ERG
5608 Parkcrest Drive, Suite 100
Austin, TX 78731

MEMORANDUM

TO: Allen Lyons, ARB

FROM: Tim DeFries, Sandeep Kishan

DATE: March 7, 2008

SUBJECT: ERG Responses to Peer Reviewer Comments on the California RSD Pilot Study Reports

---

As part of the final stages of the California RSD Pilot Study, ARB requested that Roger Atkinson, Brett Singer, and Michael Rogers peer review the following four reports produced by Eastern Research Group. These reports are respectively known as the literature search, modeling, implementation, and final reports:

1. Tom Austin, Andrew D. Burnette, Rob Klausmeier, Bob Slott, "Review of Literature on Remote Sensing Devices," Report, submitted to the California Air Resources Board and the California Bureau of Automotive Repair, ARB-040826, Eastern Research Group, Inc., Austin, TX, August 26, 2004.

2. Timothy H. DeFries, Thomas J. Petroski, Meredith F. Weatherby, Boonsiri Limsakul, Hugh J. Williamson, "Estimating Benefits of Improvement Strategies (including RSD) for the California I/M Program: An Inspection and Emissions Forecasting System," Report, Version 6, submitted to the California Air Resources Board and the California Bureau of Automotive Repair, ARB-070210, Eastern Research Group, Inc., Austin, TX, February 10, 2007.

3. Timothy H. DeFries, Andrew D. Burnette, Sandeep Kishan, Thomas J. Petroski, "Estimating Benefits and Costs of Improvement Strategies for the California I/M Program: Implementation Options for Using RSD," Report, Version 7, submitted to the California Air Resources Board and the California Bureau of Automotive Repair, ARB-070409, Eastern Research Group, Inc., Austin, TX, April 9, 2007.

4. Andrew D. Burnette, Sandeep Kishan, Timothy H. DeFries, "Evaluation of Remote Sensing for Improving California's Smog Check Program," Report, Version 11, submitted to the California Air Resources Board and the California Bureau of Automotive Repair, ARB-070504, Eastern Research Group, Inc., Austin, TX, May 4, 2007.

The peer reviewer comments and suggestions are attached to this memo as Appendices A, S, and R. We have inserted markers into the text of each submission to designate the start and end of each specific comment that we will address.

The large table in this memo has four columns. The first is an identifier that matches the markers inserted into the text. The second column contains a brief description of the peer reviewer's comment. The third column contains our brief response. The fourth column contains a description of the possible action that we took to address the comment.

As a result of this peer review process, we produced new versions of the four reports. These are the final report versions for the California RSD Pilot Study:

1. Tom Austin, Andrew D. Burnette, Rob Klausmeier, Bob Slott, "Review of Studies and Data Relevant to the Seven Questions of the California RSD Pilot Study," Report, submitted to the California Air Resources Board and the California Bureau of Automotive Repair, ARB-040827, Eastern Research Group, Inc., Austin, TX, August 27, 2004.

2. Timothy H. DeFries, Thomas J. Petroski, Meredith F. Weatherby, Boonsiri Limsakul, Hugh J. Williamson, "Estimating Benefits of Improvement Strategies (including RSD) for the California I/M Program: An Inspection and Emissions Forecasting System," Report, Version 7 (final), submitted to the California Air Resources Board and the California Bureau of Automotive Repair, ARB-080301, Eastern Research Group, Inc., Austin, TX, March 1, 2008.

3. Timothy H. DeFries, Andrew D. Burnette, Sandeep Kishan, Thomas J. Petroski, "Estimating Benefits and Costs of Improvement Strategies for the California I/M Program: Implementation Options for Using RSD," Report, Version 9 (final), submitted to the California Air Resources Board and the California Bureau of Automotive Repair, ARB-080302, Eastern Research Group, Inc., Austin, TX, March 2, 2008.

4. Andrew D. Burnette, Sandeep Kishan, Timothy H. DeFries, "Evaluation of Remote Sensing for Improving California's Smog Check Program," Report, Version 15 (final), submitted to the California Air Resources Board and the California Bureau of Automotive Repair, ARB-080303, Eastern Research Group, Inc., Austin, TX, March 3, 2008.

Comment ID	Brief Description	Brief ERG Response	Action Taken
A1	Modeling approach should be viewed as conceptual until independently reviewed.	This analysis was a first attempt using a new approach that, we believe, has many advantages. We encourage review.	None.
A2	Clarify in the executive summary why 24 months is used as the time scale.	We state in the modeling report body on page 1-2 that we used 24 months because the I/M program is biennial.	Modeling report: Footnote added on page ES-5.
A3	Models C and D produce widely different inventory estimates.	We used the different data sources (RSD and VID), through models, to determine the percent effect on total emissions of the different strategies. The percents of total emissions were applied to inventory estimates from EMFAC. Table 4-9 of the modeling report shows that while the different models produced different inventory estimates, NoFIM is about the same percent of NIM emissions. Thus, we have some confidence that the models estimate the changes in mass emissions produced by a strategy consistently. (Similar to S29) In any case, the primary focus of this study is to estimate relative changes in inventory and not to estimate the size of the inventory itself.	Modeling report: Text added on page 4-46 in the five lines just above Table 4-9.
A4	The size of the effects of pre-inspection repairs is not measurable by the VID, but they have a positive emissions impact.	We acknowledge that pre-inspection repairs have an impact on the effectiveness of an I/M program and that VID data cannot be used to quantify the size of these emissions benefits. However, every data source, including the VID, has shortcomings. In addition, there are other factors, besides pre-inspection repairs, (such as IM station accuracy) that affect the measured and real effectiveness of an I/M	None.

Comment ID	Brief Description	Brief ERG Response	Action Taken
		program.	
A5	Will deficiencies in C and D affect vehicle rankings?	Yes, model deficiencies will affect the rankings. However, we spent considerable effort building models and have documented the development procedures so that they could be evaluated by reviewers. Keep in mind that rankings for Directing, Exempting, and Calling-In are not based on emissions but are based on failed miles driven, which is less susceptible to modeling uncertainty. And in the case of Scrapping, rankings are based on the change in emissions after normalizing the emissions to the EMFAC-calculated inventory value.	None.
A6	RSD and ASM cannot be interconverted.	Please refer to final report Section 9 on variability. The average RSD and ASM emissions of large groups of vehicles are proportional to each other. For individual vehicles, substantial variability contributes to the ASM/RSD relationship. Nevertheless, a clear trend is still apparent. Consequently, we believe they can be interconverted. We do agree that an RSD test and an ASM test are not equivalent.	None.
A7	Either RSD or ASM must be chosen as the regulatory standard.	We don't necessarily agree since a third standard could be used. However, we used the I/M program ASM since we expect that California would want to confirm vehicle selection regardless of the method used to make the selection. We used ASM as the standard because the project questions were framed with respect to the existing IM program.	None.
A8	A mandatory ASM is the best	We think that this comment is not related to	None.

Comment ID	Brief Description	Brief ERG Response	Action Taken
	strategy for enforcement.	the goals of the project.	
A9	Agreement between RSD and ASM is described differently in final report executive summary and in Section 9.	Scatter of RSD versus immediate ASM is wide, as shown in final report Figure 9-4. Scatter of RSD versus a much later IM ASM can only be even wider. This increased scatter is reflected in the lower IM ASM fail rates calculated by the Fprob models - even for extremely high RSD values. Using high RSD cutpoints as a strategy in getting high fail rates in the targeted sample results in low capture rates of the failing vehicles in the fleet.	Final report: Two paragraphs (bottom of page 1-7 and top of page 1-8) modified and added.
S1	Securing RSD siting permits was under-emphasized in the report.	I think the reviewer is saying we should have raised more of an alarm about our inability to get rush-hour data. That problem has been worked through in the current SCAQMD study. It is no longer an issue in California.	None.
S2	Perhaps RSD costs could be lower due to bulk pricing.	We considered that in our cost estimates. Of course, the RSD data collection vendor is always free to lower his prices below those that we used in the calculations.	None.
S3	40% of VSPs are in VSP range.	This error was pointed out by every peer reviewer and several public reviewers as well. California-specific data indicate 40% of single RSD hits are in VSP range. However, we made a mistake in that multiple hits on individual vehicles would increase the likelihood that at least one of the RSD hits for a vehicle would have an in range VSP. Maximizing the number of unique vehicles covered conflicts with obtaining multiple hits to maximize the in-VSP-range percentage. As described in our report, it is very useful to refer to other programs when developing assumptions for the California situation.	Implementation report: Section 3 (Conditions for Calculating Cost-Effectiveness) has been almost entirely re-written with an entirely new analysis. The results of the new analysis cause changes to the cost-effectiveness numbers presented after Section 3 and those in the Final Report. However, the conclusions of the Final Report remain the same.

Comment ID	Brief Description	Brief ERG Response	Action Taken
		However, one must be careful in some of these comparisons because, in many ways California is quite a different situation. The former Missouri RSD program provides a good example of this. Relative to the LA Basin, the Bay Area, and Sacramento (where our team surveyed most or all freeway ramps), the St. Louis area provides a relative abundance of good RSD measurement sites. In the above areas of California, multi-lane, metered on-ramps predominate. A small fraction of these provide appropriate, adequate, and safe space for an RSD site. Obtaining more data from surface streets will help to a point, but again, in a large program the best of these sites will be quickly used up, requiring the use of less desirable sites for more fleet coverage.	
S4	Use different Fprob degradation slopes following initial passes and initial fails.	Actually, we do. Degradation slopes are parallel in logit space and therefore not parallel in linear space. In linear space degradation occurs faster following initial fails than following initial passes.	Modeling report: Re-wrote paragraphs on pages 2-24 and 2-26. Added trend lines to Figures 2-1 and 2-2 to demonstrate that modeled slopes are non-parallel failure rate trends.
S5	Validation check of RSD in Fprob models	Appendices D, E, F, and G show that models that used RSD were built on 2/3 of the data and validated on the remaining 1/3.	Modeling report: All of the models that use the RSD data are detailed in Appendices D, E, F, and G. These appendices show in detail how the RSD data is used. Added a paragraph at the end of Section 2.7 on page 2-30 that refers to the appendices.
S6	Define the type of stations used for the NIM path.	NIM is based on all stations in the VID dataset. Therefore, NIM represents the average station. We do not see how the reviewer got the impression that the NIM path was determined from high-performing	Modeling report: Added new list item (fifth bullet on the page) discussion on page 2-7. Added the third paragraph on page 4-13.

Comment ID	Brief Description	Brief ERG Response	Action Taken
		stations. We believe the use of the various pathways is correct.	
S7	Use different Fprobs following initial passes and initial fails.	Same as S4.	Modeling report: Re-wrote paragraphs on pages 2-24 and 2-26. Added trend lines to Figures 2-1 and 2-2 to demonstrate that modeled slopes are non-parallel failure rate trends.
S8	Various logical inconsistencies with regard to repair costs.	The reviewer is correct on this point. In the report, the repair cost estimates look forward in the future to one cycle beyond the most recent ASM inspection. To be correct, we would have made Fprob and cost calculations over a constant calendar period. We could do this, but the calculations are complex because of the several contingent Fprobs that are two or more cycles into the future. We have estimated the repair costs using the more accurate constant-calendar-period method and found them to be near the approximate method results in the report.	Modeling report: Added paragraph at bottom of page 4-54 and top of page 4-55 and footnote on page 4-61.
S9	Clarify how emissions variability is covered by the models and carry throughout modeling report.	Emissions variability is inherently part of the Fprob models. For example, in Models C, D, and E, emission rate, emission rate variability, and cutpoint work together to determine Fprob.	Modeling report: Added last bullet on page 2-7 and the corresponding item in Table 2-1.
S10	Examine or discuss how RSD can evaluate station performance.	I believe this is covered in the last two question sections.	Final report: Added a paragraph just before Figure 8-2.
S11	Examine or discuss how RSD can identify intermittently high-emitting vehicles.	We did not discuss any use of repeat RSD measurements even though we know that, because of high emissions variability, multiple RSD measurements are valuable. For this first analysis, we chose to use just a single RSD measurement. Nevertheless, intermittently-high vehicles would be captured by the Fprob	None.

Comment ID	Brief Description	Brief ERG Response	Action Taken
		models. Our models reflect the high variability of vehicles emissions since the Fprob models were built on all available data. That is, highly variable vehicles were not removed from the dataset. Many high-variability vehicles would not be able to be verified by ASM even if they were selected by RSD. This is not an area of deficiency.	
S12	Logic defining NIM, DX, CN, CS, SP paths	We believe the logic is correct. The reviewer's thinking that the logic is incorrect follows from his misunderstanding that the NIM pathway is based on data only from high-performing stations. NIM is based on data from all stations. It therefore represents the performance of the average station.	Modeling report: Added a paragraph on page 4-13 just before the section "Failed Miles Driven for Directing/Exempting."
S13	The text suggests that NIM is for high-performing stations.	NIM is based on all stations in the VID dataset. Therefore, NIM represents the average station. We do not see how the reviewer got the impression that the NIM path was determined from high-performing stations. We believe the use of the various pathways is correct. (Similar to S12.)	Modeling report: Added new list item as fifth bullet on page 2-7. Added a paragraph on page 4-13 just before the section "Failed Miles Driven for Directing/Exempting."
S14	We can't assume that changes in benefits using VID only and using RSD are the same.	We shouldn't assume that, and we didn't. We don't see where the reviewer got this impression.	None.
S15	Removing the existing effects of Directing in VID data	We decided not to "break out" the effects of the existing Directing program in the VID data because it would probably not be possible to do. We do not believe, as the reviewer does, that breaking it out is critically important. Our estimates of Directing gives the benefits to the existing program, which already includes some level of Directing. Since the existing IM program includes some Directing, the VID	None.

