I/M Programs: A Look at Where We've Been and Where We're Going

Presentation to the National Academy of Sciences Panel on Vehicle Inspection and Maintenance Programs

Joel Schwartz

February 15, 2000

Overview of Presentation

- What we know about I/M and Vehicle Emissions
  - Focus on careful, real-world studies
- I/M Policy and Regulation Divorced from Real World Knowledge
  - Case studies of flawed research and analysis
- How the air pollution planning and regulatory systems foster these problems
- Where the NRC could help take us

I/M and Vehicle Emissions Results

- Emission reduction effectiveness
- Underlying causes of observed results
  - Human behavior in I/M

Characterizing Good Studies

- Broad view of problem
  - People and vehicles are dynamic
  - Many variables, including program structure, motorist responses
- Careful attention to design and methodology
  - Data and methods appropriate to answer real-world questions
- Review, cite and compare with other work on similar problems

Measuring Emissions Reductions

- On-Road Studies
- Data
  - ambient measurements
  - roadside I/M tests
  - remote sensing
  - tampering rates
- States: Arizona, Colorado, Illinois, California, Minnesota, Atlanta

Overall Results

- Small effect of pre-enhanced I/M programs
  – Atlanta may be exception
- Larger effect for enhanced, but lower than official predictions
- Significant deterioration of benefits within an I/M cycle

Underlying Factors -- Human Behavior

- Avoidance
  - Preparing for the test
  - Multiple tests
  - Change registration
  - Don’t register
  - Test fraud

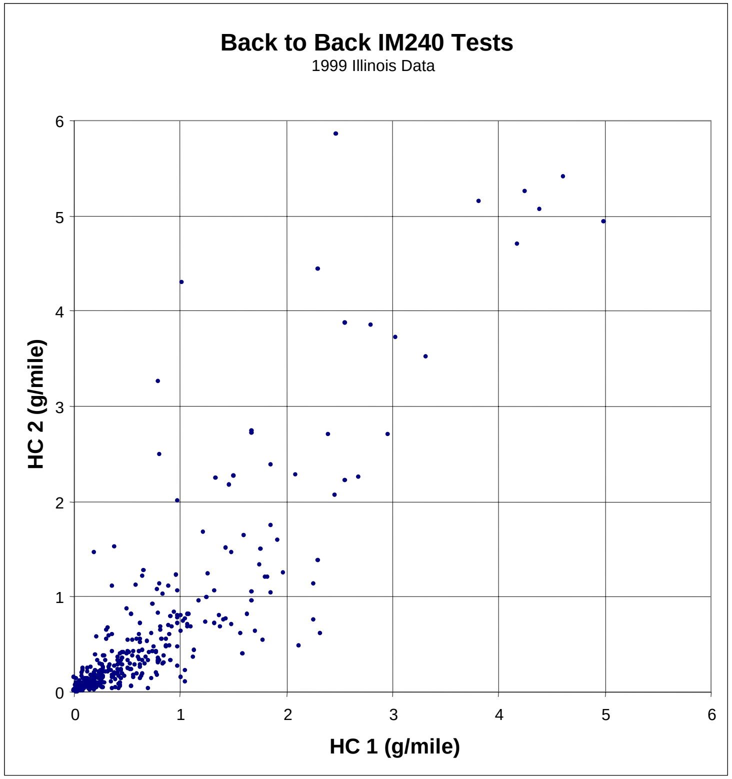
Many Failing Vehicles Not Repaired

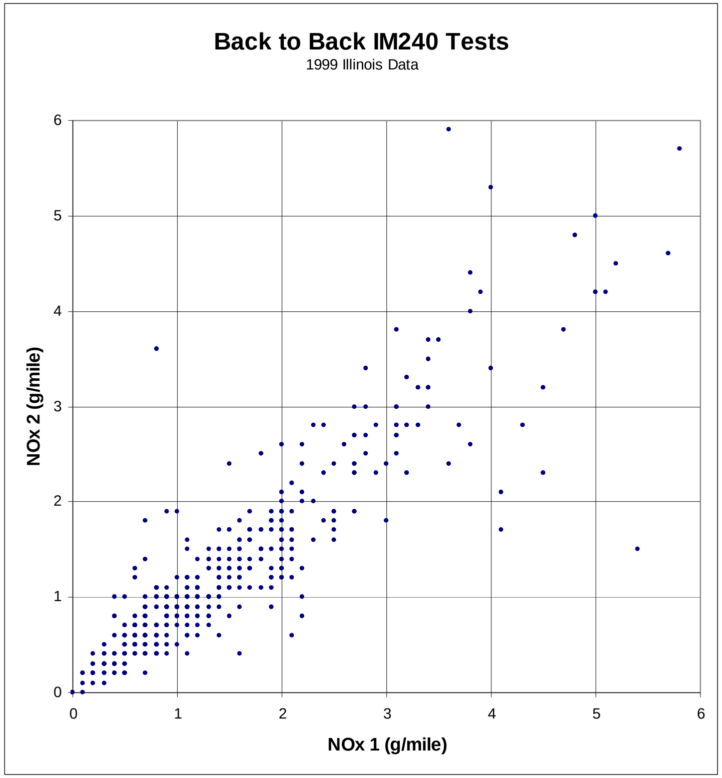
- Many motorists fail and never pass, but still on the road
  - Arizona: 30% (half still on road after 12 months)
  - Denver: 20%
  - California: 10%

Fail Then Pass w/o Repair?

- Some failing vehicles pass within short period of time
  - 10% pass within 2 hours in Phoenix
    - Half of these report $0 repair expenditures
  - 12% pass on same day in Denver
    - More than half of these pass within 2 hours
- Might indicate motorists hoping to take advantage of variability

Back-to-Back Tests Suggest This Could Be A Good Strategy





- HC and NOx $r^2 = 0.86$ (EPA 1990 Indiana data gave $r^2 = 0.66$ when a few days passed between tests)
- Using USEPA phase-in cut points and considering second test as "true" emissions
  - 20% of failing cars are false fails
  - 26% of all "should-fail" cars actually pass

Registration Changes

- Denver
  - High emitters re-registered outside Denver but still drive there
- Ohio
  - More than 10% drop in registration renewals after enhanced program begins
    - Corresponding increase in non-enhanced areas

