Peer Review of
Estimates of the Fraction of the Fleet with High Evaporative Emissions based on the Ken Caryl Station (Denver, Colorado) Field Study

Prepared by:

H. Christopher Frey, Ph.D.
Independent Consultant
Raleigh, NC

Prepared for:

SRA International, Inc.

December 30, 2011

This is a peer review of a report drafted by Eastern Research Group, Inc. of Austin, TX for the U.S. Environmental Protection Agency. The review was commissioned by SRA International, Inc. on behalf of U.S. EPA.

Synopsis

The subject report is a summary of measurements of evaporative emissions made on samples of vehicles in the Denver, Colorado area using a portable "SHED" method. The term SHED is undefined in the ERG report but refers to Sealed Housing for Evaporative Determination (SHED), and is an existing and established method for measuring evaporative emissions from vehicles. A portable SHED (PSHED) was used at a particular inspection and maintenance (IM) station in the Denver area, known as the Ken Caryl IM Station, to measure a sample of vehicles. Vehicles were selected for PSHED measurements based on a screening measurement made with a Remote Sensing Device (RSD). Remote sensing measurements were used to estimate a quantity "EI23." Depending on the value of the EI23 quantity, each vehicle passing through the screening RSD site was assigned to an EI23 "Bin." A prior estimate was made of the probability that a vehicle in each bin would be a 'high' emitter of evaporative emissions. Based on the assignment to an EI23 bin, and the prior estimated probability of being a high emitter, vehicles were invited at random for additional measurements. If the driver accepted the invitation, the vehicle underwent additional remote sensing measurements and also underwent a PSHED measurement.

In prior work, measurements were made on a different vehicle fleet to assess the concordance between SHED and PSHED measurements. The two methods were found to produce similar results on average, with a small bias in the PSHED measurement compared to the SHED measurement. However, there appears to be substantial random deviation of individual PSHED measurements compared to SHED measurements of the same vehicle under approximately similar conditions. A key assumption in the analysis of Ken Caryl IM Station data is that the PSHED measurement is a suitable substitute for the more widely accepted SHED measurement.

Based on the 'stratified' sample of vehicles that underwent additional measurements at the Ken Caryl IM Station, inferences were made regarding the observed fraction of vehicles in each EI23 Bin that had

'high' evaporative emissions rates according to PSHED measurements. These data were used to estimate the overall fraction of high evaporative emitters for all vehicles that entered the Ken Caryl IM Station during the study period. The report concludes with some comments about uncertainties in the estimates and regarding a possible future Monte Carlo study to more thoroughly quantify such uncertainties.

Report Writing

The review of this report was significantly hampered by the very poor quality of the report organization and writing.

A key question when writing any report is: Who is the intended audience for this report? The intended audience should include all stakeholders of the MOVES model, since this effort appears to be aimed at providing a technical basis for quantification of the fraction of the on-road fleet that has high evaporative emissions rates. However, as written, the report is aimed at fellow technicians who are familiar with the undefined shop jargon used by the authors. This report contains repeated sloppy use of jargon that may be meaningful to the report authors, but that make the report difficult to read by anyone else. Table 1 is a list of terms that are introduced in the text without definition, without adequate definition, or that should be introduced, defined, and used in the text. The list of terms in Table 1 should be used to construct a glossary for this report. When a term is first used in the text, it should be defined in the text.

The report needs substantial copy editing by a competent technical writer. For example, the report contains frequent use of the first person, which is inappropriate in formal technical writing. In numerous places, statements of belief are made (e.g., "we believe"). The reader does not care what the authors 'believe.' The reader cares about what is known and what is not known, and reasonable interpretations based on evidence. The report contains numerous metaphors, which are inappropriate for formal technical writing. For example, several times the authors describe what an instrument 'sees.' Aside from these problems there are numerous instances of unclear yet repetitive statements. If a student had handed me this draft report, I would have read a few pages and then handed it back as unacceptable.

As an example of poor writing, consider the last paragraph on page 4-2.