Comment ID	Brief Description	Brief ERG Response	Action Taken
		data sees the current program as more efficient than it would actually be without its current Directing. However, this is OK since our charge was to evaluate the benefits of adding Directing on the current IM program. We took this to mean adding a "Level 2" Directing strategy. Since the existing program already has a "Level 1" Directing, we are evaluating adding "Level 2" Directing.	
S16	Use different Fprobs following initial passes and initial fails.	Same as S4.	Modeling report: Re-wrote paragraphs on pages 2-24 and 2-26. Added trend lines to Figures 2-1 and 2-2 to demonstrate that modeled slopes are non-parallel failure rate trends.
S17	Clarify assumptions for forecast of NIM path	We believe the logic is correct. The reviewer's thinking that the logic is incorrect follows from his misunderstanding that the NIM pathway is based on data only from high-performing stations. NIM is based on data from all stations. It therefore represents the performance of the average station. (similar to S12)	Modeling report: Added a paragraph on page 4-13 just before the section "Failed Miles Driven for Directing/Exempting."
S18	Forecast of NIM path should be weighted average of all stations.	We believe the logic is correct. The reviewer's thinking that the logic is incorrect may follow from his misunderstanding that the NIM pathway is based on data only from high-performing stations. NIM is based on data from all stations. It therefore represents the performance of the average station. (similar to S12)	Modeling report: Added a paragraph on page 4-13 just before the section "Failed Miles Driven for Directing/Exempting."
S19	Calculating the ranking value for Directing	We believe the calculation is correct. Keep in mind that in the modeling report, we are calculating the ranking value - not the benefit. The calculations assume that directed vehicles follow the NIM pathway and that non-directed	Modeling report: The use of the DX curve for Directing is explained beginning on page 4-13 in the section "Failed Miles Driven for Directing/Exempting" and in the section "Directing and Exempting" beginning on page 4-27 and on

Comment ID	Brief Description	Brief ERG Response	Action Taken
		vehicles follow the Exempted pathway. Then, in the implementation report the calculations "moderate" the difference between the two pathways by using the result of a California study that found that average-performing stations have 80% of the fail rate of high-performing stations. We believe this is explained in the modeling report (top paragraph on page 5-15) and implementation report.	page 4-28. Added text at the bottom of page 4-28 and top of page 4-29.
S20	Calculating the ranking value for Exempting	We believe the calculation is correct. Keep in mind that in the modeling report, we are calculating the ranking value - not the benefit. We properly compare the benefits of the Exempting path with the NIM path.	Modeling report: The use of the DX curve for Exempting is explained on page 4-13 in the section "Failed Miles Driven for Directing/Exempting" and in three paragraphs on page 4-27 in the section "Directing and Exempting."
S21	Calculate the effect of Calling-In and Scrapping by comparing with the Normal I/M Process path.	Yes, this is what we did.	None.
S22	Use different Fprobs following initial passes and initial fails.	Same as S4.	Modeling report: Re-wrote paragraphs on pages 2-24 and 2-26. Added trend lines to Figures 2-1 and 2-2 to demonstrate that modeled slopes are non-parallel failure rate trends.
S23	Clarify how emissions variability is covered by the models and carry throughout modeling report.	Emissions variability is inherently part of the Fprob models. For example, in Models C, D, and E, emission rate, emission rate variability, and cutpoint work together to determine Fprob.	Modeling report: Added new item as the bottom bullet at the bottom of page 2-7 and in Table 2-1.
S24	Back-casting of trends in Figures ES-1 and ES-2.	The calculations are correct. The decrease for this specific vehicle is caused by the known ASM2525 NX failure in the previous cycle vs. the probability of overall failure in the future cycle. The car in the previous cycle has higher emissions because we know that it failed. The car in the future has lower probable emissions	Modeling report: Added the large paragraph on page ES-16.

Comment ID	Brief Description	Brief ERG Response	Action Taken
		because we don't know whether it will pass or fail at the Decision Point. Thus, the emissions for this car, which is known to have failed the previous-cycle ASM, are expected to get lower. That's the whole point of the IM program.	
S25	The next I/M cycle should have an effect for Scrapping	It does. Keep in mind that there is only a small chance that the vehicle that passed the Scrapping ASM will fail the next regular IM ASM. This is revealed by a small inflection is in each of the monthly component curves, but the effects for each month are so smoothed by the averaging effect of the probability of the next inspection in each month that the inflection in modeling report Figures ES-1, ES-2, 4-6, and 4-11 are difficult to see, but they can be seen. The calculations therefore indicate that the effect is much smaller than the reviewer expects.	None.
S26	Why not just rank all vehicles by ΔFTP instead of ΔFMD for Calling-In, Directing, and Exempting?	The intention of the IM program is to reduce emissions, but the reality is that the IM program tries to keep all vehicles in a passing status. This notion is discussed in the modeling report on page ES-17 (Ranking Vehicles Using Forecasted Benefits) of the executive summary and in Section 4.1 and on pages 5-1 through 5-3 and by Section 6.3 of the final report. The I/M program does not actually go after emissions. If it did, all vehicles above a certain single concentration would fail. Instead, I/M goes after failures by applying a cutpoint to emissions. The I/M program just goes after vehicles that fail their respective cutpoints regardless of what their	None.

Comment ID	Brief Description	Brief ERG Response	Action Taken
		emissions are. Thus using ΔFMD is correct for Directing, Exempting, and Calling-In because it focuses on the extent of a vehicle's failing status. (Note that when one set of RSD cutpoints is used for all vehicles, something that would be unthinkable in an I/M program, RSD is going after vehicles regardless of their ASM cutpoints.) On the other hand, Scrapping does go after emissions, so ranking by ΔFTP/$ is correct.	
S27	ASM to FTP conversions	The ASM (as well as the RSD) measurements are concentrations. But the study demanded effects on tons of emissions to the airshed. Thus, a conversion from ASM to FTP was needed. Any such conversion will probably always draw criticism, but we still had to use a conversion. We used the most recent one adopted by ARB. Using a modeling dataset of individual vehicles that are not representative of the distribution of emissions in the fleet does not bias the ASM to FTP equations. We are not using the modeling dataset to determine the average emissions of the fleet, so it does not need to be representative. As long as the range of emissions of the fleet and the model set are similar, the conversion can produce reasonable results - assuming no other major defects exist.	None.
S28	Clarify calculation of ΔFTP from repairs	The Fprobs that follow an initial fail, a repair, and a certified pass includes the effect of the repair, since the vehicle was ultimately certified. The effect of the repair is reflected in the Fprob and then carried via the chain to the ASM and to the FTP.	Modeling report: Added paragraph near the end of Section 2.8 at bottom of page 2-34 and top of page 2-35.

Comment ID	Brief Description	Brief ERG Response	Action Taken
S29	Wild inventories in Table 4-9 of modeling report and potential biases	We used the different data sources (RSD and VID), through models, to determine the percent effect on total emissions of the different strategies. The percent of total emissions were applied to inventory estimates from EMFAC. Table 4-9 shows that while the different models produced different inventory estimates, NoFIM is about the same percent of NIM emissions. Thus, we have some confidence that the models estimate the changes in mass emissions produced by a strategy consistently. (Similar to A3) In any case, the primary focus of this study is to estimate relative changes in inventory and not to estimate the size of the inventory itself.	Modeling report: Text added on page 4-46.
S30	Counter-intuitive repair costs	The reviewer is correct on this point. In the report, the repair cost estimates look forward in the future to one cycle beyond the most recent ASM inspection. To be correct, we would have made Fprob and cost calculations over a constant calendar period. We could do this, but the calculations are complex because of the several contingent Fprobs that are two or more cycles into the future. We have estimated the repair costs using the more accurate constant-calendar-period method and found them to be near the approximate method results in the report. (Similar to S8)	Modeling report: Added paragraph at bottom of page 4-54 and top of page 4-55 and footnote on page 4-61.
S31	The uncertainty associated with FTP emissions estimates is not adequately conveyed.	We don't mean to indicate that FTPs are exactly what a vehicle would emit. Several steps are used to estimate FTPs. We hope that FTP estimates are unbiased.	We have inserted “estimated” before FTP in all places where appropriate in the modeling, implementation, and final reports. Have modified the paragraph in the modeling report at the bottom of page ES-9 and top of ES-10 and in the final report at the bottom of page 6-10 and top of 6-11.

Comment ID	Brief Description	Brief ERG Response	Action Taken
S32	Advantages of vehicle-specific monthly mileages	We are trying to "sell" the idea that using mileages is really important, and the current IM program ignores it. Accounting for mileage in the special strategies should make the IM program more effective. We used EMFAC model year average mileages in our calculations simply because we could not complete writing the code that could estimate vehicle-specific mileages from a vehicle's VID entries. Using vehicle-specific mileage accumulation should be done in the future. Using individual vehicle mileage would improve the vehicle selection performance, but we could not finish the odometer-correction code.	Modeling report: Added footnotes on page ES-9 and 4-4.
S33	Define high-performing station	Nobody has been able to devise a way of finding high-performing stations. Nevertheless, we use the notion conceptually to investigate the benefit of Directing vehicles to them. The term "high-performing station" is used generically.	Modeling report: Added "average-performing station" and "high-performing station" to Glossary, and added footnotes on pages 4-3 and 4-28. Implementation report: Added "average-performing station" and "high-performing station" to Glossary. Final report: Added footnotes on pages 1-10 and 6-4.
S34	When should Call-Ins be implemented in the absence of an RSD measurement.	This is addressed indirectly in the modeling report text. We do not recommend a best time for call-in, but acknowledge that there is no event that "triggers" a call-in.	Modeling report: Added text to the Glossary definition of Decision Point. Discussion of Decision Point (including the case for no supplemental RSD) is located in the large paragraph on page 2-19.
S35	Depreciation of vehicle value also varies with make/model.	We agree completely, but our main point is that the value of a vehicle (as well as its emissions) is key for selecting it for scrapping. The main reason for using some estimate of vehicle value is to distinguish low-value high-emitters from high-value high-emitters. Certainly, using better estimates of vehicle	Modeling report: Added paragraph at the end of Section 3.3 on page 3-16.

Comment ID	Brief Description	Brief ERG Response	Action Taken
		value would help make scrapping even more cost-effective.	
S36	Should the No Further I/M path follow the same degradation behavior as the Exempting path since the lower-degradation effects of pre-COO-inspection repairs would be present in the Exempting path but not in the No Further I/M path?	This is a reasonable concern, but it is beyond the scope we have chosen. We have assumed that the Change-of-Ownership inspections reflect the degradation of vehicles between their regular inspections. But there are reasons to believe that pre-inspection-repair behavior is different before COO inspections and before regular biennial inspections. So, making the assumption introduces an uncertainty.	Modeling report: Added footnote on page 2-24.
S37	Replicate RSD measurements	No, we did not model the effects of multiple RSD measurements. RSD data collection specifically tried to minimize multiple RSD measurements so that a larger portion of the fleet could be covered. We acknowledge that multiple RSDs could help better target vehicles. However, all multiple hits in the field data were used in the analysis. We treated them as separate vehicles. Consider the example of two vehicles that each received three RSD hits. The first has three high RSDs; the second has only one high RSD. The first vehicle is more likely to be a true high emitter. Because our Fprob models were built on all of the data from these vehicles, our models would be more likely to select all three of the records from the first vehicle, but just one record of the second vehicle. Therefore, the models emulate the real application of the models - although not perfectly.	Modeling report: Added list item in Table 2-1 and as the sixth bullet on page 2-7 that discusses that multiple RSD measurements on a single vehicle were treated as a single observation on multiple identical vehicles to build models. The method for ranking multiple RSDs from a single vehicle is referred to on pages 4-1 and 4-2. Added footnote on page 4-2.
S38	Better captions for final report illustrations		None.
S42	The models and analysis techniques	We agree completely with this comment. We	None.

Comment ID	Brief Description	Brief ERG Response	Action Taken
	are innovative and potentially valuable in analyzing Smog Check program components. However, the methods require validation before results are assumed to accurately represent or predict real results.	believe that almost everything about our analysis is new. It presents and uses techniques new to the I/M evaluation field and new to the RSD field. The analysis does not use the “standard” techniques of the past; we judged the old techniques to be unable to adequately answer the questions that ARB and BAR asked. Because our techniques are new, they may not yet be fully developed. And they are new to reviewers. We point out in several places in the reports where biases may be present and where additional investigation needs to be performed. We point out that like any models they may not be a perfect representation of the data. However, as the reviewer apparently recognizes, we have made a substantial effort to develop a workable, logical framework that can be used to evaluate the costs and benefits of I/M program components - including supplemental RSD. At a minimum, the results of this detailed analysis should cast doubt on the ability of a supplemental RSD program to be cost-effective. We welcome constructive comments on the techniques that can improve the estimates of costs and emissions. We think that such comments will help improve the understanding of I/M programs and the capabilities of RSD.	
S43	The ASM to FTP model includes a systematic bias for high-emitting vehicles.	Developing a new ASM-to-FTP equation requires a substantial effort. ERG last did this in 1999 for BAR. (DeFries, Palacios, Kishan, and Williamson, “Models for Estimating	None.

Comment ID	Brief Description	Brief ERG Response	Action Taken
		California Fleet FTP Emissions from ASM Measurements," December 25, 1999) Proper development of an unbiased and useful ASM-to-FTP conversion model needs to consider variance inhomogeneity, uncertainty in the predicting variables, skewness of the dependent and independent variables, the inherent relationship between the dependent and independent variables, and the application that the relationship will be used in. Evaluation of any candidate relationship must also consider these same factors. Because of this, a simple linear-space evaluation of an ASM-to-FTP conversion equation can easily reach the wrong conclusions about the equation's biases. In this study, rather than devoting effort to developing an updated equation, we chose instead to use the equations developed more recently for and adopted by ARB.	
S44	The log transformation approach infuses a systematic bias in the treatment of high-emitting vehicles.	Any transformation can introduce a bias if it is not corrected for. We used the standard statistical technique to calculate a bias correction factor (BCF) for the log transformations that we used. Not using a transformation when one is called for can easily produce a fit that is entirely inappropriate for the problem. (See S43.)	None.
S45	Inconsistencies between VID and RSD results are treated in a way that systematically biases against the capability of RSD at identifying vehicles with high on-road emissions.	Throughout the report we stated that study evaluated using RSD as a supplemental component of the existing IM program – not as a stand-alone RSD program. In this context, the ASM emissions test made in the IM station environment is the reference for validating	None.