External Factors

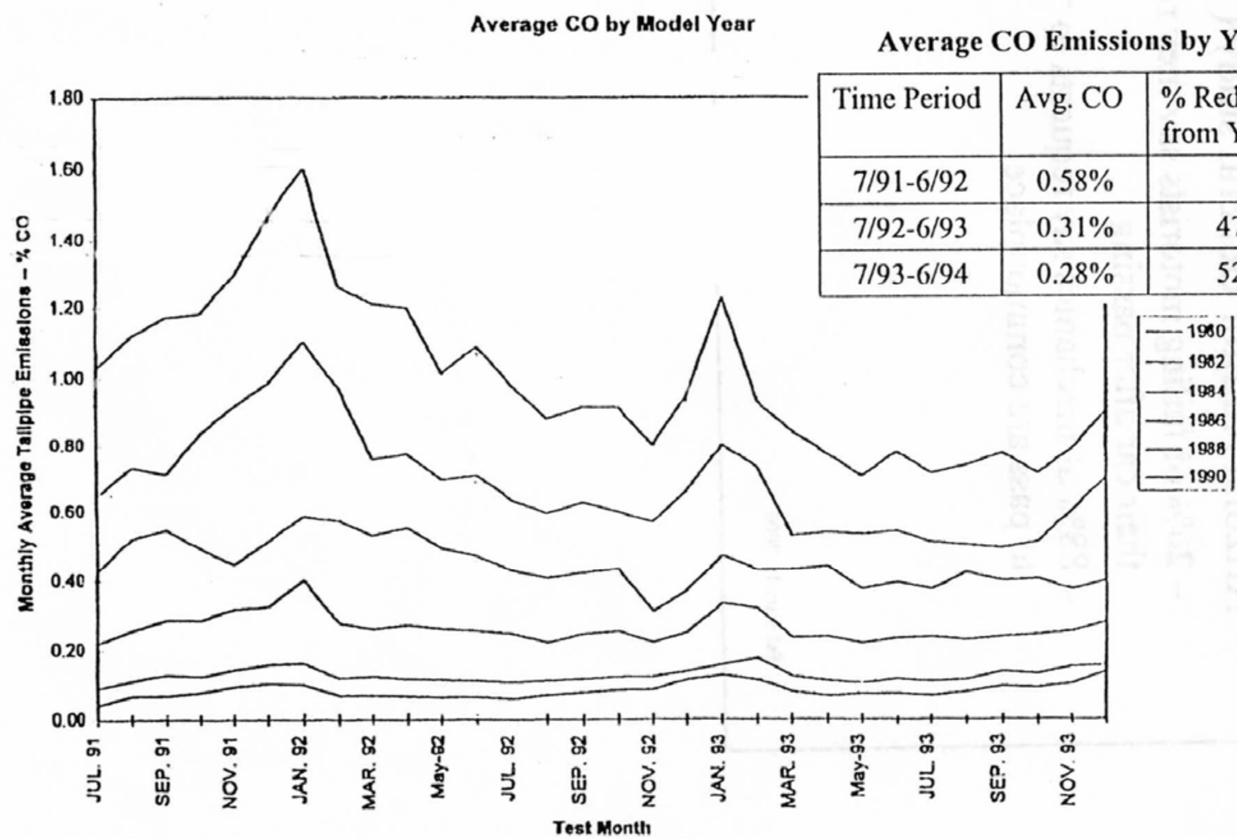
- Increase of 500,000 in registration non-renewals when California required proof of insurance for registration.

Preparing for the Test

- Arizona Auditor General study (1989)
  - 26% of failing motorists say they readjusted their car after passing
  - 88% of mechanics say requests to “set-up” cars to pass are commonplace

Evidence from Minnesota

- CO emissions drop 47% at test lane from first to second year of program, but no measurable benefit on road
  - Motorists preparing for the test?



Average CO Emissions by Year

Time Period	Avg. CO	% Reduction from Year 1
7/91-6/92	0.58%	
7/92-6/93	0.31%	47%
7/93-6/94	0.28%	52%

Skimping on the Test

- Reporter took pickup truck to 29 Long Beach, CA area Smog Check stations (Long Beach Press Telegram, 1993)
  - “Most simply announced that for a small fee - usually between $25 and $60 - they would get the truck to pass”
  - Collected 13 Smog certificates even though car continued to fail for same reason at each station

Fraud

- Los Angeles DA shuts down 24 “certificate mills” in 1992 that accounted for more than 90,000 fraudulent certificates
- After three month undercover investigation, Arizona AG arrests 13 staffers at a centralized test facility (half the facility’s staff) in 1999 for soliciting bribes to pass high emitters

Disconnect Between Real-World Results and Actual I/M Policy and Programs

- Problems in framing the issues
- Problems in the collection, use and reporting of information

Types of Problems Common in I/M Policy and Official I/M Studies

- Overarching Issue: How do we think about the problem?
  - Linear, static, separable, predictable vs. dynamic, subtle, with many simultaneously interacting factors
- Specific problems
  - Inappropriate analytical boundaries
    - Missing variables
    - Ignoring confounding variables
    - Inappropriate time frame
  - Apples and oranges comparisons
  - Lack of context
  - Omitting dynamic effects
  - Misreporting or selective reporting of data or results
  - Unsound data analysis methods

Case Study: 50% Discount

- EPA asserts decentralized programs are half as effective as centralized
- Based on:
  - Audits of individual I/M programs
    - Covert "set-to-fail" tests; Gas analyzer calibration; etc.
  - Vehicle tampering surveys
  - MOBILE Model

USEPA on What Makes I/M Effective

- Proper testing is the primary factor:
  - "For the enhanced I/M rulemaking, EPA used data from over 10,000 covert audits to assess the effectiveness of I/M programs: These results, along with the tampering survey data form the basis for EPA's 50% effectiveness discount for test-and-repair programs."
  - "…EPA found in audits of I/M programs, that emission testing was done objectively in test-only programs…On the other hand, the data shows that inspectors in test-and-repair programs routinely attempted to get failing cars to pass the initial test…These data led EPA to reduce the emission test credits by 50% in MOBILE5a for test-and-repair programs."
  - USEPA (1993), "Quantitative assessments of Test-Only and Test-and-Repair I/M Programs," EPA-AA-EPSD-I/M-93-1

Problems with Data and Methods

- Results misreported or selectively reported
- Audit protocols varied from state to state
- Key variables omitted from analysis
  - on-road emissions
  - behavioral factors
- Emissions reductions not measured

Actual Improper Testing Rates

	Improper Test Rate		
State	**EPA Report**	**Actual Audit Finding**	**Comments**
Maryland	0%	40%	Total of ten covert tests performed
Arizona	11%	11%	Total of nine covert tests performed, all with same car
New Jersey	50%	66%	Total of six covert tests performed
New York	47%	34%	Decentralized program

- Maryland Audit Report: "The EPA auditors observed many inconsistencies in the way tests are conducted."
- Only 3 of 13 audits of centralized programs were covert

Variable Protocols

- Arizona Centralized Audit: The catalytic converter was "removed and replaced with a section of rusty straight pipe."
- Georgia Decentralized Audit: "The van had been set up with subtle deception in mind. The catalyst had been removed and replaced with a very small pre-converter."
  - All 5 stations improperly passed vehicle
  - On another Georgia audit with different vehicle with missing air pump, all five stations properly failed the vehicle
  - Based on these 10 audits, EPA reported 50% improper test rate for Georgia

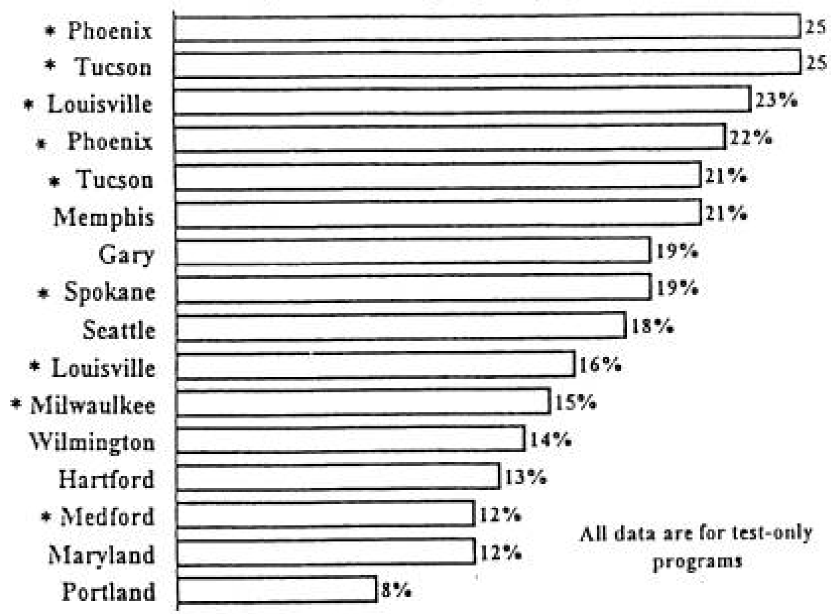
Audits Can't Tell Us About I/M Effectiveness

- Reporting and design problems
- No systematic emissions measurement
- No real-world data on program outcomes
- No analytical linkage between audits and discount
- Assumes proper testing = effective program