> What are 'pretesting data'? 'All of that pretesting data was' could simply be "These data were." "receive RSDs" - this doesn't make sense. How does a vehicle receive a remote sensing device? The intended meaning seems to be "were measured using remote sensing." Having read the appendix, I cannot figure out the basis for the statement "Analysis of the EI23 index… " "to allow the EI23 to be less dependent on an exhaust emissions, we developed EI23 Bin" is awkward – should be "To reduce dependence on exhaust emissions, EI23 Bins were developed." Do not use first person. And so on. Aside from the poor wording, the key technical concepts are unclear. What are the dependences and how have they been inferred? It is frustrating to the reader to be told to go elsewhere for definition of EI23 and EI23 Bin but to be provided with details based on knowing what these concepts are, such as "EI23 Bin has integer values of 1 through 7…" These concepts and terms should be defined, developed, explained, etc., in a

Table 1. Terms Introduced in Draft Report Without Definition or Explanation: these terms should be defined/explained when first mentioned. A glossary of these terms with definitions should also be created.

Terms that Need to be Defined	Comment
Aging enhanced evaporative emissions vehicles	? given lack of definition of 'aging,' and 'enhanced,' the meaning of this is unclear to the readers.
Aging enhanced vehicles	Undefined. Explain this.
Approximate algorithm	No idea what this means. Needs to be explained.
As-received condition	Should explain what this means.
Beam block	This is shop jargon. The intended meaning seems to be exhaust plume measurement. Needs to be defined/explained when first used.
Bench purged	Presumably, this implies that the canister was removed from the vehicle and purged (how) on a lab bench. Needs more explanation for clarity.
Bias (systematic error, inaccuracy)	See comments
Bin de-stratification	De-stratification with respect to what? How?
De-stratify (and de-stratifications)	This term is used without definition. Not clear what this is.
Detection limit	Mentioned on p 4-31 but not defined.
EI23	Mentioned numerous times without any explanation
EI23 Bins	Define when first mentioned. Introduce in a new methodology chapter prior to using this term in results chapters
Electronic HC sniffer	Is this relevant to the content of the report? If not, delete. If so, then explain.
ESP	? Seems to be the name of a company. ESP, Inc.?
Evaporative emissions canister	Is this a canister that produces evaporative emissions? Need to explain to the reader what this is. A corresponding conceptual diagram of the source of evaporative emissions and methods for prevention and control of evaporative emissions would help in explaining what this (and other relevant vehicle systems or components) is.
g/Qhr	Is not defined until page 4-4, although it is used in earlier parts of the report.
Gross liquid "leakers"	Is there a quantitative definition of this, or at least a working definition? Explain.
HC	As good practice any abbreviation should be defined when first used.
high evaps	This is shop jargon. A formal technical report should have thoughtfully developed and carefully defined terminology.
High running loss emissions	What constitutes "high"? by what criterion or criteria?

High-PSHED, and "high-PSHED fraction"	This term is shop jargon. The intended meaning appears to be "vehicles with high evaporative emissions as measured using the Portable SHED
Hot 505	This is undefined. Presumably, this is a hot stabilized dyno test cycle. If so, then give the graph of speed versus time and provide some explanation.
Hot soak	define
IM	Yes, most readers will know what this is, but as good practice any abbreviation should be defined when first used.
Implanted leak	Undefined. Explain this. Give an example.
Index/PSHED	This term is unclear
Intrusive pressure test	What is this?
Ken Caryl	Introduced asif the name of a person, this should be consistently termed "Ken Caryl IM Station" or something similarly descriptive (e.g., Caryl Station).
leakers	Is there a quantitative definition of this, or at least a working definition? Explain.
Low evap	More shop jargon. A formal technical report should have thoughtfully developed and carefully defined terminology.
Low running loss emissions	What constitutes "low"? by what criterion or criteria?
Method A	Define. Introduce in methods chapter.
Method B	Define. Introduce in methods chapter.
Modified California Method	Define. If not relevant to this report, delete.
MOVES	MOVES is mentioned but never introduced or explained.
Near-zero vehicle	Undefined. Explain this
Noise, noisy	Used on page 4-31 without definition
OBD code to flag	I know what the authors are trying to say, but many readers will have no idea. First, explain OBD and what is an OBD code. Explain what is meant by 'flag'.
OBD evaporative codes	What are these? Needs to be explained
Odometer Resolution	What is the meaning of the codes given in Table 4-1?
ORVR	?
Precision (imprecision)	See comments
Pre-enhanced vehicles	Undefined. Explain this.
Pretesting	Page 4-2
PSHED	PSHED is defined on page 1-1 as "portable SHED", but "SHED" is not defined.
RSD	The term RSD is used on page 1-1 without definition.