Comment ID	Brief Description	Brief ERG Response	Action Taken
		vehicle selection by any intervention strategy. The issue is complicated by the inability of RSD and ASM (even roadside ASM) to agree on the emissions status of individual vehicles (See Section 9 of the final report.).	
S46	RSD fleet coverage is under-estimated. RSD costs are over-estimated. Revisions of these should have a large impact on RSD program cost-effectiveness.	Same as S3.	Same as S3. The new analysis uses results from several earlier major RSD studies. The revisions cause RSD coverage to increase from 17% to 40% for a large RSD program in the five largest AQMDs. The revisions cause the costs for the RSD data collection to be cut approximately in half. While these changes cause the cost-effectiveness for a supplemental RSD program to improve, the overall results indicate that it is still not cost-effective. The reason for this is that other costs shift, for example repair costs increase, thus moderating the effect of fleet coverage increases and RSD data collection cost decreases. The analysis demonstrates that the benefits and costs of a supplemental RSD program are more complex and more interactive than “back of the envelope” calculations would indicate.
S47	ERG has a potential conflict of interest because it offers a high emitter profiling service as an ERG product.	The appearance of a potential conflict of interest is difficult to argue against. When taking on this project, we knew that the topic of the evaluation of RSD – regardless of the findings – would be controversial for two reasons. First, an annual contract award to ESP for a large RSD data collection contract worth millions of dollars annually might depend on our findings. Second, we perceived that expectations in the mobile source community might be that RSD should work well in any application. However, we did not	None.

Comment ID	Brief Description	Brief ERG Response	Action Taken
		anticipate that we would be suspected of conflict of interest for a minor service (emitter profiling). We consider all of the methods developed in this study for identifying high-emitting or low-emitting vehicles as profiling methods – whether they use only VID information, only RSD information, or both. Because we had developed three generations of emitter profilers for California in earlier projects, we naturally used that experience to develop the models and the approach to answer the specific questions for ARB and BAR. We believe that we can provide added value to identifying vehicles even if only RSD is used. This will result in more effective and more cost-effective identification of vehicles for California's I/M program. As a consultant to state and federal agencies, ERG's interest is in using whatever measurement and analysis techniques are needed to answer its client's questions as honestly and accurately as the project budget and schedule allow. Nevertheless, we can make errors. We depend on peer review and the general technical communication in the mobile source community to help us avoid errors.	
S48	Lack of input to the study from ESP.	We had substantial input from ESP in those areas where ESP, as an RSD data collection vendor, had knowledge. This included RSD data collection; identification of variables in	None.

Comment ID	Brief Description	Brief ERG Response	Action Taken
		the study's RSD data records; discussion of factors affecting RSD, speed, acceleration, and license plate reading accuracy; discussion of factors affecting coverage, and estimates of RSD data collection unit-hit costs and for the bulk cost of large RSD efforts. However, because we did not want to bias our results with that of an RSD vendor, who clearly would like to get a large California contract, we did not consult ESP for how to perform the effectiveness and cost-effectiveness analysis or what results we "should" get. While on occasion we talked with ESP, we purposely wanted to keep our analysis independent of ESP's analysis methods.	
R1, R2, R3, R4, R5	No peer-reviewed references, Poor editing and good info in the appendices, Non-useful comparison of studies, No Stedman or Atlanta references, and No references to other RSD work.	Most if not all of the objections raised seem to be related to the reviewer's reasonable assumption that a report with 'Review' and 'Literature' in the title would be a traditional, academic literature review. Normally such a review would comprehensively cover articles in peer reviewed and other reputable publications. Also it's format would more closely follow that of traditional, academic literature reviews. The authors of this report state very clearly in the introductory paragraphs that the report is no such review. It's purpose was very different. Perhaps a change of title to "Review of Studies and Data Relevant to the Seven Questions of the California RSD Pilot Study" would be appropriate.	Literature review report: Title of report changed to "Review of Studies and Data Relevant to the Seven Questions of the California RSD Pilot Study" and date changed to August 27, 2004.
R6	What do we want to model?	ARB/BAR asked us to evaluate and perform calculations as needed to answer the 7	None.

Comment ID	Brief Description	Brief ERG Response	Action Taken
		questions of the study. ARB/BAR determined what the 7 questions were. We chose the data to be collected and the modeling technique to answer the questions. Other methods could be used to answer the questions, but since ARB/BAR followed the project as it proceeded we believe the approach was acceptable to them.	
R7	ERG assumed that current ASM I/M program is an accurate baseline to evaluate all other programs, and this leads to an assumption that RSD must replicate the results of the ASM program.	We did not assume that the ASM I/M program was an accurate baseline. Going into the project we all knew that, like all I/M programs, an ASM I/M program has limitations that make it less than perfect. Nevertheless, for this project it was a given that vehicles selected by ANY method would need to be confirmed using the IM program test. That is, in the current California I/M program, the ASM test is the reference emissions test. The Directing, Exempting, Calling-In, and Scrapping questions all asked us to quantify the incremental benefit of adding RSD to the existing program - not what a separate RSD program would be capable of doing. Neither did the study ask us to determine the benefit of adding RSD calculated using RSD as the reference test.	None.
R9	Imagine a new cheap, accurate test	The problem with this argument is that the reviewer's hypothetical new test is designated more accurate than ASM without specifying what test is used as the reference test to determine accuracy. Presumably, the reference test is a third test - neither RSD nor ASM. Unfortunately, for the study's calculations we don't have a third reference test available to	None.

Comment ID	Brief Description	Brief ERG Response	Action Taken
		use to as the perfect standard for individual vehicle emissions. Section 9 of the Final Report demonstrates that RSD is substantially less repeatable than the ASM test. But we acknowledge that doesn't make the ASM as performed in an I/M program more accurate - I/M stations have inaccuracies - they don't always get the correct inspection result that represents the vehicle's on-road emissions over time. In this context and as explained in R7, we used the ASM IM procedure as the reference since all vehicles selected must pass the IM test.	
R10	RSD tests are low cost	In context this is more of a general statement and not a comment on our work. However, we agree that each valid measurement is low cost. We assumed they cost from between $1 and $0.75 per valid, DMV-matched test (measurement).	None.
R11	IM ASM testing is intrinsically biased	We tend to agree. However, our strategy was to mimic the (flawed) behavior of the IM program. Once that was done, it was possible to evaluate the effect of adding an RSD component to the IM program. IM ASM testing is the standard of comparison for the existing IM program. Vehicle selections will be verified by the ASM test performed at IM stations. This was a "given" going into this project.	None.
R12	We assumed that the <u>proper</u> way to use RSD measurements is to predict ASM results.	We investigated <u>one</u> way to use RSD measurements based on the following. Since ARB wanted to verify vehicle selection against the IM ASM test, the way we chose to use RSD measurements was to improve the	None.

Comment ID	Brief Description	Brief ERG Response	Action Taken
		way that vehicles were selected such that the selection improved the long-term ASM failure status (FMD) of vehicles in the IM program. Of course, this is not the only way to use RSD. Another way to set up RSD is with a parallel activity, which is independent of the existing IM program. But that is not what the 7 questions asked us to do.	
R13	The approach used to build the RSD-to-ASM correlation, while proper and justified, results in a substantial loss of information content and ability to judge relative emissions contributions from various strategies.	In our view, the sources of variability that dominate the RSD-to-ASM connection are the RSD-to-RSD sources of variability. The approach that we used is based on probability mainly because we believe it is important to keep track of the uncertainties associated with individual vehicles. Tracking uncertainties would have been difficult if we had simply used ppmRSD to ppmASM models. As we said in the report, we can never forecast for certain whether a vehicle will pass or fail in the future, but we can estimate the probability that it will fail. By selecting the vehicles that have the highest probable forecasted Failed Miles Driven, we are selecting vehicles that are the most likely to be the high emitters - according to the failing-a-cutpoint definition. Whether you predict ASM concentration from RSD or predict ASM failure probability from RSD, ultimately you will need to select or not select (a dichotomous decision) in order to call vehicles in or not. It is the lack of full-duration ASMs that restrict the availability of emissions information to the problem. Since full-duration ASM tests will probably never be part of the California IM program, using ASM	None.

Comment ID	Brief Description	Brief ERG Response	Action Taken
		failure probability may be the next best alternative.	
R14	Vehicles do not recognize FMD; emission rates are the real contributor to the airshed.	The reports spend substantial space explaining that IM program does not directly go after emissions but instead goes after vehicles that fail vehicle-specific cutpoints. Thus, while "vehicles do not recognize FMD," the IM program does recognize only whether a vehicle fails its cutpoints. Note that the way that RSD measurements are generally used is by applying RSD cutpoints, which are the same for all vehicles. Thus, as it is used, RSD does not recognize the differing emissions-reduction capabilities and limitations of different ECS technologies. In our opinion, this is a major disconnect between the guiding concept of the IM program and the way RSD measurements are currently used. In this study, we chose to go with the IM program concept because we were charged with looking at the incremental benefit of adding RSD to the IM program - not replacing the IM program. To do this we had to develop ways to use the RSD measurements that were consistent with the notion of ASM cutpoints.	Modeling report Section 5.1 discusses evaluating fleet benefits.
R15	RSD strategies targeting only the highest emitters would receive no more credit for finding these vehicles than for finding vehicles only slightly exceeding the ASM test standard.	This is not true. Our models do not predict whether a vehicle will pass or fail the ASM test. Instead, the models predict the probability that a vehicle will fail the ASM test. The calculation of the probability is based on the measured RSD HC, CO, and NX concentrations and the ASM cutpoints. Thus, if a vehicle has high measured RSD concentrations (relative to the ASM	None.

Comment ID	Brief Description	Brief ERG Response	Action Taken
		cutpoints), it will have a substantially higher probability of failing an ASM test. Thus, our models do give more credit to vehicles with very high RSDs in comparison to those with borderline RSDs. In addition, our models also give more repair benefits to vehicles with higher RSD emissions because the estimated FTP emissions associated with high RSD vehicles are higher than the estimated FTP emissions associated with a moderate RSD vehicle.	
R16	Are these models appropriate for modeling the impact of RSD in the California program? Why not just select the top 1% of RSD emitters of each model year?	For a strategy to be demonstrated to be effective, it needs to go after all "high-risk" vehicles. Because the selection must be confirmed by the existing IM program, which is based on an ASM fail, and because the "risk" exposure is over the two years between inspections (not just at the time of the inspection), one way to identify the high-risk vehicles is to consider those that have a high forecasted FMD. Selecting just a rarified slice of high-RSD vehicles doesn't go after the bulk of the vehicles that will have high FMD over the next two years. Thus, with that approach, many high-risk vehicles would not be selected. Note that in the modeling report we evaluated this sort of approach of selecting vehicles based solely on their RSD measurements (See the thin black lines in Figures 5-4 through 5-34). We found that in all of the cases tested, selecting vehicles simply based on elevated RSD measurements was inferior at capturing emissions in comparison with other methods - some of	None.

Comment ID	Brief Description	Brief ERG Response	Action Taken
		which used RSD measurements in other ways. That is, we found better ways to more effectively use RSD measurements for vehicle selection than simply using the RSD measurements directly.	
R17, R18	In-range-VSP UVD to UVD should be 70-90% not 40%. The ratio of Raw to Valid readings should be higher since many of these same vehicles have invalid readings due to out of range VSP.	This error was pointed out by every peer reviewer and several public reviewers as well. California-specific data indicate 40% of single RSD hits are in VSP range. However, we made a mistake in that multiple hits on individual vehicles would increase the likelihood that at least one of the RSD hits for a vehicle would have an in range VSP. Maximizing the number of unique vehicles covered conflicts with obtaining multiple hits to maximize the in-VSP-range percentage. As described in our report, it is very useful to refer to other programs when developing assumptions for the California situation. However, one must be careful in some of these comparisons because, in many ways California is quite a different situation. The former Missouri RSD program provides a good example of this. Relative to the LA Basin, the Bay Area, and Sacramento (where our team surveyed most or all freeway ramps), the St. Louis area provides a relative abundance of good RSD measurement sites. In the above areas of California, multi-lane, metered on-ramps predominate. A small fraction of these provide appropriate, adequate, and safe space for an RSD site. Obtaining more data from surface streets will help to a point, but again, in a large program	Implementation report: Section 3 (Conditions for Calculating Cost-Effectiveness) has been almost entirely re-written with an entirely new analysis. The results of the new analysis cause changes to the cost-effectiveness numbers presented after Section 3 and those in the Final Report. However, the conclusions of the Final Report remain the same. Table 3-7 now shows that the ratio of in-range-VSP UVD to UVD is now 79%, which is in agreement with the reviewer's expectation. Table 3-7 now shows that the ratio of Raw to Valid readings is now 1.54 instead of 1.33.

Comment ID	Brief Description	Brief ERG Response	Action Taken
		the best of these sites will be quickly used up, requiring the use of less desirable sites for more fleet coverage. (Same as S3.)	
R19	The rest-of-state vehicles have different VMTs than those that are registered in the area.	The reviewer said our assumption is unlikely. We agree that it is a rough assumption. It is a small part of the fleet and because these are commuting vehicles they are probably better maintained. We said we assumed that vehicles from 'other' areas would be measured at the same rate as local vehicles. This is basically saying that all vehicles from 'other' areas are daily commuting vehicles. It is an optimistic assumption from the perspective that it will increase the statewide coverage we calculate. We think the reviewer thinks we should make this assumption less optimistic.	None.
R20	Other ways to use RSD were not considered, e.g., targeting high emitters, voluntary clean-screening sites, smart signs, dedicated RSD facilities.	There are many ways to use RSD. We had to investigate ways for which we could solidly quantify the mass emissions benefits. Some of these come down to logistics, e.g. advertising so that people go to voluntary clean-screen RSD sites. We did specifically consider targeting high emitters using RSD. That's what calling-in and directing using RSD only was. We did estimate voluntary clean-screening because we included publicity costs.	Added text in the first paragraph under "Intervention Activities Evaluated in This Report" on page ES-6 of the modeling report.
R21	Assuming the VID is free gives unfair advantage to non-RSD targeting methods.	There will be a VID - even with an RSD-only IM program. It is important to remember that we were told to determine the incremental benefit of adding an RSD component to the IM program. Thus, for this analysis, RSD never replaces the existing IM components. So, the VID and its associated costs must be included in both the without-RSD and with-	None.