Tampering Rates Selectively Reported

USEPA's Actual Test-Only Tampering Results

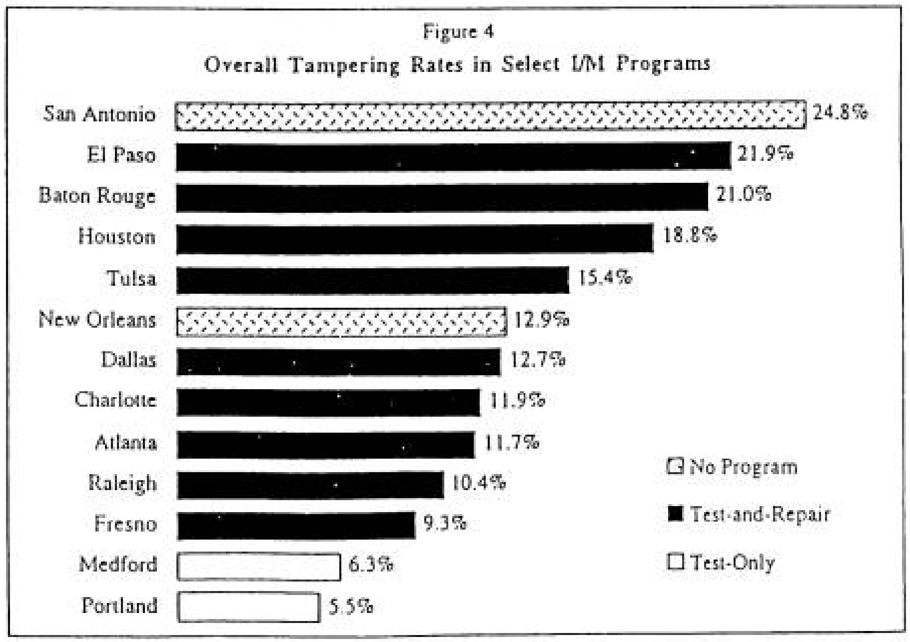
(Source: USEPA tampering survey reports: 1985 - 1990)



* = survey conducted at test lane

But USEPA Presents Data Only for "Good" Test-Only Programs

Source: USEPA, "I/M Briefing Book." February, 1995. Originally published in USEPA, 1993, "Quantitative Assessments of Test-Only and Test-and-Repair I/M Programs."



Tampering Data Show Little I/M Effect

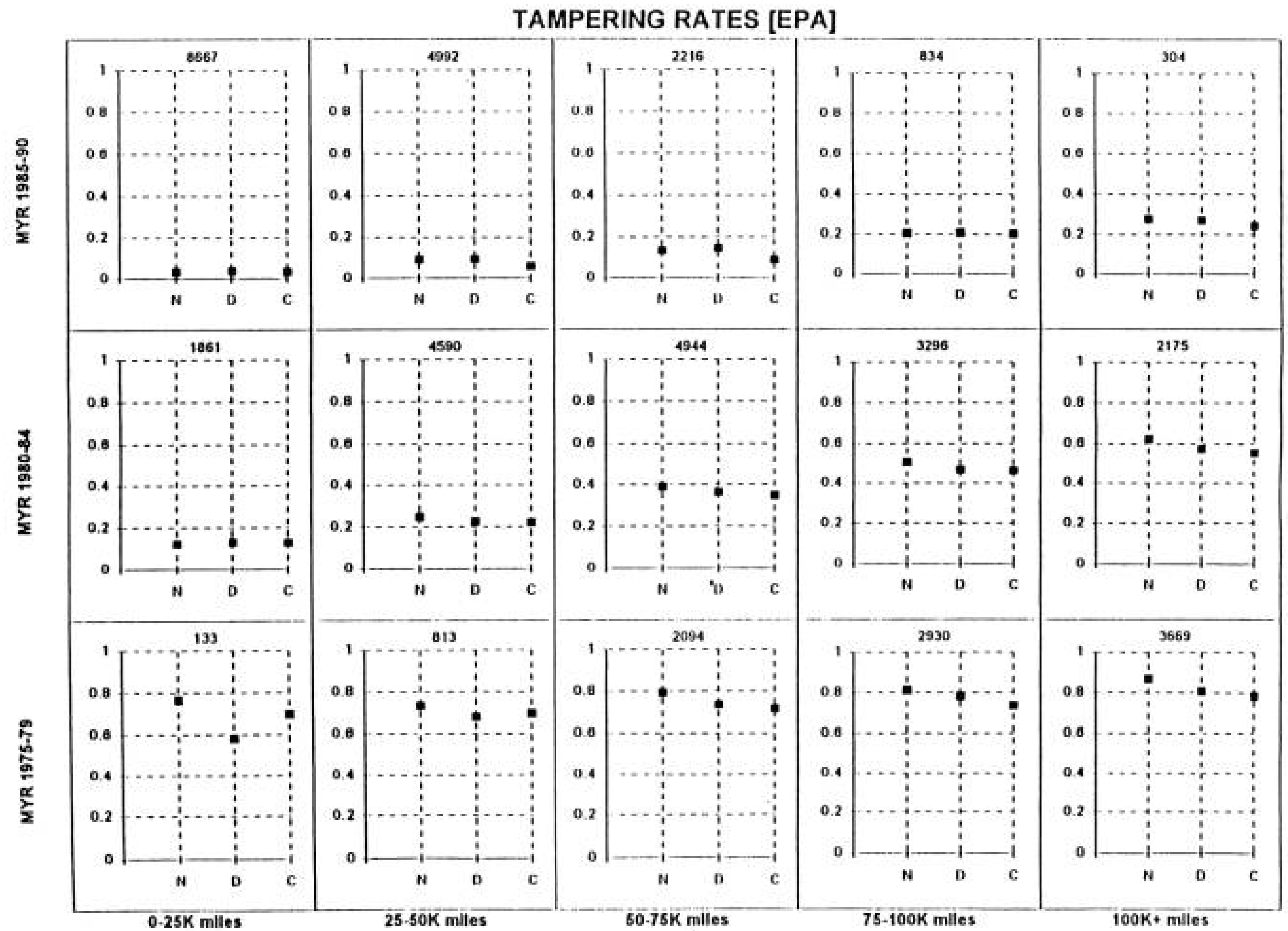


Figure 7-9. Cross-classification diagram using EPA Motor Vehicle Tampering Survey data for the 1985-1990 surveys. Tampering rates for 44,000 vehicles are shown by I/M program type as a function of year of manufacture and odometer readings (N = non-I/M region; D = decentralized I/M program; C = centralized I/M program). The number of vehicles in each model year and mileage group is shown in the figure.

MOBILE Model “Proves” Discount

- EPA asserts MOBILE output shows that decentralized is half as effective.
- “Output” resulted from statement in model that said (in FORTRAN code): “If the program is decentralized, multiply the emission reduction credit by 0.5”
- So 50% discount is not an *output* of the model, but an *input* to the model

Summary of 50% Discount Issue

- Inappropriate framing of problem
  - "accurate and proper testing = effective program"
- Apples and oranges
  - Audits not relevant to the problem
- Inappropriate analytical boundaries
  - Many relevant variables not considered
- Selective reporting of data
  - Tampering rates
- Selective citation of other work
  - Contrary results not considered

Case Study: Official I/M Evaluations by EPA, States, and Their Consultants

- Idealized studies
  - Key variables, such as human behavior excluded
- Divorced from real-world effects
- Methodologies not tied to on-road emissions
- Often don’t cite or compare with other work

EPA's Official I/M Evaluation Method

- "Sierra Research Method," designed by Sierra for EPA
  - Compare average post-I/M fleet emissions in state X to average post-I/M fleet emissions in the Arizona "benchmark" program
    - Using only "in-program" emissions data biases results
    - Assumes *only* I/M program causes differences between regions
    - Never actually measures emissions reductions
- EPA ignored its own peer reviewers' critiques of the method

El Monte Pilot Project (CA,1994)

