legally required to have the safety model peer reviewed).

If the agencies choose to redesign the models, then given the significant changes that would be needed and the central role of the models in the Proposed Rule and the cost-benefit analysis, a subsequent peer review of the revised models and their projections, including public participation, would be required. However, even if the agencies attempt to address the problems that have been identified, they may have to accept that—given the state of the economic literature and available modeling approaches—it is currently impossible to develop point estimates of these effects with any reasonable certainty. This is what the agencies concluded in the previous fuel economy and GHG emissions rulemakings for light-duty vehicles, and it is very likely that there are simply insufficient grounds to justify a change in that position.

In either event (whether the agencies revert to the previous modeling approach or fundamentally redesign the new models), given the significant changes that would be needed and the central role of the models in the Proposed Rule and the cost-benefit analysis, the core justification for the Proposed Rule would be undermined, and the revisions would inevitably lead to a different proposal. As a result, the agencies must withdraw the fatally flawed Proposed Rule and reinitiate the rulemaking process. Any new proposed rule, based on the revised analysis, will require a new public comment period so that interested parties can review and comment upon any new models and the revised proposed rule. In light of the central role of the sales, scrappage, and safety models in the justification for the Proposed Rule, these steps are essential before the agencies can issue a final rule.