Comment ID	Brief Description	Brief ERG Response	Action Taken
		RSD programs.	
R22	The behavioral aspects of RSD monitoring and enforcement were not considered.	This subject was considered during the study. A balanced discussion of the influences of enforcement via RSD would be interesting, but speculative. Apart from a multi-year study, that included actual high emitter detection (via RSD) and enforcement (which would require legislation) the only method for obtaining data on these effects would be opinion surveys, focus groups, discussions with experts, etc. These all require interpretation of people's opinions. To properly compare the influences and costs of enforcement via RSD to the current situation with Smog Check, it would be necessary to study the influence of other methods of enhanced enforcement. For example, in addition to fining and suspending inspectors, mechanics, and station owners, California could explore the impact of issuing citations to motorists who knowingly avoid program requirements. Another approach to study would be greatly expanded public outreach and education campaigns. The Pilot Project budget would not support an adequate assessment using this approach. A balanced discussion of these issues would also require the exploration of all types of influences the presence of RSD would have on behavior – positive and negative. For example, it is likely that many people would indeed be made more aware of their	None.

Comment ID	Brief Description	Brief ERG Response	Action Taken
		responsibilities for driving a clean vehicle and would respond positively. However, those most incentivized to avoid detection would be easily able to do so. Unlike a motorcycle policeman, RSD did not (at the time of the study) have the ability to hide under a bridge, and good locations for speed traps greatly outnumber those for RSD. Also, methods available to vehicle drivers for invalidating RSD readings are simple and quite difficult to mitigate against. For example, removing one's foot from the throttle and decelerating through the RSD site, or the opposite; accelerating at wide open throttle will cause an invalid reading in most cases and will cause unrepresentative emissions in all cases.	
R23	Most states can get 50-80% RSD coverage.	When it was operating, the Missouri RSD Clean Screen program routinely obtained 50% coverage of the St. Louis area fleet. More than 95% of RSD data in Missouri came from 60 sites. Some claim that 50% coverage is easily obtained and even higher coverage is possible. However, in actual programs the highest coverage so far has been around 50%. The Missouri program would certainly have liked to have 75% or even 80% coverage (as the Greeley project predicted was possible), but it did not. There were factors in play that limited their coverage to around 50%. The Greeley project obtained higher coverage, but it is even less like California than St. Louis and it wasn't a real program. In comparing this to California we looked at Sacramento data. Sacramento is representative of California.	Implementation report: Added Section 3.6 describing California RSD coverage issues.

Comment ID	Brief Description	Brief ERG Response	Action Taken
		Daily VMT in the St. Louis area is almost double that in the Sacramento area. In proportion to its VMT, Sacramento has far fewer freeway ramps appropriate for RSD. During its RSD program, the St. Louis area had no multi-lane, metered ramps. In comparison, 90% of Sacramento's freeway traffic uses a multi-lane, metered ramp to enter the freeway. The large majority of these types of ramps are not suitable for RSD use. From a total of 471 visually surveyed on-ramps in the Sacramento area, contractor survey teams identified 22 suitable, high-volume ramps. Five of these were rejected by the RSD data collection teams, mainly due to stricter safety criteria, leaving 17 sites. Following the pattern of the Missouri program (i.e., suitable, high-use sites per VMT), Sacramento would have little more than half the number of suitable sites as the St. Louis area from which to obtain 95% or more of its RSD data. At the least this shows that the Sacramento area would have a much tougher time obtaining 50% coverage than the St. Louis area.	
R24	The 50-80% coverage corresponds to 90% of the light-duty VMT.	1) We did include the effect of VMT in our analysis. We used the distribution of 69,629 vehicles, which were collected based on RSD hits. The model year distribution (shown in the report) is biased newer than the registration distribution. 2) The estimated VMT of each vehicle was an input to the selection method. 3) The IM program is vehicle-based not VMT- or mass-emissions-based. 4) It seems more accurate to say that 50 - 70% coverage	None.

Comment ID	Brief Description	Brief ERG Response	Action Taken
		corresponds to 55 - 77% of light duty VMT. According to Missouri RSD data the relationship is a little over 0.9 to 1. Our task did not include looking at ways to replace Smog Check. We had to assume that Smog Check was as effective as the state claims. Unlike RSD, Smog Check data points are proportional to vehicles, not VMT. We had to look at RSD this way. We accounted for VMT in the benefits.	
R25	The major benefit of emissions enforcement is how it changes the behavior for vehicles that witness enforcement activities. So, using IM station inspection results under-estimates the RSD benefits.	Same as R22.	None.
R26	Reducing mass emissions is more important than reducing failed miles driven.	We agree. That's why we spent considerable pages discussing this comparison. For example, see pages 5-1 through 5-3 of the modeling report. Unfortunately, the IM program only goes after failures - not mass emissions. A vehicle that passes the I/M test may produce higher emissions over the biennium that a failing vehicle does. So, we had to take into account not simply the biennial mass emissions of a vehicle but also the likelihood that the vehicle would fail the IM test. This led us to rank vehicle for selection based on Failed Miles Driven. But when we evaluated the vehicle rankings, we calculated the biennial mass emissions because we know that emissions are most important to the airshed. We think that the reviewer simply misunderstood what we did.	None.

Comment ID	Brief Description	Brief ERG Response	Action Taken
R27	VID information cannot be used to evaluate the newest model year high emitters since little data exists for them in the VID.	While this is true, we don't see that the newest vehicles represent a large emissions risk. While they contribute the most to VMT, the newest vehicles are the cleanest with low probabilities of being broken.	None.
R28	The experience of other states with RSD programs indicates that 50% of the RSD fleet can be covered with complete data. The resulting cost estimates in the reports are at least 10 times too high.	See S3.	Implementation report: Section 3 (Conditions for Calculating Cost-Effectiveness) has been almost entirely re-written with an entirely new analysis. Implementation report: Added Section 3.6 describing California RSD coverage issues. The results of the new analysis cause changes to the cost-effectiveness numbers presented after Section 3 and those in the Final Report. The new RSD costs are about half of the previous version's RSD costs. However, the conclusions of the Final Report remain the same.
R29	The cost of maintaining and updating the VID needs to be considered in evaluating relative program costs.	Same as R21: It is important to remember that we were told to determine the incremental benefit of adding an RSD component to the IM program. Thus, for this analysis, RSD never replaces the existing IM components. So, the VID and its associated costs must be included in both the without-RSD and with-RSD programs.	None.
R30	Careful planning and execution of the program is required and this activity may need to be independent of the use of RSD for differing purposes (e.g. fleet characterization vs. special strategies) due to potentially conflicting measurement requirements.	We agree. Also, we see little benefit to be obtained by targeting a few vehicles found in a small RSD program that is designed primarily to estimate fleet emissions.	None.
R31	The reports fall short of the goal of	This study had a defined scope and not an	None.

Comment ID	Brief Description	Brief ERG Response	Action Taken
	finding ways - in a neutral context - to use RSD to supplement the existing IM program.	unlimited budget. We were to focus only on the seven questions presented to us. Additionally, we were expected to solidly quantify the emissions benefits and costs of certain special strategies in the context of the existing IM program as defined by the four questions that dealt with Directing, Exempting, Calling-In, and Scrapping. We found that to be able to do this, we could not definitively address all possible RSD uses. We addressed those that we could most clearly envision. We had to leave others for other projects. We have used RSD data and VID data in many different projects. We went out of our way to build new and improved methods for selecting vehicles using RSD data as well as for selecting vehicles using non-RSD data. We may have made errors, but we have been neutral.	
R32	Costs are over-estimated by over an order of magnitude.	If this is true, then ESP should be willing to annually collect at least one in-range-VSP, valid RSD measurement on 2,271,613 unique, California I/M vehicles that have matching DMV records for an annual fee of $3,163,813. Order of magnitude is an overstatement. (See S3.)	Implementation report: Section 3 (Conditions for Calculating Cost-Effectiveness) has been almost entirely re-written with an entirely new analysis. Implementation report: Added Section 3.6 describing California RSD coverage issues. The results of the new analysis cause changes to the cost-effectiveness numbers presented after Section 3 and those in the Final Report. The new RSD costs are about half of the previous version's RSD costs. However, the conclusions of the Final Report remain the same.
R33	The analysis framework assumes that the major benefit of RSD is early detection of high emitters that would be detected by the Smog	Yes, we used early detection by RSD as the framework for the analysis, but it is not true that the IM program would have caught the high emitters anyway. That is one of the	None.

Comment ID	Brief Description	Brief ERG Response	Action Taken
	Check program anyway. This undervalues the benefits of a real RSD program.	problems of supplementing the IM program with RSD: The IM program and RSD will largely disagree on which vehicles are the high emitters. Since the IM test would be used to confirm proper selection, RSD's reputation will suffer. As we said in the report, it's not RSD's fault. No test (not even a roadside ASM) would be able to select vehicles that would substantially agree with the IM station test results. We believe that any roadside test (including RSD) when used as a supplemental test for special vehicle selection strategies will produce results that are inconsistent with the IM program results. To be able to effectively use RSD, we believe may require throwing out the entire IM program, and this was not an option that California allowed us for performing the study.	
R34	VID costs are not attributed to ASM while many other costs are attributed to RSD.	We do not treat VID costs as sunk for RSD only. VID costs are the same for both alternatives: selection by VID and selection by VID + RSD. That is why we do not consider VID costs. We were asked to evaluate the costs of adding RSD to the existing IM program; we were not asked to evaluate a stand-alone RSD program.	None.
Public 1	The calculations used 60% percent as the fraction valid RSD records that could be matched to the DMV. This value is too low.	ESP cites a fraction of 90%. We believe the difference in values may be because of the basis for the calculation. Our 60% value is based on DMV matches of all valid RSDs. ESP's 90%+ value may be based on the all valid RSDs that have readable California license plates, which would not include out-of-state, obscured, and fuzzy-image plates. In	None.

Comment ID	Brief Description	Brief ERG Response	Action Taken
		any case, since the cost of the RSD measurements is based on the number of valid, DMV-matched RSD readings, the percent of valid RSDs that can be DMV-matched would be irrelevant.	
Public 2	Did we take into account that RSD data collection collects data with respect to VMT? The RSD data will therefore favor newer model year vehicles compared to the vehicle registration distribution.	We have accounted for this. Figure 4-1 of the implementation report shows a comparison of model year distributions of the 69,629-vehicle dataset used to build models and evaluate effects and the IM fleet. For observations to be in the 69,629-vehicle dataset, the data had to be complete. Thus, they had to have RSD measurements. That's why the figure shows a bias toward newer model years. We considered adjusting the model year distribution prior to analysis so that the set would reflect the IM fleet. However, then the dataset would not reflect the data that would be collected in a real RSD effort.	None.

Appendix A

Comments of Roger Atkinson

Review of the following reports to CARB and BAR:

#1 **Estimating Benefits of Improvement Strategies (including RSD) for the California I/M Program: an Inspection and Emissions Forecasting System**

#2 **Evaluation of Remote Sensing for Improving California's Smog Check Program**

#3 **Estimating Benefits and Cost of Improvement Strategies for the California I/M Program: Implementation Options for using RSD**

Firstly, let me point out that I am not a statistician, nor do I have any expertise in vehicle emissions monitoring or modeling. My comments and evaluation of these three reports are therefore those of an interested bystander, working in the area of atmospheric chemistry.

I found the reports interesting and informative. I first make some general comments about what I perceive is the take-home message from these reports, noting that there is significant overlap (and in some cases, redundancy) between these three reports:

1) Remote sensing measurements of on-road vehicle emissions and the emissions measured by the California I/M program test (hereafter termed the ASM test) are very different and not easy to compare:

- Remote sensing (RSD) captures a vehicle's emission over a ~0.5 sec duration, and only a fraction of these measurements are under engine load conditions which are considered acceptable for emissions use. The data do, however, pertain to emissions from on-road vehicles under real-world conditions.

- The ASM test is of 90 sec duration, with engine loads within the "acceptable' limits. As noted in the reports, these data may not be representative of the vehicle shortly before the test, because of pre-test repair and/or maintenance.

2) Replicate ASM tests on a given vehicle (or set of vehicles) shows a significant degree of variability, significantly outside of the instrumental measurement uncertainties (Report #2, page 9-8).

3) Replicate RSD tests on a given vehicle (or a set of vehicles) results in much larger scatter than is the case for replicate ASM tests (Report #2, page 9-7). This in part can be explained by the 0.5 sec RSD test duration versus the longer (90 sec) ASM test duration and the increased averaging in the longer test.

4) Consistent with these conclusions, comparison of RSD test data with ASM test data for a suite of vehicles also shows a high degree of scatter (Report #2, page 9-10).

[A6 start] These "findings" immediately show that RSD test data and ASM test data are not equivalent and cannot be interconverted [A6 end], and that [A7 start] one or the other has to be chosen as the standard for regulatory purposes. [A7 end] It appears that for the time-scales relevant to RSD measurements, emissions from vehicles are inherently time-dependent, even for constant engine operating conditions. RSD is viewed in these reports as being a passive system (i.e., non obligatory for the vehicle owner), and as such can achieve only a relatively small vehicle fleet coverage with a law of diminishing returns (the number of RSD measurements increases more rapidly than the number of vehicles observed). For example, the report states, consistent with data from California and other states, that getting a number of RSD measurements which equals 50% of the registered vehicles in California will translate into usable data for 17% of California's vehicle fleet. [A8 start] These observations all suggest that the mandatory ASM test is the correct strategy for enforcement purposes. [A8 end]

However, RSD is a very valuable additional resource. As noted above it captures real-world emissions and can rapidly collect emissions from a large number of vehicles. As such, RSD appears optimum for capturing the emissions of a fleet of vehicles representative of those passing the RSD site (which may differ from the California average fleet in terms of age, etc.).

The contractors have used the California vehicle inspection database (VID) to significantly extend its predictive abilities beyond the present I/M pass/fail predictions. Of regulatory interest, this includes being able to predict the miles driven in failure mode and pollutants emitted as a consequence of this. The basic *framework* involved in these extensions of the present usage of the VID should be of interest to CARB and the BAR – while there may well be arguments about the precise statistical derivations and numerical values used, I believe the basic framework presented in these reports can provide the necessary information for future advances in better and more sophisticated use of the available database for vehicle emissions, which is of course continually expanding.