- Purpose: compare testing and repair using IM240 and ASM
  - Cars recruited by ARB with promise of free repairs
  - Cars tested on ASM, IM240, FTP
  - ARB hired mechanics to repair cars at ARB repair bays
- Required by EPA to see if ASM could get same reductions as IM240

Apples and Oranges Comparisons

- CA Pilot characterized as a simulation of an enhanced, centralized program:
  - "Theoretically, the results obtained during the California Pilot Project would be expected to represent a centralized I/M program using ASM or IM240 testing." (Sierra Research, 1995)
    - Sierra acknowledged that real benefits would be "somewhat" lower because private garages would not be as effective. But no discussion of omission of human variables that make this far from a simulation of a real program.
      - Sierra appears not to see human factors as important: "The available data do not support a conclusion that motorist cheating is a significant factor that undermines the potential for I/M programs based on periodic, scheduled inspections." (Sierra Research, letter to JAWMA, 1994)

More Apples and Oranges in CA Pilot

- EPA and ARB see test type as key factor in program "credit"
- Radian developed emission reduction credits using EMFAC emissions model
  – Used observed high emitter ID rates and repair effectiveness for different tests
- Once again, program effectiveness "determined" without looking at real-world dynamic factors such as human behavior

ARB Enhanced Eval. Method

- Use random roadside ASM to get baseline fleet emissions
- Calculate emissions reductions achieved by applying current ASM cut points
- Discount by ratio of in-program to on-road failure rate
  - Separate calculation for test-only and test-and-repair cars

Method Not Tied to Real-World

- Start with real on-road emissions as baseline, but all subsequent steps are divorced from real-world effectiveness
- But could use roadside data to directly measure I/M effect
  - Data collected in middle of first enhanced cycle so look for difference between tested and untested vehicles

Many Studies Use Only "In-Program" Data

- Colorado Audit (1998)
- EPA Arizona Evaluation (1997)
- Wisconsin (1993)
- British Columbia (1996)
  - Don't see pre-repairs, avoidance, deterioration

Case Study: Emissions Models

- Concerns With Emissions Models
  - Unrepresentative input data
  - Unsound data analysis techniques
  - Structural limitations: *Ad hoc* assumptions rather than actual "modeling"
  - Real-world measurements contradict model output

Recruitment Bias

- Vehicles different on day of test
- Low response rates in mail solicitations
  - Non-respondents likely have higher emissions
- Don’t see unregistered vehicles
  - I/M, insurance requirements, fees, etc.

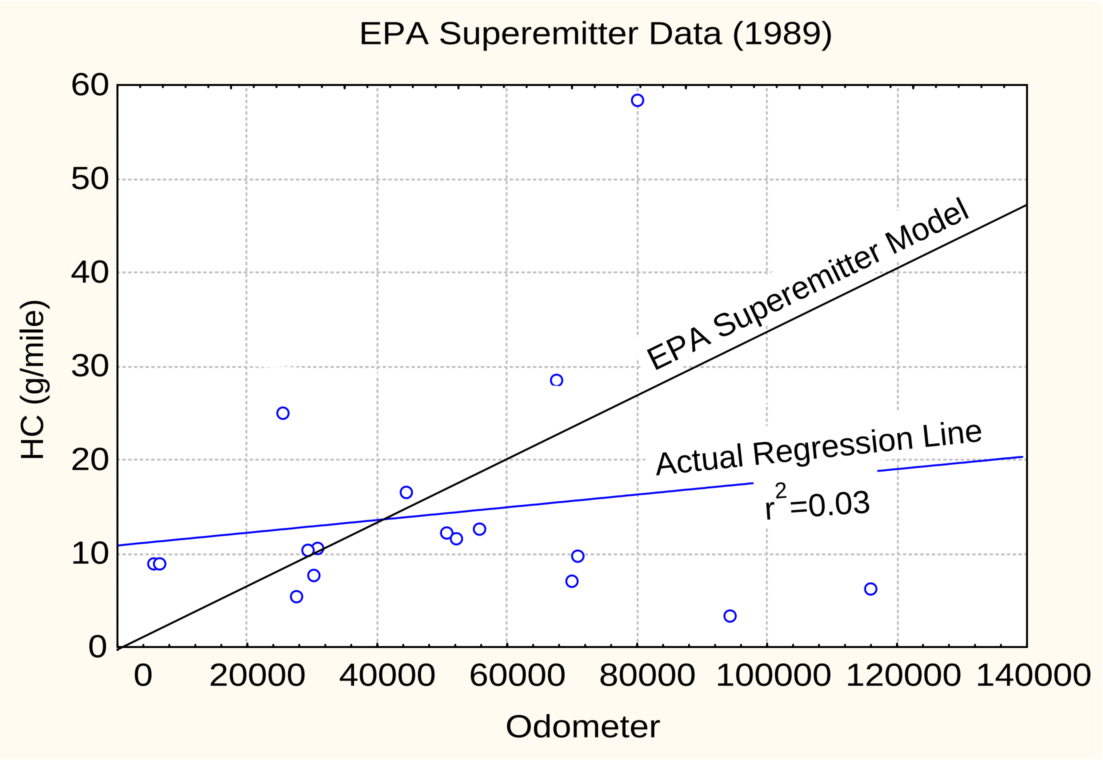
Small Samples

- MOBILE5 databases
  - IM240 ID Rates: 274 vehicles
  - Repair Effectiveness: 266 vehicles
- Both datasets divided into technology/emitter sub-groups, some with very few vehicles

Artificial Studies: Missing Variables

- Repair effectiveness based on results with EPA-paid contractors, rather than real-world mechanics and real-world motorist-mechanic interaction
- 46% of repaired vehicles did not meet standard
  – Post-repair emission values artificially adjusted downward to meet EPA assumptions about what “would” happen in a real program

Unsound Data Analysis



- EPA super-emitter model
  - 17 car data set
  - r^2 of zero, but EPA drew arbitrary line through origin and centroid

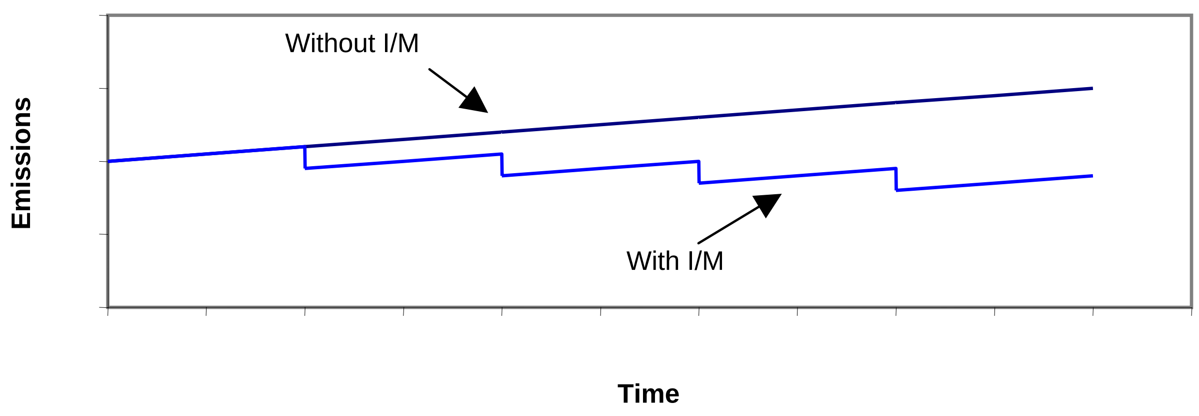
Human Behavior Not Modeled

- All behavioral effects subsumed in *ad hoc* factors added after emission calculations
- $BER_{I/M} = B_o(1 – (CRED\ (1\text{-}w)\ f(c)\ d))$
  - d = discount; f(c) = compliance rate
- Static factors can't encompass dynamic responses to wide range of variables
- Fudge factors subvert any presumed accuracy of basic emission rate calculations