RSD Method B	This method should be introduced and explained in a methodology section of the report.
Running loss emissions	Define/explain
Seal Barometric Pressure	Table 4-2: this term is undefined. There needs to be a footnote to explain what this is.
Seal Temperature (F)	Table 4-2: this term is undefined. There needs to be a footnote to explain what this is.
Selection RSD	Mentioned on page 4-3. An "RSD" is a measurement device, but the term "RSD" is used inappropriately to refer to a measurement of a specific vehicle. The intended meaning of "Selection RSD" is "screening remote sensing measurement." The screening measurement is used to determine whether the vehicle will be recruited for addition RSD measurements and PSHED measurements.
SHED	Amazingly, SHED is not defined the first time it is mentioned, on page 1-1.
Slow vapor leaks	What is a "slow" leak? Does this refer to a low emissions leak? Of vapor? Of evaporating liquid? Needs to be defined and explained.
Standard de-stratification techniques	? undefined.
Standard I/M inspection	Explain. Or, if not relevant, delete.
Stratified sample	With respect to what? This term needs to be explained when first used.
Stratified set	Explain in new methods chapter.
Uncertainty	Should be defined – see comments
VDF	Table 4-2: this term is undefined. There needs to be a footnote to explain what this is.
VECI Engine Family	Table 4-1. needs to be defined in a footnote.
VECI Evap Family	Table 4-1. needs to be defined in a footnote.

methods chapter prior to producing results based on these. The paragraph introduces, perhaps for the first time, the term "running loss emissions," without definition. If EI23 Bin is central to the methods and interpretation, it is simply unacceptable to push it to an appendix and to give such short and uninformative treatment to it in the main body of the report.

Aside from the poor writing, the organization of this report should be reconsidered. Methods and results appear to be mixed together. A good technical report will have a chapter devoted to methods, organized in a manner consistent with the order in which the methods are used later in the report. Furthermore, this report tends to have too much of 'here's what we did' without first introducing the purpose, key concepts, or basis/foundation.

A technical report should have the following elements:

- Introduction

    - states the challenge, problem, issue being addressed,
    - establishes the need for new work
    - **clearly** states the objectives of the work (note: objectives are not a list of tasks – they relate to the purpose of the work)
- Background: Survey of relevant prior work, if needed. Also, a brief review of the types of evaporative emissions and factors to which they are sensitive is needed. For example, evaporative emissions are sensitive to ambient temperature.
- Methods
    - For each major component of the analysis, state the following:
        - Overall purpose
        - Basic concept
        - Empirical or theoretical basis established in prior work (with citations)
    - Provide sufficient information regarding the methods so that someone else could reproduce the work – include definitions of key terms, variables, equations, algorithms, and so on
    - Examples of content for this chapter (illustrative)
        - Schematic of the vehicle path through the various RSDs and PSHED
        - Methods A and B for estimating plume concentrations from remote sensing measurements
        - EI23 definition and definition of EI23 bins
        - Approach to 'stratification'
- Results
    - Results could be organized into more than one chapter if the subject matter is too much for one chapter
    - Results should include a clearly summary of all input data and assumptions
    - Results obtained should be from application of methods described in the methods chapter.
    - Results should be appropriately interpreted
- Conclusions
    - What are the key findings that are related to the objectives stated in Chapter 1?
    - What are the key conclusions that are related to the objectives stated in Chapter 1?
    - What are the key recommendations that are related to the objectives stated in Chapter 1?