I find no obvious (at least to me) omissions or errors in the contractors' conclusions that RSD is not a cost-effective way to obtain additional information for directing likely-to-fail vehicles to test-only centers, for exempting vehicles from the mandatory biyearly inspections, for identifying vehicles for additional out-of-cycle ASM testing, or for identifying vehicles for out-of-cycle ASM testing and potential scrapping. These conclusions follow logically from the limited coverage of RSD measurements and the high variability of those measurements for a single vehicle, plus the estimated high cost of manned RSD measurement stations.

Additional comments on the individual reports are:

#1. Estimating Benefits of Improvement Strategies (including RSD) for the California I/M Program: an Inspection and Emissions Forecasting System

This report is basically a review of I/M and an extension of present methodology to better use the VID to explore avenues to reduce vehicle miles in failure mode and associated pollutant emissions due to operation in failure mode. I found this report interesting and informative, and view the proposed (or potential) extension of the use of VID as a definite advance. [A1 start] However, at this point in time I would suggest that the "models" developed by the contractors be viewed as conceptual, and that further independent work needs to be carried out to see how robust they, and the contractors' general approach, really is. [A1 end]

- [A2 start] Page ES-5, last few lines. The authors need to explain why 24 months is chosen as the time-scale for changes. While it becomes evident later on that this is the time between regularly scheduled I/M inspections, some clarification here would be useful to the reader. [A2 end]

- [A3 start] Page 4-45. Use of Model D leads to a significantly higher emissions inventory than does use of Model C – why? The authors need to discuss the reasons for this difference and for the other differences in emissions inventories for the California vehicle fleet estimated from the subset of 69269 vehicles studied. Do these differences indicate fundamental problems in Models C and/or D, or not? [The emissions inventory obtained using Model C is a factor of ~1.5 lower than the official inventory for HC and NO_x, and that using Model D is a factor of 1.2 higher for NO_x and a factor of 2.1 higher for HC than the official inventory] Significant additional discussion is definitely needed here. As I read the Table and associated text, this emissions inventory is for the 1976-1998 vehicles in California which were subject to I/M inspections and testing in 2004. The South Coast Air Quality Management District website lists 2002 emissions inventories for on-road vehicles of 360 tons/day of HC and 611 tons/day of NO_x. Since luckily not all Californians live or drive in the South Coast air basin, a rough guess would be that the on-road California emissions inventory would be a factor of ~2 higher than the South Coast air basin inventory. The newer, <6 year old, vehicles are presumably a relatively small fraction of the total inventory, so the inventory in Table 4-9 for California (256 tons/day of HC and 423 tons/day of NO_x) seems very low. Some comment or additional discussion is in order here. See also Report #2, page 1-13, which notes that an emissions inventory derived from RSD data differs significantly from the "official" inventory. There is a related section in Report #2, page 8-8, which compares an emissions inventory for the South Coast air basin from RSD measurements combined with fuel sales with the official inventory (144 tons/day of HC and 177 tons/day of NO_x), which again is markedly lower than the on-road emissions listed at http://www.aqmd.gov/aqmp/07aqmp/draftfinal/Chapter_03.pdf (see numbers cited above). [A3 end]

- [A4 start] Page 5-3, last half of page. The authors note here (and elsewhere) that there is a higher failure rate for unannounced emission tests compared to regularly scheduled ASM tests, and this is (very reasonably) attributed to pre-test repair and maintenance prior to the regularly scheduled ASM tests. This would seem to be a positive, since vehicle emissions are being reduced more than the ASM test results indicate and hence additional emissions reductions arising from the Contractors' extension of VID usage would occur. [A4 end]

- [A5 start]Page 5-10. first full paragraph. How dependent are these rankings on Model (C or D) and how dependent are they on the correctness of the model. If Models C or D have serious deficiencies (see comment above concerning comparison with the emissions inventory), would all of this change? [A5 end]

#2. Evaluation of Remote Sensing for Improving California's Smog Check Program

This report specifically discusses the RSD pilot study and its potential uses for enhancing the I/M process, and to some extent can be viewed as an appendix to Report #1. The contractors conclude that in practical terms RSD is limited to no more than about 50% coverage of the California vehicle fleet, and that the implementation costs make RSD not cost-effective for emissions reductions purposes if an integral part of a regulatory I/M program. In fact, obtaining RSD measurements on 9.5 million vehicles (50% of California's fleet) is estimated to result in usable RSD data for 17% of California's vehicle fleet subject to I/M inspections. This, together with the estimated cost of such a program, is the key problem in any attempt to cost-effectively integrate an RSD program into the regulatory I/M program; the relatively small benefits which might be obtained are not cost-effective. It is quite possible that this conclusion could be altered if automatic (unmanned) RSD stations could be implemented and be shown to be reliable.

However, RSD can play an important role in monitoring on-road emissions from sufficient numbers of vehicles to act as a real-world check on vehicle emissions (and hence emission standards and I/M program efficiency) as a function of time (i.e., trends as a function of time). In this mode, RSD can be separate from the I/M program.

- [A9 start] Page 1-7. The discussion in the last paragraph (good consistency between RSD readings and immediate roadside ASM test) seems at odds with the section in the same report on pages 9-9 through 9-14. On page 1-7 it is stated that there is good agreement between RSD tests and immediate roadside ASM test data, whereas on page 9-11 it is stated that "the wide scatter of the points is troublesome". These two apparently conflicting conclusions need to be resolved – the data presented suggest that the statements on page 9-11 are more appropriate. [A9 end]

- The conclusions to this report on page 10-1 appear eminently reasonable and are supported by the data presented in the body of this report.

#3. Estimating Benefits and Cost of Improvement Strategies for the California I/M Program: Implementation Options for using RSD

This report is largely concerned with the costing of the various potential strategies for use of RSD in an integrated I/M program. I only deal with issues that pertain to the science as it affects Reports #1 and 2.

- Table 4-2 and associated text on page 4-4 again deal with emissions inventories for California's vehicle fleet. Here the numbers are (tons/day) 392 for HC and 601 for NO_x for all vehicles, corresponding to 282 tons/day for HC and 466 tons/day for NO_x for those vehicles which are in the I/M program. These are actually the same numbers as in Report #1, page 4-45, except that they are in "English tons" [which appears to be 2000 lb, often termed a 'short ton", whereas the now-obsolete British ton was 2240 lb] rather than metric tons. The authors need to use a consistent set of units, and metric tons would seem preferable since they relate directly to emissions in terms of gram per mile.

Roger Atkinson

Air Pollution Research Center
University of California
Riverside, CA 92521

Appendix S

Comments of Brett Singer

External Review of Study on the Potential for Using Remote Sensing Devices (RSD) to Improve California's Smog Check Program.

Reviewer: Dr. Brett C. Singer
Date: 06 August 2007

I. Summary Assessment

ERG has conducted a serious and extensive analysis of the potential benefits of using remote sensing to enhance the effectiveness of California's Smog Check Vehicle Inspection and Maintenance (I/M) Program. The researchers who led and conducted this review have extensive experience and a reputation for high quality work in this field; this study will add to the excellent reputation. They have produced an impressive collection of reports and appendices that are well organized and describe their methods and results in a transparent manner. These reports deal with the advantages and limitations RSD in a balanced and objective manner. The study examines most of the ways in which remote sensing can be used enhance an I/M program, and addresses in a comprehensive manner all of the specific program enhancements proposed for analysis by ARB and BAR. The approaches that constituted the basis for analyzing benefits of program changes are fundamentally valid and appropriate. ERG correctly identified the importance of I/M benefits occurring over time. The performance metrics on which they based their analysis – ΔFMD (change in the number of miles vehicles travel while in a condition that would cause a failing Smog Check test), and ΔFTP (the estimated change in full cycle emissions, i.e. that would be measured if vehicles were operated in a manner analogous to federal test procedure) – are highly relevant to the objectives of the study. The analysis and modeling methods that ERG had to develop in order to use these metrics are innovative, statistically sound, and generally well supported by relevant data (see specific comments below for some caveats). The approaches outlined in this study thus represent an important and valuable step forward in our collective ability to forecast and thus analyze potential I/M program variants. This review raises questions about several specific elements of the analysis implementation, including key assumptions and the logic used to develop benefit estimates for the program options. These concerns, if valid, suggest the need for changes in the analysis that could produce different results for estimated benefits under all ranking schemes. I don't believe it can be determined a priori if these changes would in turn affect the overall conclusions of the study. The conclusion that RSD cannot be cost-effectively used to enhance I/M program benefits appears to be primarily rooted in the high cost and low overall fleet coverage potential of employing RSD for I/M purposes. The analysis on which these findings are based appears to be robust and well supported. Nevertheless, the concerns raised in subsequent sections of this review should be considered and potentially addressed. The value of addressing these methodological issues is not limited to evaluating the potential for RSD; it is also important to arrive at robust and valid estimates of the potential benefits of employing program enhancements based on VID only data.

My review focused most intently on the February 10 "Modeling" report as that describes the base methodology including models of failure probability, ranking of vehicles for inclusion in the various program options, and estimation of benefits and costs per vehicle included. The other reports drew heavily on the base methods explained in this report. The next focus was on the

Final Report, specifically on the latter sections pertaining to RSD coverage issues, and emissions variability issues including RSD-ASM and ASM-ASM correlations. (Earlier sections mostly summarized material from the Modeling and Implementation Reports.) The literature review and the April 9 "Implementation" reports were reviewed in a relatively cursory manner due to time limitations and necessary prioritization.

Just below I provide summary comments on the specific set of scientific issues on which reviewers were asked to focus. Following this, I provide additional comments, concerns, and questions that pertain to methodology.

II. Specific Issues to be addressed by Peer Reviewers

1. RSD Fleet Coverage.

The Final Report presents a convincing case for the limitations of fleet coverage. The inclusion of data from other states was illustrative and helpful. [S1 start] The securing of permits to use RSD more widely than it has previously been allowed in California is an important implementation challenge that was perhaps under-emphasized in this report. [S1 end] The slope of increasing costs for increasing fleet coverage makes sense from the perspective of having to utilize decreasingly suitable sites. [S2 start] It seems there could be some savings associated with increasing the overall size of the data collection contract (bulk pricing!) and that this could counter-balance the increased cost of needing to make more and more measurements to capture vehicles that haven't yet been measured; but I have confidence in ERG's experience and knowledge in making this evaluation. [S2 end] [S3 start] The least-well-supported element of the coverage discussion is the assertion that 40% of the measurements will be within a valid VSP range. First, this percentage is supported by limited data. Second, I would think that the percentage could be higher at low coverage (as the best VSP sites would be used first) and should decrease as the fleet coverage rate increases (need to use less suitable VSP sites). [S3 end] I have no firsthand experience with this and cannot offer data supporting that an estimate of how the actual rate of valid-range VSPs is likely to vary with fleet coverage.

2. Calculated Smog Check Failure Probabilities.

The failure probability models are reasonable and at least for the VID data seem to be supported by analysis. I identified two potential concerns about these models. [S4 start] The first relates to the assumed degradation relationships (increasing failure rates) following an initial pass (IP) versus an initial fail (followed by a pass, or FP) ASM test. In short, I believe the data indicates that different slopes should be used for these two situations. Minimally, there should be a more complete and transparent analysis of this issue; the current discussion (in place of quantitative analysis) around Figures 2-1 and 2-2 is insufficient and may reach the wrong conclusion (see below for additional comments on this point). [S4 end] [S5 start] My second concern regards the analysis of the relationship between RSD and subsequent failure probability. Section 9 of the Final Report provides convincing data and analysis showing the poor correlations between RSD and subsequent ASM for a small number of vehicles. However, since the central focus of this study is on the potential for RSD to enhance Smog Check, it would be valuable to see a more complete analysis of the relationship between RSD and subsequent ASM tests for a larger fleet. I may have missed this in my reading of the documents, but it is not clear to me exactly how or whether there was any validation check of the use of RSD data in the Fprob models. Since this is an issue that is central to the analysis and should be highlighted in the method descriptions. [S5 end]

3. Methodology to Estimate Emission Benefits.

This is the area on which many of my comments and concerns are focused. Please refer to the comments in the following section. As noted above, I believe the basic modeling approach to be valid, appropriate and in some ways truly innovative. I have concerns about several of the specific model assumptions; these mostly focus on the logic used in setting up the scenarios including the normal I/M (NIM) pathway. [S6 start] The current NIM appears to assume that all (or the maximum number of) vehicles are tested in so-called “high performing stations” (note that this overlaps the Call-In pathway). And the calculation of Directing benefits assumes that in the absence of Directing, all of those vehicles would be tested at low performing stations with no associated benefits (same as Exempting). Neither of these assumptions is reasonable or consistent with the logic of the program changes being considered. A change in the NIM pathway would impact the benefits calculated for all policy options. I am additionally concerned about the pathways projected for other options including Scrapping and the models used to calculate costs for each pathway. [S6 end] [S7 start] The use of parallel slopes of increasing failure probability for initial pass and initial fail vehicles (incorrect to my understanding, as noted above) also impacts the modeling of benefits. [S7 end] [S8 start] There are some other, more minor, logical inconsistencies, such as higher repair costs being associated with options having fewer ASM tests (e.g. Exempting vs. NIM and CS vs. CN). [S8 end]

4. Methodology to Estimate Cost Effectiveness.

The results for cost effectiveness appear to be robust. The changes to the implementation of benefits calculations indicated above and in the specific comments that follow in subsequent sections may lead to substantial changes in benefits estimates for all scenarios. There is no way to evaluate at this point how these changes to benefits estimates will impact the cost effectiveness results and conclusions.

5. Using RSD to Analyze Smog Check Benefits and Characterize Fleet Emissions.

In Chapters 8 and 9 of the Final Report, the contractor provides a nice, if abbreviated discussion of the potential for using RSD data to characterize fleet emissions and conduct alternative, non-VID-based evaluations of Smog Check benefits. The discussion of vehicle emissions variability with supporting analysis is an especially valuable (and necessary) component of this report.