Unrealistic assumptions

- **Only one chance to fail a high emitter**
  - **Overemphasizes test ID rate**
- **Assumes reductions persist forever, but look at real world:**
  - **40% of cars that fail then pass in Arizona fail again next cycle**
  - **On-road RSD in AZ shows faster deterioration for fail-then-pass vehicles**

Figure 1: EPA "Sawtooth" Model of I/M Effect on Emissions



“Model” is a Misnomer

- No mathematical representation of the interactions among the social and physical processes that affect emissions
  - Instead, *ad hoc* combination of emissions data and hardwired input assumptions strung together with arithmetic
- Not capable of evaluating policies
  - Many real-world factors could change without affecting output of MOBILE

Model Validation

- On-road studies: models underpredict urban fleet HC emissions by factors of 2 to 4
  - Known since 1987
- Unlikely to predict for the right reasons
  - Many free parameters
  - Input data time/place specific
  - Structural problems
    - Key real-world effects not modeled

The Model is Reality in the Regulatory System

- SIP and I/M regulatory targets are set by model rather than real-world measurements.
- Regulators use model to “determine” how programs will work in future.
- “Recent tunnel studies and measurements of tailpipe emissions suggest that actual VOC emissions from highway vehicles are much higher than traditional estimates, but as of yet, these suggestions are unsubstantiated by emission models”
(EPA (1991) National Air Pollutant Emission Estimates, EPA-450/4-91-026)

ARB on Benefits of I/M and OBD

A Study Of The Relative Benefits Of On-Board Diagnostics And Inspection And Maintenance In California

Dilip Patel, Mark A. Carlock

California Air Resources Board, SAE Paper, 1995

ABSTRACT

California is considering adopting an enhanced Inspection and Maintenance (I&M) program (commonly referred to as Smog Check II) beginning with the 1996 calendar year. This program will utilize a targeting scheme to identify vehicles likely to be high emitters and send these vehicles to centralized testing facilities. The remaining fleet of vehicles will be sent to decentralized testing facilities. At these facilities, vehicles will be subjected to steady state loaded mode dynamometer based tests. Simultaneously, all 1996 and later model year passenger cars, light- and medium-duty trucks sold in California will be equipped with an On-Board Diagnostic (OBDII) system. This system is designed to monitor critical emission related components and activate a Malfunction Indicator Light (MIL) when a failure or a drift in calibration is likely to cause emissions to exceed 1.5 times the vehicle certification standards. **The main objective of this paper is to ascertain what percentage of the emission benefits are attributable to either I&M or OBDII in order to assist regulators in making near term programmatic decisions. This paper also addresses the potential emission benefits of incorporating a radio transponder into the vehicle's OBDII system that is capable of transmitting fault codes when queried, resulting in prompt identification and repair.**

- No data collected or analyzed for this “study”
- Based solely on EMFAC model runs with input assumptions for enhanced I/M and OBD effect

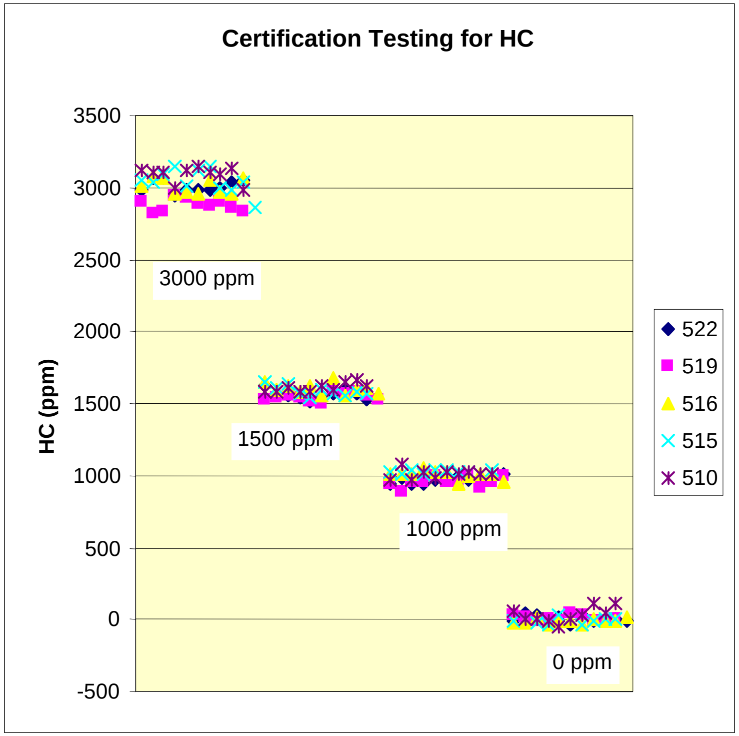
Summary of Emission Modeling

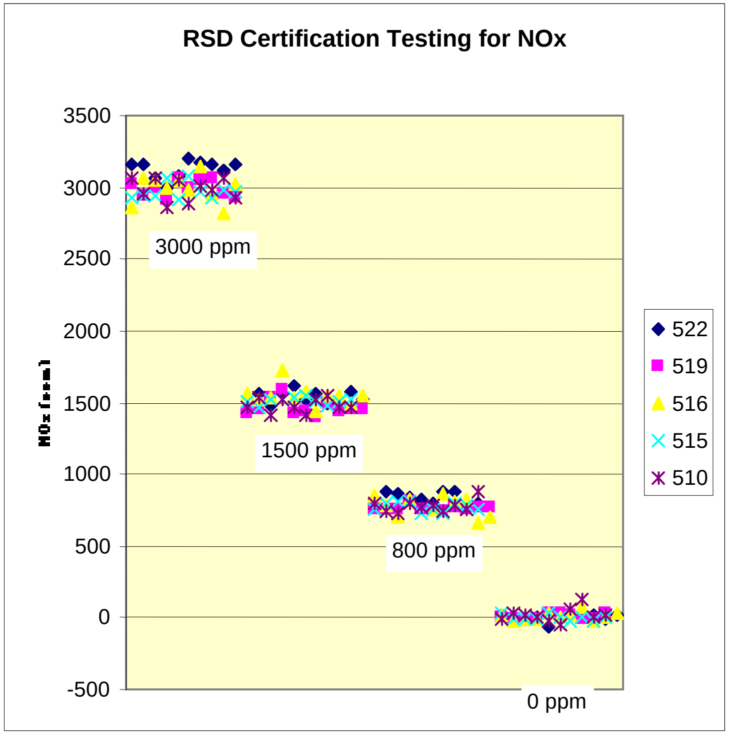
- The model is only as good as its inputs and structure
  - Inputs reflect the full range of problems in analysis and reporting of data outlined above
    - Static view of world: no feedback effects
    - Inappropriate analytical boundaries: idealized studies exclude key real-world effects
    - Unsound data analysis methods
  - Invalidation by real-world measurements both overall and in detail
- Nevertheless, models continue as drivers of regulatory policy

Case Study: Remote Sensing

- Overall results of various evaluations of the technology:
  - Instantaneous measurements are accurate
  - Selecting highest RSD emitters yields low “false fail” rates in contemporaneous comparisons with IM240
  - Average fleet emissions measurements (say by model year) show excellent correlation between RSD and other tests (r^2 generally above 0.95)