Use of Proprietary Methods

Over the years, EPA has been criticized for making public policy and developing modeling tools to support public policy that are based on proprietary data and methods. The use of proprietary methods precludes a full understanding and review of the underlying science. A case in point are the "Method A" and "Method B" exhaust plume analysis methods associated with the ESP remote sensing instrumentation. Since the distinction between Method A and Method B appears to be an important technical consideration in this study, the lack of disclosure of what these methods are is unacceptable.

Fundamental Questions

There are some fundamental questions related to this work that should be part of the objectives and that should be addressed in the technical results and conclusions:

1. Is PSHED a good surrogate for SHED?
2. Can an RSD, if appropriately interpreted, be a good surrogate for a PSHED measurement?

The first question presumes that SHED is the reference method to which all other methods should be compared. What, however, is really measured in a SHED measurement? There are many evaporative processes. Some, such as refueling, are not addressed by SHED. Which processes are addressed?

In what ways are PSHED measurements similar to those of SHED measurements, and in what ways do they differ? Is PSHED effectively just as good as SHED?

What kinds of evaporative processes can be measured using RSD? There is an unstated hypothesis in this report that RSD measurements can provide information on evaporative emissions in a manner comparable to that of PSHED, if only the RSD measurement is appropriate interpreted. What is the basis for this hypothesis? What evaporative processes affect the quantity of HC that is detected by remote sensing? If there was no error in the measurement, would strong concordance be expected between RSD and PSHED? If so, why? A clearer statement of hypothesis and the theoretical underpinning for it would be helpful when interpreting results.

Specific Technical Comments

Page 1-1. The first sentence refers to 'further developing' something that has not yet been defined in this report. Please, hire a technical editor and have them go through this report very carefully. The first line is poorly written, and the report that follows is also very poorly written.

The purpose of the report is to estimate, not develop, fractions of various levels of high evaporative emissions. However, no where is any justification or rationale given as to why this report is focusing on the Denver fleet. Since Denver is at high altitude, and barometric pressure is a factor in evaporation, it is not clear that data from Denver would be representative of other parts of the U.S.

A number of terms are mention on this page without definition, including SHED, Ken Caryl, RSD. This is a bad way to start a report.

Page 1-2 "the real investigation in this study happens in…" this kind of colloquial writing has no place in a formal technical report by what is supposed to be one of the top environmental engineering consulting firms in the country to the Federal agency charged with quantifying and regulating air quality. This report needs to be taken more seriously by the authors.

Background Chapter: this chapter is plagued with undefined jargon, lack of clarity of concepts, and is poorly organized. It is very qualitative and vague and provides little to no insight on the topics being addressed. Examples of content missing from this report include a brief review of the types of evaporative emissions, factors to which such emissions are sensitive, the SHED measurement approach, how PSHED works, what is remote sensing, and how can remote sensing be used to infer information

about evaporative emissions. What does the RSD actually measure that is representative of evaporative emissions, and is this similar to what is measured in PSHED? Why is there an expectation that there should be agreement between evaporative emissions inferred from RSD measurements versus those inferred from PSHEDS? Are they measuring the same processes under similar conditions? How might they differ?

The background chapter should be followed by a new chapter 3 that provides an overview of the methods used in this report, including a schematic of the Ken Caryl IM station, the specific instruments deployed, the analysis methods used, etc. Material that is now in Appendix A and B should be rewritten into the methods chapter.

The current Chapter 3 should be rewritten as "Assessment of Concordance Between Portable and Fixed Location Evaporative Emissions Measurements." This chapter needs technical editing. The basic information is useful and interesting. The technical analysis should include quantification of the statistical significance of each parameter in the regression equation, the standard error of the estimate, the distribution of the residuals, a normality check for the residuals, the coefficient of determination, and other basic information that would commonly be reported as diagnostic goodness-of-fit indicators when developing a regression model. To what extent are results such as in Figures 3-4 and 3-5 actually providing an indication of repeatability of the test – are the conditions really the same in each test? If the repeatability is really this poor, what are the implications for selecting a threshold for what constitutes a 'high evap' vehicle? It is more common to report 95% probability ranges, not 68% probability ranges.