6. The Big Picture / Other Issues.

There are a few issues that were not adequately addressed in the analysis and/or reporting. [S9 start] First is the question of how emissions variability is addressed in the modeling of benefits. Past studies of Smog Check effectiveness have made estimates of how this variability impacts effectiveness (see, for example, the 2000 report of the Inspection and Maintenance Review

Committee). In Chapter 9 of the Final Report, the contractor does a nice job of explaining the issue of emissions variability. This issue also should be addressed throughout the modeling report. The contractor should make clear whether the modeling approach does or does not account for emissions variability. For example, a non-zero offset of failure probabilities following a passing ASM test accounts for at least one element of vehicle emission variability. [S9 end]

[S10, S11 start] The other area of deficiency is that the study failed to substantively examine (or even discuss) two additional ways in which RSD could potentially enhance and I/M program: (1) as an on-road check of station performance, and (2) as a means to identify intermittently high-emitting vehicles. Both of these benefits would increase under an RSD program that was ongoing and from an analysis approach that considered repeat measurements over time, not just the most recent RSD measurement. It is reasonable to argue that these deficiencies are beyond the scope of the current study. But they should be mentioned and discussed in the context of the overall question of the potential benefits of RSD. [S10, S11 end]

III. Major Concerns that Could Impact Overall Results

[S12 start] My first major concern relates to the logic employed in selecting and modeling the normal I/M projection (NIM) and some of the intervention options. This issue affects many, if not all of the results and thus, potentially, the overall findings. This concern relates to the logic of comparing potential pathways that a vehicle can take following the decision point as demonstrated in Figures ES-1, ES-2 and similar figures in Chapter 4 of the Modeling report. The first part of this concern relates to what assumptions are inherent in the NIM pathway. The descriptions related to the Directing and Call-In pathway suggest that the NIM pathway is based on vehicles going to high-performing stations. Directing benefits are calculated as the difference between the Exempting pathway – which is assume to correspond to no benefits of tests at low-performing station – and the NIM pathway, which overlaps the Call-In pathway associated with sending vehicles to high-performing station. [S13 start] (The text also clearly suggests that the NIM is associated with high-performing stations). [S13 end] There are at least two major deficiencies in this logic that will lead to bias in calculating benefits (in ΔFMD and ΔFTP) for all alternative program options (Directing, Call-In, and Scrapping). In general, the NIM pathway should reflect the possibility that a vehicle could go to any station, and thus should be based on a weighted average of results for high performing and lower performing stations. These curves would be above the NIM curves currently shown in Figs ES-1, ES-2 and corresponding figures in Chapter 4. The Directing and Call-In pathways (assuming all vehicles that are called in early are directed to high performing stations) should logically be below the NIM pathway reflecting the benefit of the high performing stations. The shift in the NIM pathway will affect the calculation of benefits for all options. Regarding the approach currently used for calculating Directing benefits, the assumption that a vehicle tested at a lower performing station will on average receive no benefit from the I/M cycle (setting this equivalent to the exempted vehicle) does not appear to be supported by any data and it may not be reasonably assumed. It must be emphasized that changes to the logic of the various pathways will affect the calculation of benefits of various program options. I would expect Directing benefits to decrease and Call-In benefits to increase, while the penalty of Exempting would also decrease (as the NIM and

Exempting curves are brought closer together). [S12 end] [S14 start] A point of critical importance is that we cannot assume offhandedly that changes in the calculated benefits would be the same for program implementations that use VID only data verses those that use RSD data. [S14 end]

[S15 start] In raising this issue, I recognize that there may not be a way to address this concern in a straightforward manner. My understanding is that the data used to model the failure probabilities that form the basis of the NIM pathway include inherent biases in that some vehicle descriptions have been directed in the past to higher performing stations. Additional analysis of the primary data would be required to develop relationships between performance of high performing stations and the average station (representing a combination of high- and low-performing stations). This analysis would require additional funding. But this is a critically important element to understanding the potential benefit of Directing at the time of a regularly scheduled ASM or at an earlier time (Call-In), and it must be based on solid analysis of real data. [S15 end]

[S16 start] My second major concern relates to the assumption of parallel deterioration following a pass or fail on the previous ASM test. This may seem to be a minor point, but it impacts almost all of the analyses conducted in this study. As far as I can tell, this assumption is supported solely by Figures 2-1, 2-2, and a few paragraphs of accompanying text in the February 10 Modeling report. The specific figures shown are entirely unconvincing and to my eye paint a much different picture than what is described. To me the figures show distinct trends for initial pass and initial fail results. Both have indeterminate behavior over the first 3 months that can be inferred only by the trend which follows. For the initial pass, there appears to be a trend of failure that projects back close to zero at the time of the initial test, increases linearly over the first year, then levels off for the next year. In contrast, the trend for initial fail vehicles starts at point above a zero failure probability (for the vehicle shown approximately 15% would fail a retest at time zero) then increases steadily over the first two years at a rate that is intermediate between the first and second year rates for the initial pass vehicle. While these figures apply to only one vehicle description, a previous analysis of Smog Check program data (IMRC 2000 report) found similar trends program wide, for at least the first year. Figures 3 and 5, in Chapter 3 of that report show next test failure rates for initial fail and initial pass vehicles for 13 months following the initial test (these are also based on change of ownership tests). These results suggest the need for the Fprob models to assume different trends following an initial pass or initial fail. Ideally, these would be based on analysis of the more recent data available since the 2000 IMRC report. [S16 end]

II. Specific Comments and Concerns

Unless otherwise noted, all chapter and page references in this section pertain to the February 10 Modeling report. Since the issue of assumptions about pathways is critically important to the analysis, I have divided my first major concern into specific elements.

1A. Clarify Assumptions for Forecast of NIM pathway.

[S17 start] Based on the discussion surrounding figures ES-1 and ES-2 and the corresponding method descriptions in Chapter 4, it is not entirely clear if the curve for NIM reflects the weighted average performance of all stations, or assumes testing in a high performing station. The descriptions and visual presentation of how the other curves interact with this curve suggest the latter. For example, the ES discussion about Directing notes that the benefit of Directing results from the difference between the D/E curve – which assumes normal(poor) station performance – and the better performance of high performing stations (to which the vehicle would be directed), as indicted by the NIM pathway. Both the ES and Chapter 4 must clarify what is being assumed (high performing station or weighted average of all stations) in the NIM pathway. The specific meaning of high performing station should also be clarified (i.e. test only or inclusion of other stations as well). [S17 end]

1B. Forecast of NIM pathway should be weighted average for all stations.

[S18 start] For the analysis of the various program options to be valid, it seems that the NIM pathway should be based on an average of all stations, weighted by the likelihood of a vehicle being tested in a higher performing station. This pathway should be between the DI/EX and NIM pathways currently shown in Figures ES-1 and ES-2 (and corresponding Figs in Chapter 4). Alternately, if the current NIM reflects an average of all stations, then the DI and CS/CN curves should be below this. The setting of this new base NIM pathway will reduce the benefit of Directing, increase the benefit of Calling-In (Sticker and No-Sticker) and may also affect Scrapping. The disbenefit (penalty) of Exempting would be lower if the new NIM curve is higher (i.e. if the current NIM curve really does correspond to high performing stations). While the direction of these trends will apply to the effects calculated for both models C and D (i.e. to the use of VID only or VID + RSD data), they may not apply equally to both. Also, the expected significant reduction in the benefit of Directing for all cases and the reduced penalty for Exempting in all cases would change the overall balance of program options. [S18 end]

1C. Calculating effect of Directing.

[S19 start] As described in the ES and Chapter 4, the benefit of Directing is calculated based on comparison to a vehicle that follows the same path as the Exempt vehicle. This pathway assumes no benefit of an I/M test cycle conducted through a non-directed (i.e. non-high-performing) station. This is an unreasonable assumption that leads to an over-estimation of the benefits of Directing. As noted above, the setting of an NIM that weights all stations will reduce the overall benefits of Directing and potentially change the relative value of RSD + VID versus VID only in producing benefits related to Directing. [S19 end]

1D. Calculating effect of Exempting.

[S20 start] Calculation of the penalty for exempting should compare the exempted vehicle path to the weighted average NIM path, not the Directing vehicle path. This change will reduce the overall penalty of Exempting vehicles and may result in a change in the difference between RSD + VID versus VID only Exempting selection. [S20 end]

1E. Calculating effect of Call-In and Scrapping.

[S21 start] Call In and Scrapping pathways also should be compared to a weighted average NIM path, not the Directing vehicle path. This is expected to increase the benefit of each of these pathways and may result in a change in the difference between RSD + VID versus VID only for these program options. [S21 end]

2. Deterioration and failure rates have different trends following ASM pass or fail.

[S22 start] Refer to comments made about this concern in the overall evaluation above. [S22 end]

3. Effect of vehicle emissions variability.

[S23 start] As noted in the summary evaluation, the phenomenon of emissions variability for many vehicles with non-negligible emissions is well established. Such variability can lead to legitimate passing of vehicles with intermittent high emissions or with emissions that vary around an I/M cut point. This phenomenon is especially important to the performance of an I/M program. Some of the methods used in this analysis may implicitly address the impact of individual vehicle emissions variability. Minimally, the reports describing this work should discuss this issue explicitly and clarify the extent to which the methods used do or do not account for the phenomenon of vehicle emissions variability. [S23 end]

4. Back-casting of trends in Figures ES-1 and ES-2.

[S24 start] The trends shown in Figs ES-1 and ES-2 (and corresponding Chapter 4 figures) suggest that ΔFMD and ΔFTP are both lower for the NIM scenario on the AFD (ASM test following decision point) than they were 24 months prior to this test, i.e. following the last I/M cycle. This is counter-intuitive because it suggests that this vehicle is getting progressively cleaner over time, and specifically that vehicles of this description are cleaner following a subsequent I/M cycle than they were following the previous cycle. It is not clear to me what part of the analysis leads to this result, but it raises questions about the analysis. [S24 end]

5. Projection of emissions for Scrapping pathway –next I/M should have an effect.

[S25 start] The Scrapping pathway shown in Figs ES-1 and ES-2 (and corresponding Chapter 4 figures) indicates no effect of the next I/M cycle. I understand that some fraction of vehicles called in for a Scrapping ASM would pass and that these would represent some finite quantity of FMD and FTP emissions. But it seems that some fraction of those that passed at the decision point, would fail at the time of their next regular I/M cycle (in this case at 3 months after the Decision Point). Analogous to the CN curve, there should be a dip in the SP curve at this time. Those failing at this time could either be scrapped or would need to be repaired; either would represent an emission benefit on this pathway. [S25 end]

6. *Use of ΔFMD and ΔFTP/$.*

[S26 start] ERG used different ranking schemes for including vehicles in various program enhancements. It makes great sense that the ranking for Scrapping would need to include a metric for the cost of the vehicle. However, it seems that the Scrapping ranking could just as well have used ΔFMD/$ instead of ΔFTP/$. Likewise, once you are set up for the ΔFTP model, why not just use that metric in place of ΔFMD? Putting aside concerns about the accuracy of the ASM to FTP conversion, the metric of mass emissions is in theory better than the metric of increased failed miles driven. Two vehicles could drive the same number of miles in a condition that would lead to them both failing an I/M test, but one could have higher mass emissions per mile. In this case, you would want to rank the higher emitter higher for inclusion in your program (e.g. for Directing or Call In). [S26 end]

7. *ASM to FTP conversion equations.*

While I recognize the value of getting to a full cycle mass emissions estimate, the use of ASM to FTP conversion equations for I/M program evaluation may invoke some special concerns. (Note: I have not had the opportunity to review the Sierra updates to these equation – these concerns are based on the original equations developed by ERG.) [S27 start] My first concern is that the population of vehicles used to develop the original equations was not representative of the overall fleet of vehicles. Specifically, the ARB surveillance fleet has traditionally included a lower percentage of high emitters. While the ASM to FTP conversion equations may produce decent results for the overall fleet on which they are based, it is unclear how these translate to the California I/M fleet. [S27 end] [S28 start] It is also unclear to me how the changes that result when vehicles are repaired in an I/M program are carried through the ASM to FTP conversion. [S28 end] [S29 start] On this point, it was very interesting to see the results presented in Table 4-9. It is a credit to the contractor that they conceived, conducted and presented this analysis. To my reading, this table gives an indication of just how wildly uncertain are the estimated FTP values. Regarding the specific issue being discussed around Table 4-9, the contractor makes good use of these results using the reasonable approach of normalization. But the larger point about the uncertainty of these numbers should not be lost. It would be helpful if the contractor could address, perhaps in the appendix, the potential biases (or justify no concern about biases) associated with the use of ASM to FTP conversion factors. [S29 end] Please also refer to the comments below about the language used to describe the estimated/calculated/forecasted/modeled FTP emissions.

8. *Calculation of repair costs for various program options (Chapter 4 of Modeling report).*

[S30 start] It is counter-intuitive (and I believe incorrect) that a pathway involving fewer I/M tests over a set period of time would lead a vehicle to have higher repair costs. This logic is echoed in a paragraph on p. 4-67 that explains that "Calling-In strategies will always be associated with increased" costs relative to NIM because "the call-in ASM is an 'extra' test relative to the" NIM. Yet the analysis contained in the same section suggests that repair costs for Exempting will be higher than for NIM and costs for CS (calling-in with sticker) will be higher than CN (no sticker). In both of these cases, the pathway with fewer ASM tests leads to higher overall costs – a result that is hard to accept as reasonable. If we start today with 1000 vehicles of the same description, we can assume that some fraction would have an emission control failure leading to the potential to fail an ASM and thus the requirement of a repair over some period of time in the future (e.g. 2 or 4 years). For simplicity we can consider all 1000 vehicles being on the same I/M schedule; let's say all being on schedule for a Smog Check in 6 months. In the NIM scenario, some fraction of the vehicles would fail that initial inspection and need to be repaired. Additional vehicles would fail along the way and would need to be repaired as they are brought in for change of ownership ASMs. At 30 months from now, another group would fail the ASM administered at that time. It doesn't make sense that in the Exempting scenario, more vehicles would fail during change of ownership and the second ASM than would fail during the first ASM, change of ownership and the second ASM in the NIM scenario. The logic that more chances to fail leads to higher likelihood is sound and especially consistent with the well established phenomenon of emissions test-to-test variability. The same is true for the CN vs. CS pathways. Although this seems like a small issue, it raised a question about the validity of the specific Fprob models used and the overall analysis. (My hunch is that this issue relates back to the need for different degradation pathways following a passing vs. a failing ASM). [S30 end]

9. Terminology and language used in method descriptions in the Modeling report.

The following points are grouped because they relate to the general issue of the language used in presenting methods. These concerns can all be addressed with editing of text of the February 10 report.

9A. Description of FTP emissions should consistently indicate uncertainty.