High Instantaneous Accuracy



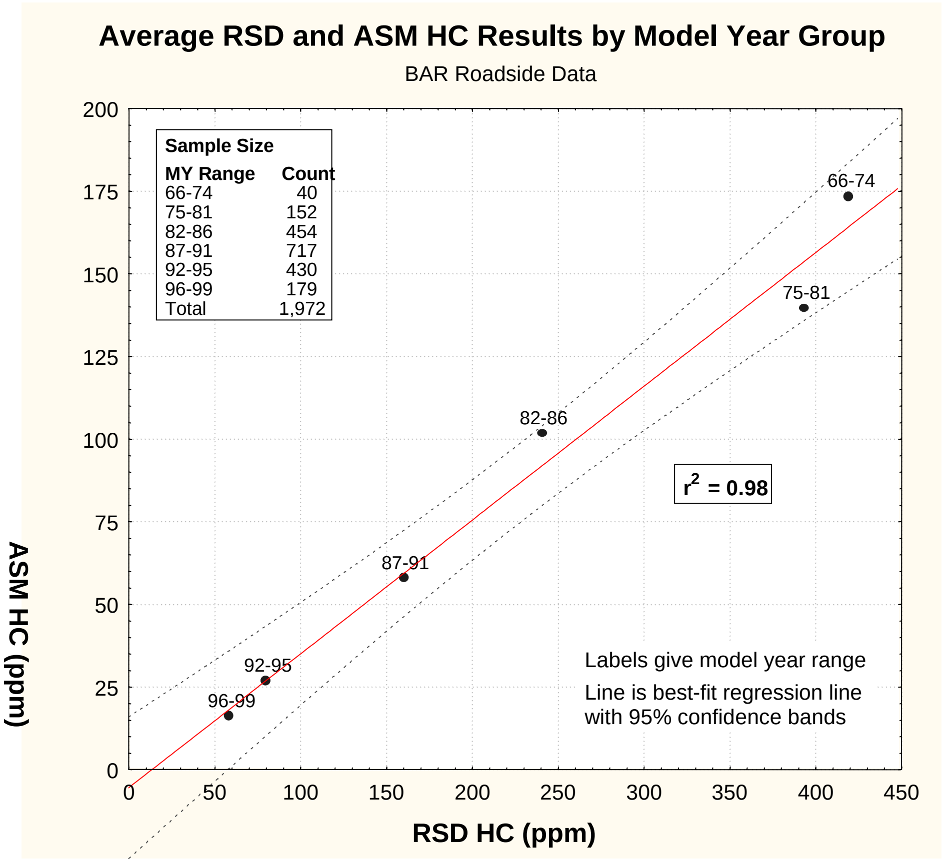


Good High Emitter ID

- Several pullover studies using RSD to select high emitters, followed by IM240 testing
  - "False fail" rates at about 8% or below when comparing RSD cut points of CO>4%, HC>0.1% with EPA Final IM240 cut points
- New techniques using vehicle load estimates likely to allow even better determination of vehicle status

Excellent Evaluation Tool

- Very large samples (easy to get > 10^6 cars) at range of sites and loads
- Find out % of cars on road but unregistered, failed without passing, etc.
- High correlation with other tests in average sub-fleet comparisons



Issues

- How will motorists respond?
- Can we follow through to ensure repair?
- What are the right cut points?
- Can we use to encourage voluntary repairs?
- How much fleet coverage can we achieve and how much is necessary?

Mischaracterization of RSD Data

- Inaccurate statements
  - "It is clear that the technology is incapable of detecting the most common forms of tampering, prone to a high rate of false failures, and unable to identify most of the excess emissions that exist in the fleet even using cutpoints that yield a relatively high false failure rate." (Sierra Research, 1995)
  - "Analyses of remote sensing data that draw conclusions about the distribution of emissions in the vehicle fleet are nonsense." (Sierra Research, 1995). (Addressing RSD results indicating that roughly 10% of the fleet produces 50% of emissions for CO and HC (a different, but overlapping 10% in each case)).
    - Actual Results: In fact, IM240 data and all other test data give the same skewed distribution. Furthermore, fleet average RSD and IM240 emissions show excellent correlation (r^2 typically greater than 0.95 for model-year averages) suggesting that, on average, RSD and other tests are measuring the same thing.
- Statements that are true but misleading due to lack of context
  - "Using mobile vans equipped with remote sensing devices, measurements could not be obtained on 75% of pre-1980 model vehicles during 500 van-days of testing." (Sierra Research, 1995)
    - Actual Results: Fleet coverage was increasing at a rate of 5.5% of the fleet per 100 van-days when study was ended. One of every three cars was a first-time read at end of study.

CONTRIBUTION OF OLD CARS VS. NEW CARS

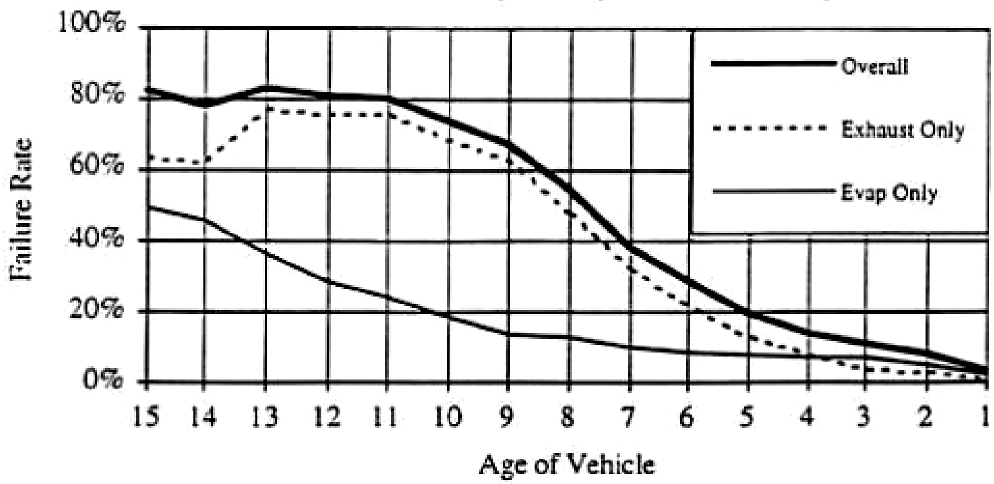
Many people have suggested that the problem with motor vehicles is limited to a small percentage of the fleet, typically 10%. The thinking goes that if we could find a way to identify the 10% worst emitters without having to test all of the vehicles, we would solve the problem. The table below shows the contribution of different fractions of the fleet ranked by CO emission levels. CO was chosen for the ranking only because the Remote Sensing Device CO channel is reasonably capable of accurately testing instantaneous emissions (although instantaneous emissions are not very representative of overall performance of a vehicle). The table shows that while the worst 10% of the CO emitters represent about 50% of the CO emissions, these vehicles only contribute 33% of the HC and 11% of the NOx. Moving down the table, you find that in order to get a significant fraction of all of the pollutants you need to fail a much larger fraction of the fleet. EPA estimates that about 40-45% of the vehicles in the fleet need repair. EPA recommends that states phase-in emission standards on the IM240 over at least two test cycles. By doing this, the overall failure rates will be in the 30-40% range. With the looser standards, the dirtiest cars will fail in the first test cycle, which will yield the biggest emission reductions and benefits.