Chapter 4:

A schematic of the Ken Caryl station is needed to illustrate what is meant by the "driveway RSD unit" and "Measurement RSDs"

What is the purpose of "stratification."? Why is achieving stratification a goal in itself? E.g., page 4-3, "to achieve stratification, a higher fraction of vehicles…" The reader can eventually figure this out, but why can't the authors communicate this more clearly? The purpose seems to be to evaluate a screening procedure for identifying vehicles with high evaporative emissions rates, but what about goals for false positives or false negatives?

Is it literally the case that six RSDs were used? i.e., six remote sensing devices at six locations? Or were the two highway "RSDs" based on repeated passes by the same RSD? The authors need to stop using the term "RSD" to refer to a measurement. RSD = Remote Sensing Device and refers to an instrument. A measurement made using an RSD could be described as a remote sensing measurement. What is an RSD beam block? This is shop jargon (I know what it means, but most readers won't).

"These two RSDs were measured on the same RSD instrument as the Selection RSDs." This sentence is extremely sloppy, using the term "RSD" where the concept of a 'remote sensing measurement' should be used instead.

How does a vehicle "receive" an "RSD"? I have done measurements with an RSD, and I have never seen a vehicle receive an RSD.

What is the 'standard I/M inspection' – for those of us not from Denver, please explain what this is. Also, explain the "Modified California Method" – both of these should be documented in the new methods chapter that needs to be written. Who does the olfactory examination? What is an 'electronic HC sniffer"? Is this relevant to the report? If not, then delete mention of these.

Page 4-4: Method A was used on ESP 4000 and 4600 instruments, and Method B was used on ESP3000 series instruments. Yet, results for both Methods A and B are reported in Table 4-2. Were two RSD instruments used at each RSD site? Or were both Methods A and B applied to the same data measured from just one RSD instrument at each site? At the end of the paragraph is it mentioned that 'code' was 'added' to the 4000 and 4600 series instruments – it would have helped if this was mentioned up front, and if there was a prior section that more clearly disclosed the study design in terms of what instruments were deployed at what locations and what the vehicle path was through each RSD site. It would help if this text were reorganized so that there was an intro paragraph, one paragraph on Method A, one paragraph on Method B, and then a paragraph that compares Methods A and B. Are the CO, NO, and CO2 results shown in Table 4-2 based on Method B? The distinction between Methods A and B with respect to how they deal with exhaust versus evaporative concentrations of HC is not clear. To merely state that "ESP believes" that one method is responsive to exhaust and another is not is quite tenuous.

Page 4-4 (bottom): regression toward the mean…. This is stated as if it is an underlying principle in a rather didactic manner, but the actual concept is poorly explained here. A measurement is biased if it is systematically high or systematically low. If the error is randomly distributed with a mean of zero, then the measurement is subject to random error, not bias. The random error can lead to false positives or false negatives if used in the context of a binary decision (e.g., vehicle is a high emitter). This context is not clearly articulated. False positives or false negatives are not necessarily a result of bias, but rather a result of imprecision (random error). The discussion here of bias is thus without sufficient context and therefore is unclear.

What role does ambient temperature have in contributing to variability in estimated evaporative emissions based on RSD measurements? Since the "Temperature" in Table 4-2 (ambient temperature at the time of each RSD measurement?) differs from the PSHED "Seal Temperature", what role might this have in confounding the results?

Table 4-2: what is the meaning of negative values for HC Method A (ppmC 3) and how are these interpreted? Table 4-2 values of CO_2 percent appear to be what one would expect in the tailpipe, but this cannot be what was actually measured in the exhaust plume. How is the air-to-fuel ratio inferred, or is it assumed to be stoichiometric? Some discussion is needed. The text barely alludes to this. More detail is needed in a methods chapter.

Table 4-2: terms PSHED and RSD in caption should be spelled out. All nomenclature in column headers need to be properly defined – e.g., use footnotes. Is RSD temperature the ambient temperature at the date and time of the measurement?