[S31 start] The uncertainty associated with FTP emissions calculated in this study is not adequately and consistently conveyed. My understanding is that FTP emissions are calculated from model-predicted ASM emissions which are calculated based on model-predicted failure probabilities. I recognize the challenge involved with writing a readable document that is also accurate in the terminology and comprehensive in the presentation of uncertainties. And I don't have what I consider to be a completely satisfactory alternative to recommend. Perhaps best is to use the generic term "estimated" to refer to the FTP values (this term is used in some places already).

The attempt at a clear description of this on page ES-9 of the Modeling report is insufficient in my opinion. First, I don't believe that the term "probable" captures the collection of uncertainties involved in calculating FTP emissions. Nowhere in this analysis is it established that "the sum of the probable values…will be close to the sum of the actual values for those vehicles". This *may* be the case for the fleet of vehicles for which the ASM to FTP conversion equations were developed (I haven't seen the report). But has such an evaluation been done for a representative sample of the ASM fleet in California? And in the current analysis, there is the additional uncertainty associated with fast pass. My understanding based on this report is that the ASM values being used are not full tests for actual vehicles, but rather modeled values based on vehicle characteristics, whether the vehicle passed the ASM, etc. This adds an additional, important level of uncertainty. At best, it can be said that the *concept* or the *intent* is that for a group of vehicles the average or sum of the estimated FTP values would be close to the values that would be measured for the same group of vehicles if FTP tests were administered. [S31 end]

9B. Vehicle use rates.

[S32 start] The analysis astutely identifies the importance of considering mileage accrual rates for individual vehicles (since we want to preferentially repair vehicles that drive more) and correctly notes that specific mileage accrual rates are available via the VID. However, in the current analysis, annual average mileage accrual is tracked only by model year / age (and car vs. truck). There are several places throughout the text that (somewhat misleadingly) refer to mileage accrual that is specific to the vehicle. All sections that refer to this issue should be reviewed and the language should consistently note that mileage accrual is by model year and vehicle type (car vs. truck) only. Given the many uncertainties in the analysis (including centrally the failure rate to ASM emissions to FTP emissions chain) the benefit of this mileage accrual feature is oversold in my opinion. An example of a misleading section of text is in the third paragraph of page ES-9 in the Feb 10 report; the text refers to an "individual vehicle's monthly miles driven". To me, this suggests a value that is determined (e.g. directly from the VID) for the specific vehicle being studied. [S32 end]

9C. Define high performing station.

[S33 start] Does this refer to test only stations, stations that meet some other criteria, or some other metric? The term should be clearly defined in enough places, including in the glossary, to allow readers to readily access the info. [S33 end]

10. How is VID data used for Call-In program?

[S34 start] I do not recall seeing, and could not find on limited re-inspection of the Modeling report, what assumptions are made about the implementation of a Call In strategy using VID only data. With RSD, the decision point for Call-In would be following the RSD measurement. But when would a vehicle Call-In occur based on a VID-only data model? Is there a set time, e.g. at the midpoint between regular I/M tests, that the vehicle would be called in? This should be addressed clearly in the text if it is not already (I may have just missed it). [S34 end]

11. Vehicle value depreciation rates should vary by make and model.

[S35 start] The data presented in Figure 3-4 of the Modeling report appears to show that depreciation rates vary more among vehicle makes and models than they do between cars and trucks. I recognize this is a small detail, and certainly not worthy of any re-analysis at this point. But it should be considered in future analysis since there may be a correlation between ASM failure rates and depreciation (higher failure rates correlated to more depreciation). [S35 end]

12. Curve for exempting.

[S36 start] This is related to the points above but is of lesser importance. If the same deterioration slope (increase of failure rate with time) is assumed following an initial pass (IP) or fail-pass (FP) (an incorrect assumption to my understanding, as noted above), then should this slope also apply to the Exempting pathway? The data for IP and FP degradation (fail rates over time) come from change of ownership tests in the years leading up to their next regular ASM.

The observed slopes thus include repairs made in anticipation of these COO tests. If a similar effect of COO is assumed for the Exempted fleet, then the degradation should follow this path and the no future I/M pathway would have a higher degradation rate. I ask that these assumptions be reconsidered by the contractors. (I am less certain of these points than I am of those above). [S36 end]

13. Replicate RSD measurements have value; how does the model account for this value?

[S37 start] In the Final Report, there is an extensive discussion about the challenge associated with obtaining substantial fleet coverage; one of the ramifications is that many vehicles will receive multiple RSD measurements. Does the current modeling framework utilize this potentially valuable data? [S37 end]

14. Some Figures and Tables in Final Report need better captions.

[S38 start] On my first read of the Final Report, I found many of the Tables to be confusing. Some were clearer to me after reading the Modeling report. But if the Final Report is intended to stand on its own, then it is imperative that the Tables and some of the figures include much more completely descriptive captions, footnotes and discussions. [S38 end]

External Review of Study on the Potential for Using Remote Sensing Devices (RSD) to Improve California's Smog Check Program – ADDENDUM PREPARED IN CONSIDERATION OF THE COMMENTS SUBMITTED BY ARB REVIEWERS

Reviewer: Dr. Brett C. Singer
Date: 28 September 2007

Purpose of this Document

On August 6, 2007 I submitted to Allen Lyons (of the ARB) an independent peer review of the series of reports by ERG Corporation on the potential for using remote sensing devices to improve California's Smog Check program. My summary evaluation at that time was that the ERG approach to evaluate emission reductions was fundamentally valid, but that there were substantial and important concerns about several key assumptions and details of their analysis. I noted at the time that addressing these concerns would likely have a substantial impact on the calculation of emission reduction benefits attributable to RSD (and VID) and *could* change the conclusions about relative effectiveness. My review focused on the general approach and the specific details of the analysis of emission reduction benefits. I suggested in my review that most of the cost estimates seemed reasonable, but I noted that this was an area about which I had little direct expertise.

Prior to submitting my own review I did not have the time to review the series of comments that Mr. Lyons had received from others about the ERG study. I have just now done so. Prof. Donald Stedman and Dr. Doug Lawson, working as consultants to the ARB, raised a number of very important and convincing challenges to the ERG work. Their comments have caused me to reconsider my summary evaluation. Provided below is an addendum to the review that I submitted on 06 August 2007.

Addendum to Summary Assessment of ERG Study
In Consideration of Concerns Raised by Prof. Don Stedman and Dr. Doug Lawson

The concerns and criticisms raised in the Stedman & Lawson review are generally on target and appear to be very well supported with detailed references. While I do not wholly agree with their position that any assessment which is based on VID data is fundamentally invalid, I recognize and agree that ERG failed to provide suitable validation or support for their assessment methodology that is based on modeling of VID results.

I therefore retract my assessment that the ERG "approaches that constituted the basis for analyzing benefits of program changes are fundamentally valid and appropriate". [S42 start] I stand by my assessment that the models and analysis techniques developed by ERG are innovative and potentially valuable in analyzing Smog Check program components. But I agree with Stedman & Lawson that the methods require validation before results are assumed to accurately represent or predict real benefits. [S42 end] *Criticisms of the ERG benefits calculation approach* that require a substantive response include [S43 start] (1) that the ASM to FTP model includes a systematic bias for high-emitting vehicles, [S43 end] [S44 start] (2) that the log transformation approach infuses a systematic bias in the treatment of high-emitting vehicles, [S44 end] and [S45 start] (3) that inconsistencies between VID and RSD results are treated in a way that systematically biases against the capability of RSD at identifying vehicles with high on-road emissions. [S45 end] In my initial review I raised concerns relating to items (1) and (3) above; the review by Stedman & Lawson provides a much more extensive, detailed, and, frankly, damning critique on these points.

[S46 start] **The Stedman & Lawson review also calls into question the ERG analysis of RSD program costs.** Stedman & Lawson assert convincingly that ERG estimates of potential fleet coverage are greatly underestimated and RSD data collection costs are greatly overstated. The RSD coverage and cost estimates offered by Stedman & Lawson seem much more reasonable and are supported by a collection of relevant references. Taken together, revised estimates of coverage and cost should have a large impact on the total emission reduction potential and an even large impact on the evaluation of RSD program cost effectiveness. These criticisms are of central importance and appear to undermine the validity of the ERG analysis. [S46 end]

Stedman & Lawson raised two related concerns about conflict of interest that deserve scrutiny by the ARB and BAR. [S47 start] The first is the observation that the ERG report concludes that the most cost-effective method to improve Smog Check effectiveness is through use of a high emitter profiling service that ERG offers as a product. [S47 end] [S48 start] The second issue is the noted lack of input from the RSD provider ESP. While ESP also has a clear conflict of interest, it is entirely possible to carefully scrutinize and check information that this company would offer about costs, coverage and technical capabilities. It appears that ERG ignored much valuable information that could have been provided by ESP; presumably this was based on a conflict of interest concern. Yet ERG presents a report that promotes the value of their HEP approach, which in this context is a competitor with RSD as a tool to improve Smog Check effectiveness. Based on my personal knowledge of several of the ERG researchers who conducted this study, I do not believe there was any intentional attempt to manipulate the data or analysis. However, there is an important conflict of interest issues that needs to be addressed. And the wholesale exclusion of information offered by ESP inappropriately limits and appears to bias the analysis. [S48 end]

My summary assessment is that Stedman & Lawson raise a series of concerns that call into question the basic validity of the ERG analysis approach and certainly appear to invalidate the major conclusions noted in their report. The criticisms related to bias in the handling of high emitter data are particularly troublesome. Even if these challenges can be answered, there are several key concerns about specific analysis elements and assumptions (e.g. about the relative effectiveness of test only vs. standard stations) that I identified in my initial review. Finally, the ERG estimates of RSD coverage and costs appear to be inconsistent with relevant experience elsewhere. **In conclusion, I must agree with the assessment by Stedman & Lawson that the ERG study should be regarded as fundamentally flawed.**

Appendix R

Comments of Michael Rogers

A Review of:

Evaluation of Remote Sensing for Improving California's Smog Check Program

Prepared For:

California Air Resources Board
1001 I Street
P.O. Box 2815
Sacramento, CA 95812

By:

Dr. Michael Rodgers

Principal Research Scientist and Adjunct Professor
School of Civil and Environmental Engineering
Georgia Institute of Technology
&
Principal Research Scientist and Associate Director
Aerospace, Transportation and Advanced Systems Laboratory
Georgia Tech Research Institute
Atlanta, Georgia 30332

August 2007

Overview:

In late May 2007, the California Air Resources Board (CARB) requested a technical peer review of a variety of documents related to the potential use of Remote Sensing Devices (RSDs) in the California Vehicle Inspection and Maintenance (Smog Check) program. These documents included:

"Review of Literature on Remote Sensing Devices", August 26, 2004 (Literature Review)

"Estimating Benefits of Improvement Strategies (including RSD) for the California I/M Program: An Inspection and Emissions Forecasting System", February 10, 2007 (Modeling Report)

"Estimating Benefits and Costs of Improvement Strategies for the California I/M Program: Implementation Options for Using RSD", April 9, 2007 (Implementation Report).

"Evaluation of Remote Sensing for Improving California's Smog Check Program", May 4, 2007 (Final Report).

All of these reports were principally authored by the Eastern Research Group, Inc., a well-known and respected research organization with significant experience in this area. Later, CARB provided copies of the public comments on these reports although these comments are not reviewed directly. Many of these public comments address specific issues regarding these reports and these specific comments will generally not be duplicated here. Rather, this review will examine each of these reports in a broad policy-relevant sense and will attempt to answer some specific questions outlined by CARB in their review request. Certain issues that have a significant impact on the overall policy implications of these reports will be examined in more detail. The review will conclude with this reviewer's overall evaluation of these reports in the context of the application of RSD in California.

General Comments on Reports:

Literature Review

In its current form, the literature review is seriously deficient. [R1 start] Coming from an academic background, this reviewer finds the complete absence of any references to the peer-reviewed scientific literature in a "literature review" both surprising and unacceptable. [R1 end] [R2 start]The report is poorly edited and a great deal of useful information is relegated to appendices with little effort to integrate these results with the remainder of the text. [R2 end] [R3 start] Likewise, efforts to compare and contrast the findings of the various studies are largely perfunctory and not generally useful. [R3 end]

[R4 start] From both a scientific and a personal perspective, the absence of any discussion of results from either the work from the University of Denver (Don Stedman's group) on evaluation of the Denver I/M program or results from the *Continuous Atlanta Fleet*

Evaluation (CAFÉ) in Atlanta is both puzzling and troubling. In particular, the fourteen year old CAFÉ remote sensing program was designed to evaluate the Atlanta I/M program in much the same way that RSD may be used in California. Both of these omissions are surprising as experiences in these and other studies have provided the basis for much of U.S. EPA's guidance on the use of RSD for program evaluation purposes and both studies are described in the open peer-reviewed scientific literature as well as in a variety of reports, conference proceedings and graduate theses. [R4 end] [R5 start] Much of the other relevant work using remote sensing (e.g. the analyses of the Arizona RSD data and the work on fuel based inventories of pollutants based on RSD observations) are also either omitted or discussed only indirectly through other reports. [R5 end]

Modeling Report

This is the largest and most complex of the four reports and provides the most information available for review. The report, despite some deficiencies, represents a generally credible, although not entirely successful, effort to model the relationship between RSD readings and ASM test results using the ASM testing as a baseline. [R6 start] In a policy context, however, the larger question is: *"Is this what we want to model?"* [R6 end]

This approach makes two implicit assumptions. [R7 start] The first is that current ASM test results represent an accurate and minimal baseline from which to evaluate all other programs. As a consequence of this assumption, a second assumption becomes that, because of the accuracy of the baseline established by the current Smog Check program, the goal of any RSD program should be to duplicate the results of these ASM tests even though they are spatially and temporally separated from the RSD measurements. [R7 end] [R9 start] Both of these assumptions may reasonably be questioned. For example, let us imagine the development of new low-cost tailpipe test procedure that has both lower false positive and lower false negative rates than the current ASM procedure but costs the same. Due to the first assumption, this new test would be modeled by the framework as being inferior to the current system (since it has less than perfect correlation with the current approach) although this is clearly not the case. The second assumption is also questionable from a policy perspective if we acknowledge the possibilities of systematic errors in the measurement process. In this case, lack of correlation with such a measurement may be the only way in which these errors are detected and corrected. [R9 end]

From an air quality policy perspective, I/M test results are meaningful only to the extent that they accurately represent the emissions from vehicles over time and/or identify those vehicles needing repairs and ensure that these repairs are actually conducted and are effective and durable (or that the vehicles are removed from service). The extent to which this assumption is true, of course, is the underpinning of the concept of periodic scheduled testing used in current I/M programs. The validity of this assumption in actual practice has been the subject of substantial debate over for a number of years. While various experts hold different opinions, the weight of evidence seems to suggest that no single approach, including ASM, RSD or even FTP testing can claim to represent the "truth" regarding the emissions of a fleet of vehicles over time.