More Than 10% of the Fleet Needs Repair

		Percent of Total Emissions		
CO Emitters	**Fraction Of Fleet**	**HC**	**CO**	**NOx**
Top 10%	10%	33%	52%	11%
Top 20%	20%	49%	67%	24%
Top 30%	30%	60%	77%	37%
Top 40%	40%	69%	83%	48%
Top 50%	50%	75%	88%	59%

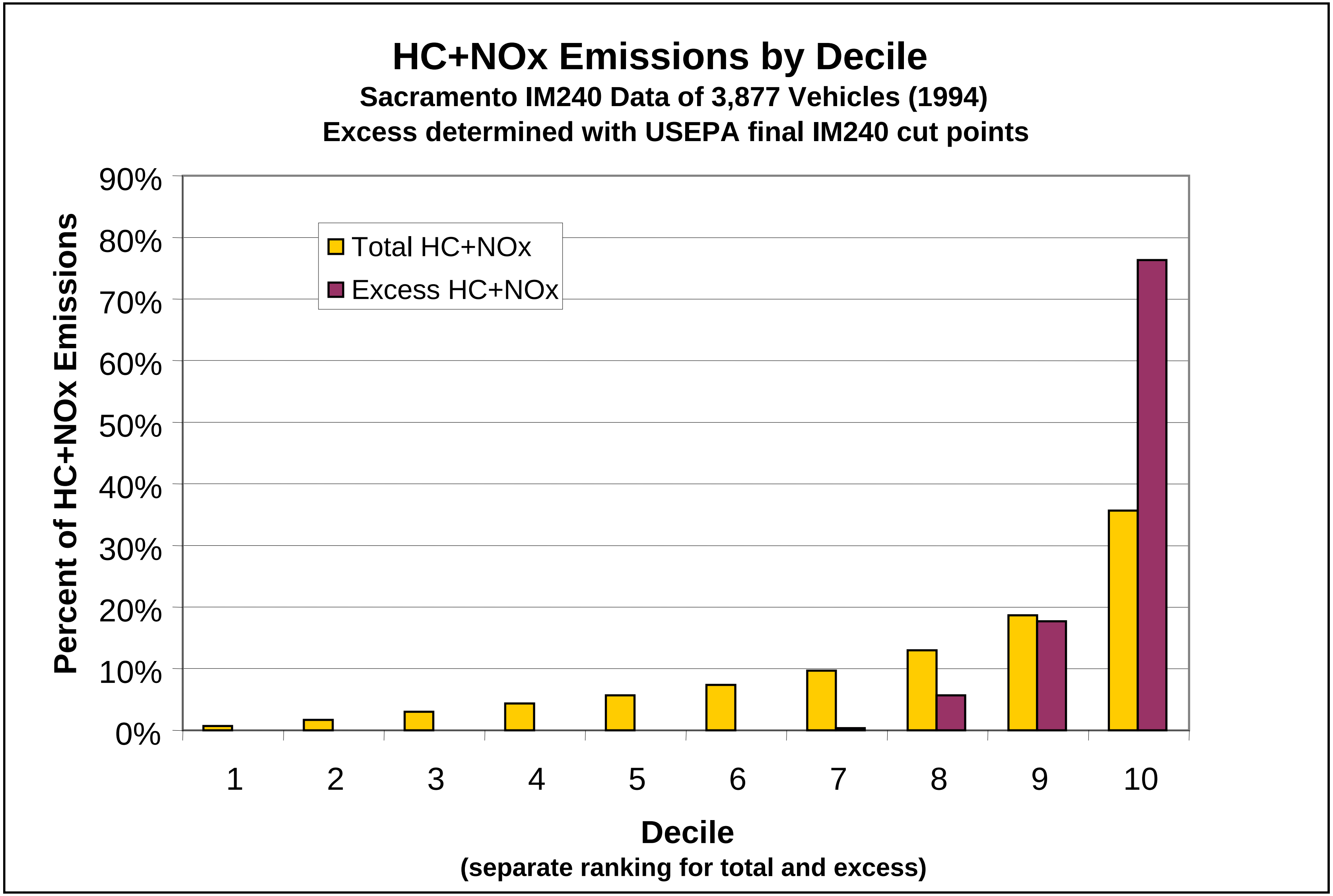
Some people argue that the problem could be solved by scrapping old vehicles or requiring only old vehicles to participate in I/M. While EPA believes that new vehicles don't need to be tested until 2 or 4 years of age, testing or scrapping only old cars will not achieve the air quality goal. The figure below shows that while most older vehicles need repair, so too do newer vehicles.

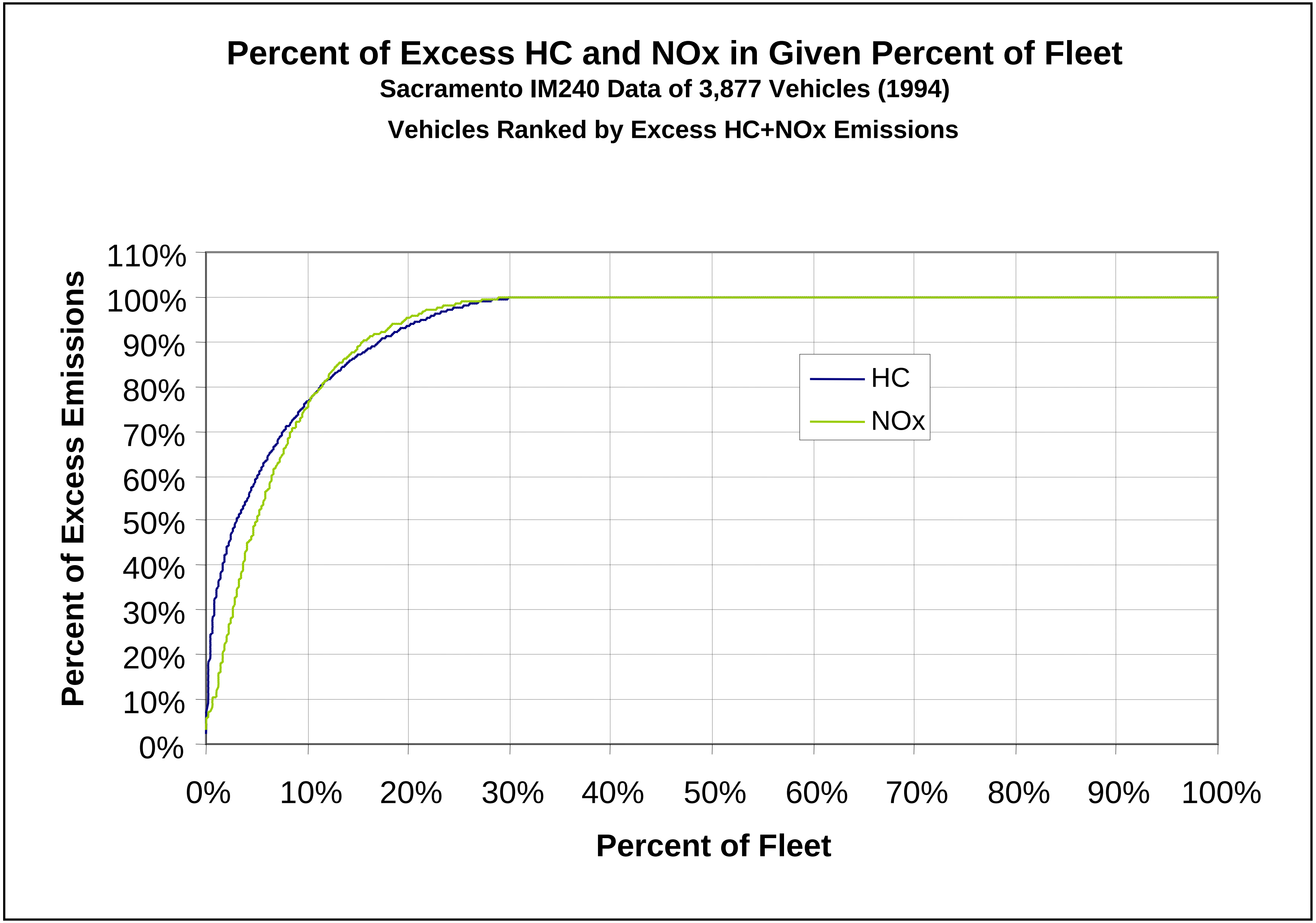
Percentage of Vehicles Failing Stringent IM240 Cutpoints