The quantity in Figure 4-1 labeled as "RSD EI23" needs to be clearly defined. Is this based on any numbers given in Table 4-2? Which specific column of Table 4-2 is "RSD EI23"? Which specific column of Table 4-2 is "PSHED Mass (g/Qhr)"? Presumably, "Measured PSHED HC at 15 Minute Soak (grams)" in Table 4-2 is the same as "PSHED Mass (g/Qhr)". However, use consistent terminology in both places to avoid ambiguity. The EI23 values need to be added to Table 4-2.

Figure 4-1 needs better formatting. Should use a much larger font size for the numbers on the axes, and consider using scientific notation rather than decimals if showing a log scale. In the caption, spell out PSHED. What is "RSD EI23"?

Table 4-3 is hardly a table and is not formatted well. Add a row for percentages of total to help in the interpretation. Please change the terminology – e.g., 'Measurement RSDs" (should be Remote Sensing Measurements).

Table 4-4 the term "high PSHEDs" is unacceptable. The intended meaning appears to be "high PSHED measurement" "High-PSHED Definition" should be "High PSHED Measurement threshold" or criterion.

What is 'de-stratifications'?

Page 4-22: what role does ambient temperature have in the estimation of EI23? The RSD measurements are made at ambient temperature. Evaporative emissions are proportional to ambient temperature (something that needs to be introduced and discussed in a background or methodology section of this report). Is the EI23 metric less responsive to evaporative emissions at lower ambient temperature? Speed is not the only factor that affects inference of evaporative emissions.

"these Selection RSDs can be used to de-stratify the stratified set and provide an estimate of the high-PSHED fraction of the fleet…" given the lack of clear definition of these terms, and the sloppy use of terminology, this sentence is unclear.

'is not an unbiased' – why not say 'is a potentially biased'… positive statements are always more clear than negative statements.

Page 4-23: "For the RSD to be useful…" should be 'for the remote sensing measurement to be useful…" however, why is model year important? Earlier, a note was made that model year was not part of the EI23 binning method.

If there are multiple EI23 bin values available for some vehicles, these data should be analyzed separately to determine the robustness with which a vehicle is assigned to an EI23 bin. Ambiguity in assignment to an EI23 bin would be a significant factor to consider in evaluating the usefulness of this method.

Table 4-5. the term "Selection RSD" needs to be changed… e.g., "screening remote sensing measurement"? But the table is actually of EI23 bins and model year groups, not screening remote sensing measurements. Thus, the caption is not consistent with the content of the table.

A table prior to Table 4-5 would be more useful… i.e. distribution of vehicles by model year groups and EI23 bins for the selected (stratified?) vehicles.

Page 4-24 "we will get started…" might be okay for a presentation but this is not how a technical report should be worded. Aside from this, the first paragraph in Section 4.4 is unclear and is hampered by repeated use of terms that are not well-defined. Methods for stratification and de-stratification should be in a prior methods chapter.

Try reading aloud the first sentence of the last paragraph on page 4-24. It needs to be rewritten. Aside from being a run-on sentence, it is awkward, contains repetitive points and yet is not very clear.

Page 4-25: The Appendix B should be part of a methods chapter given earlier in this report (could be Chapter 3). N_h is defined in Appendix B but is given a lower case symbol (n_h). To avoid confusion and ambiguity, use consistent mathematical nomenclature. "Fraction of elevated PSHEDs" is given the symbol p_h, which is defined in Appendix B as the "probability"… this is inconsistent. Either it is a frequency or it is a probability- choose one and use the concept consistently. The standard error of fraction of elevated PHEDs is given in Table 4-6 based on a definition involving s_h and N_h, but this definition is not given in Appendix B (it should be). The terms sample and population in the Appendix B need some careful re-thinking or at least more clear definition. Here, the term 'population' is implied to describe the total sample of 5,830 vehicles (which is actually a sample from a larger fleet). That is okay, but at least be clear as to the meaning of the term 'population' as used in Appendix B. W_h is the fraction of the population of vehicles that fall into each EI23 bin. It is not clear as to the definition of "n" in Appendix B – is this the total number of vehicles in the 'population'? (i.e. n=5830?). L=7 (could be stated clearly). The term σ_h is not clearly defined in appendix B in terms of other variables. Is this the standard error of the fraction of elevated measurements in each strata? Appendix B does not actually show how one estimates the estimated fraction of the population that is above the threshold. How was the value 0.127 estimated? This appears to the product $p_h W_h$ summed over all h. Based on the numbers given in Table 4-6, over 75% of the estimated 'elevated PSHEDs' (a sloppy term) are from Bins 1-4, which account for over 96% of the 'population.'