Rather, each approach provides different "pieces of the puzzle". FTP testing is intensive but its high cost precludes its widespread use and thus fleet estimates are based on extrapolations

from a limited number of tests. [R10 start] RSD measurements represent only the briefest snapshot (less than one second) of the tailpipe emissions from a vehicle but are conducted at such low cost that they may be replicated thousands or millions of times under actual field conditions. [R10 end] [R11 start] For its part, scheduled ASM testing represents an intermediate case, providing longer and more controlled testing than RSD at moderate cost but requires the cooperation of the vehicle owner and thus is intrinsically unable to identify compliance rates and scheduling bias. [R11 end]

[R12 start] This leads us to our second assumption, that the proper way to use RSD measurements is to predict ASM test results, since the techniques tend to measure different aspects of the vehicle emissions question. [R12 end] [R13 start] While RSD certainly should have some reasonable correlation to ASM test results since both measure vehicle emissions, variability in emissions due to time, repairs, differences in fuels, etc. as well as the test itself serve to limit the extent of this correlation. This assumption is enforced in the modeling (e.g. in the logistic regressions) by converting a continuous variable (RSD reading) into a dichotomy (i.e. ASM "pass" versus ASM "fail") rather than a direct comparison of emissions measurements. This is not done without reason. The use of "fast pass" algorithms in the ASM testing used in most programs (and I believe that is the case in California) limits full tests to only failing vehicles or a small number of vehicles selected to have full tests for program evaluation purposes. Thus the number of seconds of ASM testing applied to "passing" cars is normally significantly less than those of "failing" vehicles. This has implications for the use of ASM test results directly since the significance of start up transients and vehicle conditioning effects versus hot-stabilized emissions will vary with test length as will the resulting uncertainties in the mean value of the emissions rates. These and other factors significantly impact quantitative analysis of these data[1]. Thus it is quite reasonable that the authors chose to use the "pass/fail" criteria for their modeling even though this approach results in a substantial loss of information content and ability to judge relative emissions contributions from various intervention strategies[2]. The extent to which this paradigm is applied is substantial, including the use of this threshold ("failed miles driven") in one the primary evaluation criteria. [R13 end] [R14 start] Of course, actual vehicles do not recognize such criteria and real emissions rates (g/mile) are the actual contributor to air quality degradation. [R14 end]

[R16 start] These comments are not meant to demean or dismiss these modeling efforts, that as mentioned earlier, are credible within the framework of their assumptions but rather to question whether these models are, in fact, appropriate for modeling the impact of the use of RSD in the California program. For example, RSD criteria such as presence among the top 1% of emitters or recorded emissions more than three times the average for all vehicles of a certain model year and vehicle type could realistically be considered for implementation of RSD testing without reference to any ASM testing criteria. Such approaches might be shown to provide superior results in terms of total emissions reductions to the approach considered in this study. [R16 end]

1 This difficulty has lead EPA's Office of Transportation and Air Quality (MOVES Modeling workgroup, June 2007) to limit the data used in the development of the MOVES emissions model to "full tests" used for program evaluation purposes.

2 [R15 start] For example, RSD strategies targeting only the highest emitters would receive no more credit for finding these vehicles than for finding vehicles only slightly exceeding the ASM test standard. [R15 end]

Implementation Report

The implementation report is well written and presented but suffers from a relatively poor development of the scenarios that will be considered and some logical errors associated with the estimates. [R20 start] Examples of some of the logical errors are discussed below: [R20 end]

[R17 start] In Table 3.1b the relationship between "Unique, Valid, DMV-Matched Vehicles (UVD Vehicles)" and "In Range VSP, UVD Vehicles" is obtained by multiplying their number by 40% assuming that this represents the fraction of all measurements within the designated range (other programs generally report this fraction as between 60% and 80% but this does not change the argument). However, the previous two columns would indicate that the average vehicle is measured 3.3 times (31.6 million/9.5 million) that should result in a ratio of "In Range VSP, UVD" to "UVD" vehicles in the range of 70% to 90% depending on the assumed distribution of tests rather than the 40% reported here. [R17 end]

[R18 start] A similar argument may be made the ratio of Raw to Valid readings as many of these same vehicles have invalid readings due to out of range VSP. [R18 end]

[R19 start] In Table 3.1a it is assumed that the 10% of vehicles from the "rest of the state" are measured by RSD at the same rate as local vehicles. This would imply that these vehicles have similar VMT on facilities used for RSD measurements (i.e. ramps and single travel lanes) as local vehicles. This seems highly unlikely. [R19 end]

These issues result in a substantial underestimate of the fraction of I/M eligible vehicles that can be measured and to overestimate the cost per measurement per unique vehicle.

[R20 start] In terms of scenario errors, the report has only a very limited discussion on the ways that various types of RSD programs could be implemented and what the effect of such a strategy might be. For example, the various scenarios evaluated are all based on differing intensities of essentially random measurements. There is no discussion of other approaches such as targeted measurements of high emitters, non-random (e.g. announced) clean screening sites and/or public information (e.g. smart sign) programs, dedicated remote sensing facilities to ensure proper VSP ranges for measurements, etc.. In evaluating a potentially large program, a number of these options should be considered and analyzed since all have the potential for reducing costs for certain types of objectives (e.g. clean screening or high emitter identification) generally at the expense of other uses of RSD data. [R20 end]

In summary, while the report is potentially useful to the state, the errors in the cost estimates and fleet coverage assigned to the RSD programs and [R21 start] the failure to fully evaluate the programs on a "level playing field" (e.g. not assuming that the VID is "free" in evaluating vehicle scrappage programs while assigning full costs of the RSD program to this activity) [R21 end] make the current evaluation unreliable for evaluating various policy options. It would appear that errors and similar problems identified in some of the public comments could be corrected fairly quickly and more realistic estimates provided.

Final Report

The final report is the best written of the reports and clearly the authors have devoted substantial effort to refining the text. The major criticism of this report is that it summarizes results from the other reports that are often flawed, in some cases seriously so.

[R22 start] The most serious direct failing of this report is its lack of evaluation on the behavioral aspects of emissions control programs. Many investigators (including this reviewer) have argued that the most significant benefit of enforcement actions are in the behavioral impacts on those not directly involved in the enforcement action. For example, we have almost all witnessed the slowing of traffic passing by a vehicle stopped for traffic violation. Whereas such individual events have little impact on overall speeding rates within a jurisdiction, research has shown that consistent enforcement of speed laws can increase the perceived threat of being caught to a much greater extent than actual citations would indicate and have a substantial impact on the incidence of high speed driving in an area. It is thus reasonable to assume that the threat of unscheduled testing (e.g. in a high emitter identification program) may introduce similar beneficial behavioral changes in the general population that go far beyond the relatively small number of enforcement actions actually completed. For example, a RSD enforcement program might reduce the average time required for vehicle owner to repair a vehicle after a MIL illumination or could lower the likelihood of vehicle tampering in the period between vehicle inspections. However, both of these impacts would not be evaluated as benefits of the program in the current analysis, unless an enforcement action took place. This behavioral impact is, of course, not limited to RSD-based programs. For example, studies in Atlanta (which has an annual vehicle inspection program) have shown that a significant number of vehicles receive initial emissions tests immediately following emissions-related repairs as opposed to repairs occurring after a failing test. Many of these repairs would probably not have occurred in the absence of scheduled emissions test and thus the emissions benefits of scheduled ASM testing are also likely to be underestimated. In either case, such impacts should be explicitly considered in the implementation of any vehicle inspection program to maximize the air quality benefits of public investment of both time and money. [R22 end]

CARB Review Questions

RSD Fleet Coverage

[R23 start] The fleet coverage requirement is largely a function of level-of-effort desired and clearly at some point raising this fractional coverage becomes cost prohibitive. It is clear from the Missouri data that 50% coverage can be routinely achieved and it is unlikely that coverage of more than 80% could be achieved on any large air basin due to a variety of factors. So estimates in the range of 50-70% of registered vehicles are credible with the ERG estimate on the lower end of this scale. [R23 end] {R24 start] It is important to note, however, that this 50-70% of vehicles represents a much higher fraction (probably in excess of 90%) of light-duty VMT as the primary reason for not being sampled by an aggressive RSD is low vehicle mileage accrual rates. [R24 end]

Calculated Smog Check Failure Probabilities

[R25 start] It is not clear that Smog Check failure is reasonable criteria for evaluation of the success of an intervention unless the Smog Check test is conducted immediately. As discussed above, human behavior in the light of an RSD program will probably influence emissions to a much greater extent than the relatively limited number of vehicles that might be identified by a high emitter identification program. Repairs conducted because of the presence of the program or repairs conducted due to notification that an unscheduled test will be required but before its execution cannot be evaluated by ASM test results. [R25 end]

Methodology Used to Estimate Emissions Benefits of Special Interventions

[R26 start] While an interesting metric, the "failed miles driven" is far less relevant to air quality than "total excess emissions" which takes both the emissions and activity factors into account and is the standard method used by U.S. EPA for estimating the effectiveness of control programs. This is especially relevant in the case of Tier II vehicles where a vehicle can be considered driving while "failed" even though absolute emissions are still far lower than for older vehicles. [R26 end] [R27 start] It is also important to note that a high emitter identification program may choose to require tests of observed high emitters among recent model year vehicles not subject to regular testing. The benefits from such a program cannot be evaluated using the VID information. [R27 end]

Methodology to Estimate Cost Effectiveness

[R28 start] As noted by a number of the public comments the estimates of costs for the remote sensing program are probably grossly overestimated. Although the "per beam block" cost figures are reasonable, the estimated efficiency at obtaining "complete" records appears to be far too low. The ERG Implementation Report estimates that only about 3% of RSD beam block will produce valid, DMV-matched records of unique; I/M covered vehicles with in-range VSP readings in California. While it is difficult to extrapolate between programs, this estimate is far from the experience of current programs. For example, in 2006 the CAFÉ program in Atlanta conducted measurements at 37 measurement sites producing approximately 360,000 beam blocks. These "beam blocks" yielded approximately 253,000 valid in-Range VSP, DMV-matched readings for all pollutants representing about 217,000 unique vehicles. In the Atlanta measurements approximately 85% of these unique vehicles are subject to the I/M program yielding an overall efficiency of about 50% for meeting the strictest criteria presented in the report.

Even assuming a factor of three difference in efficiency due to California specific factors and program scale (approximately 50% coverage as opposed to 6-10% coverage for Atlanta) the cost estimates are likely to be at least a factor of five too high (i.e. actual efficiency of ~15% rather than 3%). The experiences in the Missouri & Virginia programs have been similar.

This overestimate of costs has serious implications for all cost effectiveness calculations presented in the report and will tend to invalidate many of the conclusions. [R28 end] [R29 start] In addition, the ERG reports tend to assume that the VID is a "sunk cost" and that the cost of maintaining and updating the VID need not be considered in evaluating relative program costs. While less significant than the overestimate of RSD program costs, the exclusion of these costs tends to underestimate the cost of ASM-type programs further distorting the relative cost-benefits of various program alternatives. [R29 end]

Using RSD to Analyze Smog Check Benefits and Fleet Emissions

Despite their differing feeling regarding the use of RSD for "clean screening" or "high emitter identification", most experts will agree that fleet-level evaluations make the most compelling case for the use of RSD. The ERG reports generally agree with this conclusion as does this reviewer. [R30 start] However, careful planning and execution of the program is required and this activity may need to be independent of the use of RSD for other purposes due to potentially conflicting measurement requirements. [R30 end]

The "Big Picture" and Overall Conclusions

For most of the last fifteen years, there has been some form of conflict between supporters of the application of RSD in, or in lieu of, scheduled tailpipe I/M testing programs and supporters of these traditional programs. Many supporters on both sides were quite strident in their positions, at least initially, and viewed any presentation from the "other side" with great suspicion. RSD were characterized as inaccurate, subject to temporal drift, and all manner of interferences. For their part, I/M testing programs were characterized as ineffective, rife with corruption, and of only minimal value in improving air quality. As time went on, improvements in RSD and RSD methodologies and continued improvements in I/M testing programs including enhanced testing (ASM and I/M 240) and OBD II, as well as significant volume of research, tended to bring the camps closer together with each group developing a greater appreciation of advantages and limitations of each approach. RSD measurements were used to demonstrate the effectiveness of I/M testing and clean fuels programs and RSD were incorporated as a regular part of several I/M programs without replacing regular testing. [R31 start] I believe that it is in this context that CARB and the State of California commissioned this study to evaluate *in a neutral context* the ways in which RSD could be used to enhance the Smog Check program without seeking to displace the major elements of the current program.

Unfortunately, as a reviewer I must conclude that in their current forms these reports fall well short of this goal. [R31 end] [R32 start] Serious errors in the calculations of costs and fleet coverage overstate RSD costs by as much as an order of magnitude. [R32 end] [R33 start] An analysis framework that assumes that the major benefit of RSD is early detection of high emitters that would be detected by the Smog Check program anyway clearly undervalues the potential emissions benefits from a real RSD program as well as a range of other benefits. [R33 end] [R34 start] Likewise, in some cases costs that might reasonably be attributed to ASM (e.g. the VID) are treated as sunk costs even though additional costs will be accrued in the future while RSD alternatives are priced at full costs. [R34 end] The combined effect of these approaches is to so distort the cost-benefit analysis as to render it virtually useless for policy analysis in its current form. This is not to assert that RSD is a cost effective way to identify high emitters, candidates for scrappage, or to “clean screen” vehicles in California. It may or may not be. In fact, for many applications it is probably not, especially for scrappage applications. But I do not believe that that determination can be made without major changes to the existing analysis.