Page 4-27. The last sentence of the first paragraph is unclear. Rewrite. Create a flow diagram or show an algorithm to make this more clear.

Table 4-8. It is not very clear as to what variable is implied by "High-PSHED Fraction…" is this based on p_h and W_h defined in some different way compared to Table 4-6?

"It is important to understand that" should be deleted. "It… that" statements are passive and contain no information. The assumption of the EI23 bins is that they are bins of EI23 values. Since no assumption is made regarding model year, it is not really correct to imply that if there is a dependency on a model year that somehow the use of EI23 is inherently inappropriate. It could be that the fraction of vehicles with high PSHEDs measurements is correlated with EI23 and with model year, but that does not imply that EI23 would not be a useful indicator. Whether EI23 is a useful indicator can be determined with or without consideration of model year. In fact, if EI23 has a trend with respect to model year that is consistent with the trend with respect to PSHED measurements, then there might be increased confidence in the utility of EI23 as an indicator.

Section 4.5: the discussion here suffers from a conceptual problem related to not clearly defining what is meant by "uncertainty." The term uncertainty is used inappropriately as if it refers only to imprecision, and the notion of bias is discussed as if it distinct from "uncertainty." Uncertainty refers to lack of knowledge regarding the true value of a quantity, and includes both random and systematic sources of error. Random error is imprecision. Systematic error is bias and also known as lack of accuracy. Thus, bias is a component of uncertainty, not distinct from it.

Uncertainties associated with small sample size are typically quantified based on random sampling error. The discussion of the role of 'chance alone' is inappropriate as written. Perhaps the intended statement is that if a different random sample of vehicles had been selected, the number of vehicles

with PSHED measurements greater than 2 g/Qhr might have been different from the 2 that were observed in the available sample. Because the fraction of vehicles with PSHED measurements greater than 2 g/Qhr is based on a sample, there is 'sampling error'in the estimate. If the sample is assumed to be random, then the error of the estimate can be estimated based on sampling distributions of the statistics (a statistic is a quantity estimated from a sample). The errors shown in Table 4-11 are of unclear basis. For example, the 'size of error for 'high PSHED Definition' of 2 is given as 0.025. There should be more detail on how this number was estimated, based on the data given in Table 4.6.

PSHED measurement error should be more clearly discussed. The text refers to 'two parts' but really only one 'measurement error' is actually addressed. Measurement error typically refers to the imprecision and bias of the measurement method itself. Propane retention and recovery tests are an incomplete indicator of the imprecision and bias of the PSHED method, because actual evaporative emissions are not pure propane. Variability in hot soak emissions is a measurement error only in the context of attempting to assess the repeatability of measurements of the same vehicle under the same conditions. However, it is not clear that such an experiment has actually been done. If there are underlying differences in the state or condition of the vehicle, then the variability in the measurements is not because of the measurement method itself but because of the state of the vehicle being measured. The concept of repeatability of the measurement should be discussed in a separate paragraph or subsection. If the repeatability is only -50% to +200%, then there is significant question as to the usefulness of any kind of PSHED test when compared to a 'brightline' threshold that is a point value.

The discussion of detection limit and how it was inferred is difficult to follow. First, it would help to define what is meant by detection limit. It is not clear how a detection limit can be inferred by making a measurement on a vehicle or any sample for which it is not known as to whether the HC concentration is actually zero. Why not use a 'zero' calibration gas that contains 0 ppm of HC? A baseline before a vehicle enters the PSHED does not guarantee that actual concentration was 0 ppm of HC. However, it does provide a background level. However, the text does not discuss what is background or the role of background in making measurements.

Pae 4-32 : the analysis of duplicate EI23 measurements is quite important, and the text refers to Appendix A. Appendix A is very poorly written and very unclear. It is not apparent that there are any data regarding the duplicate EI23 values in the main body of this report or in the appendix. The data and findings from these data should be disclosed.

The rational for the bias in the EI23 values and the implication that it would 'tend to elevate the high-PSHED fraction' needs to be more clearly articulated.

Page 4-33: the apparent confusion regarding detection limit and background level is evident in the second paragraph on this page. One does not subtract a detection limit from a measured value to impute an unbiased estimate. This would be done only for a background level. However, if the background is negligible compared to the measurement, this will have little effect on the results.

"jumps around" – this kind of informal writing needs to be expunged from this report.

The discussion of a possible Monte Carlo simulation is so vague that it hardly merits being in this report. Unless the authors can clearly define terms and propose a meaningful algorithm, the recommendation

for future Monte Carlo simulation could be stated briefly, with further development left to those competent to conduct such an analysis.

Chapter 5:

The lead paragraph here is probably the most coherent statement of the objective of this report. Such a statement is needed in the introduction.

The second paragraph is not useful because it is based on evidence not provided in earlier parts of the report.

The purpose of Chapter 5 is unclear. Is this meant to be a conclusions chapter? A summary chapter? A results chapter?

The third paragraph is awkward and overly didactic. One can make the point, for example, that the use of EI23 as an indicator of evaporative emissions was explored in this work, and state the findings, conclusions, and recommendations accruing from this work. Subsequently, a recommendation can be made that the existing data could be analyzed using other indicators for the purpose of evaluating whether other indicators might be better than EI23. Whether 'any evap index' can be used depends on what variables are critical to an 'evap index' and whether they were all measured during the study at the Ken Caryl IM station. Since the report lacks even a basic overview of factors that lead to evaporative emissions, it is not clear as to whether all useful factors have been quantified to support development of 'any' evap index.

The paragraph at the bottom of page 5-1 is sufficiently cryptic as to be useless to anyone but those involved in the data collection or project management effort. It is not very clear as to what point is being made here.

Page 5-2 "to measure the RSDs" – this makes no sense. RSDs are devices that make measurement. Why would one make a measurement on the RSD itself?

The intent of the paragraph on "RSDs of the Denver fleet" is unclear. Perhaps this is a recommendation to calculate EI23 for a wider set of vehicles and use the Ken Caryl IM station data for fraction of high emitters to estimate a fraction of high emitters for the larger fleet. If that is the case, the intent is unclear.

Last paragraph on page 5-2 – seems to be introducing a lot of new information but in an unclear manner such that the point(s) here are unclear.

What is the main contribution of this report? What are the key limitations? What additional work is needed? If the purpose is to estimate the fraction of vehicles with evaporative emissions exceeding a threshold, the method described in this report using EI23 Bins and a 'stratification' approach may be reasonable; however, the uncertainty in the estimates made using this method are unknown. Such uncertainties should be estimated as the next step. Without quantification of uncertainty, the utility of this approach is unclear.

Some key issues that should be addressed in the conclusions:

- Is PSHED a useful surrogate for SHED?
- Can RSD measurements, if appropriately interpreted, provide an indicator of evaporative emissions?
- Is EI23 a useful indicator?
- Are the trends in the results for high evaporative emissions fractions in the vehicle fleet consistent with model year? What results developed here provide some confidence that EI23 is operationally useful?
- What are limitations of EI23? What other indicators should be explored?
- What uncertainties have been quantified? What uncertainties have not yet been quantified?
- Need for further evaluation of uncertainties prior to making a decision on acceptance of this approach?
- Application of this or other approaches to fleets that are more representative of the U.S. fleet.
- Others?