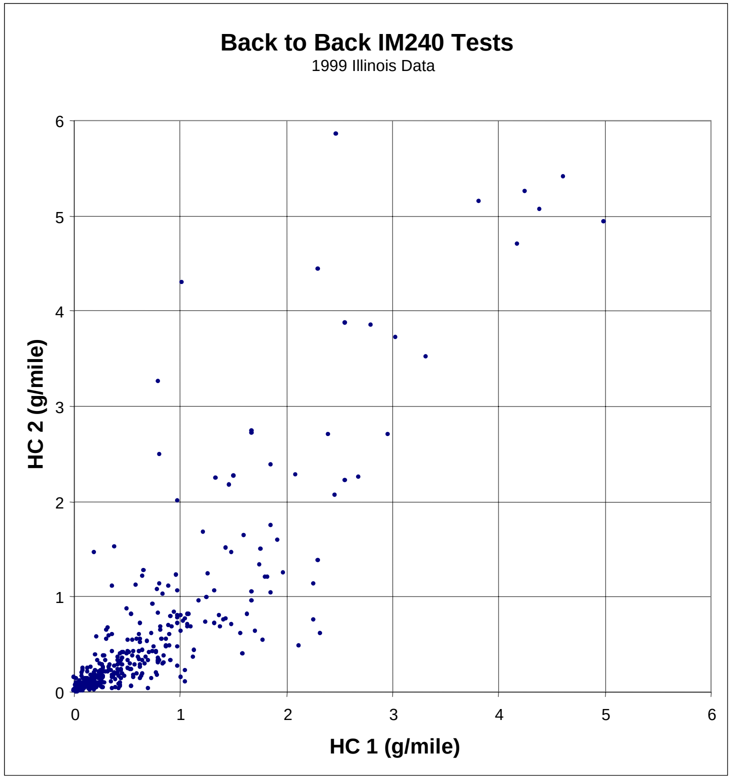
Up to 20% of five year old vehicles fail either the evaporative system test of the tailpipe exhaust test. If these newer vehicles are not fixed along with the older vehicles, then air quality goals will not be achieved. It is especially important to identify problems when vehicles are young so that they

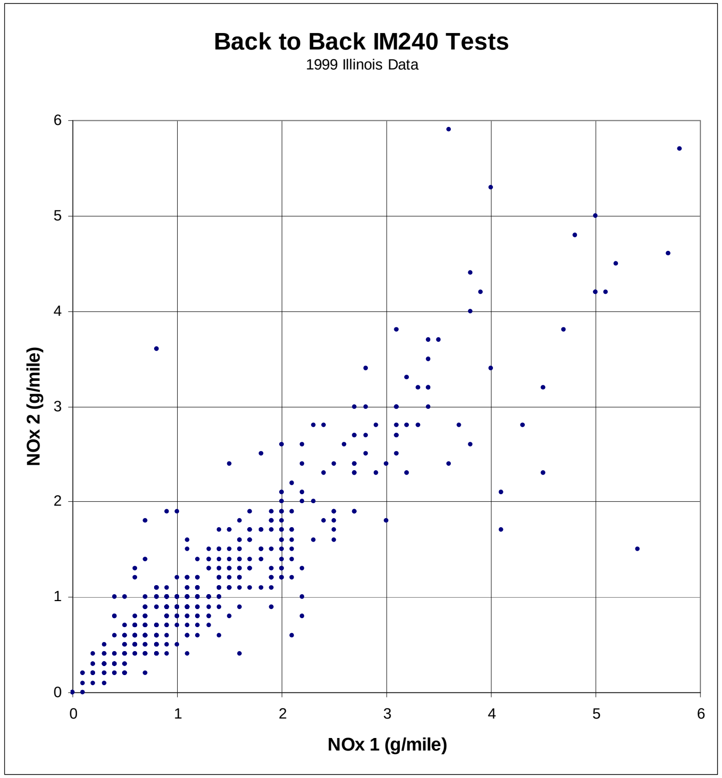
Source: EPA (1995) "I/M Briefing Book," EPA-AA-EPSD-IM-94-1226


HC+NOx Emissions by Decile
Sacramento IM240 Data of 3,877 Vehicles (1994)
Excess determined with USEPA final IM240 cut points
Percent of HC+NOx Emissions
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Total HC+NOx
Excess HC+NOx
1
2
3
4
5
6
7
8
9
10
Decile
(separate ranking for total and excess)


Percent of Excess HC and NOx in Given Percent of Fleet
Sacramento IM240 Data of 3,877 Vehicles (1994)
Vehicles Ranked by Excess HC+NOx Emissions
Percent of Excess Emissions
110%
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
HC
NOx
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Percent of Fleet

Missing Context: Vehicles Intrinsically Variable Even in Back-to-Back Tests





- HC and NOx $r^2 = 0.86$ (EPA 1990 Indiana data gave $r^2 = 0.66$ when a few days passed between tests)
- Using USEPA phase-in cut points and considering second test as "true" emissions
  - 20% of failing cars are false fails
  - 26% of all "should-fail" cars actually pass

Apples and Oranges Comparisons

- Traditional I/M is idealized
  - Loaded-mode testing with test-only network captures virtually all potential benefits available from I/M
    - Assumes little motorist avoidance, high-repair effectiveness, long-lasting repairs, no test variability.
- RSD portrayed in overly pessimistic way
  - Motorists will avoid driving by it, or accelerate or turn off engine
  - Too many false failures and false passes
  - Only fraction of fleet covered
  - RSD credited only with small fraction of potential reductions not captured by traditional program
- What's missing?
  - realistic view of problems with current programs
  - serious exploration of potential ways RSD could help

Case Study: I/M Managers and Policymakers

- I/M paradigm inertia among I/M managers, policymakers, and their consultants and contractors
- Embodied in Colorado Conference
  - Attended mainly by I/M program managers, contractors and consultants

“I/M Future” Panel

- 4 of 5 speakers don’t mention motorist behavior as an important policy variable in future I/M effectiveness
- Sierra Research:
  - “The Arizona program is currently the best example of a good program”
  - “OBD III will largely eliminate current concerns about I/M test equipment, inspectors, and contractors”
- EPA: Need to develop more sensitive tests for upcoming ULEV cars

“I/M Program Evaluation” Panel

- Focus was on how to get SIP credit from EPA
- No discussion of key variables
  - No mention of measuring real-world emissions reductions
  - No discussion of existence and measurement of motorist avoidance
- No real-world researchers on panel

“Human Dimension in I/M” Panel

- Discussed:
  - evidence for how people behave in I/M programs
  - evidence for motorist avoidance
  - evidence for voluntary repairs using “Smart Sign”
- No audience linkage re: relevance of presentations for I/M effectiveness or I/M policy

How Do I/M Policymakers Think about I/M?

- Focus on technology and not people
  - Very high emissions measurement accuracy is key
  - Technology and proper testing will cause people to behave the way we want
- I/M is a static problem
  - No feedback or dynamic effects when policies change
- Models and surrogate measurements tell us whether program is successful
- What happens at the test lane rather than on road is main focus

Case Study: OBD and Future of I/M

- The regulators' paradigm
  - OBD II and III will:
    - catch virtually all emissions problems
    - require motorists to repair their cars in order to turn off the MIL

percentage of vehicles in each regime by the average regime emission rate. The emissions benefit from one I&M cycle for a particular technology group can be estimated by comparing the average emission rates before and after repair.

The CALIMFAC model was recently modified by the Radian Corporation (under contract to BAR) to estimate the emissions benefit from the 1996 enhanced I&M program. The key assumptions used to model the 1996 I&M program are:

- A vehicle will be tested every two years (biennial program).
- Steady state loaded mode tests will be used to identify malperforming vehicles.
- Usage of a high emitter targeting profile to identify likely failures.
  Approximately 30% of the fleet will be targeted and sent to centralized testing facilities.
- The remaining fleet of vehicles will be sent to decentralized testing facilities.

The 1996 I&M program as modeled by Radian included a 50% repair disbenefit for vehicles going to decentralized test and repair facilities as dictated by the U. S. Environmental Protection Agency. This disbenefit was removed in the following analyses since the baseline emission factors for the current I&M program already implicitly contains a disbenefit. Including this disbenefit would result in double counting the repair disbenefit. Additionally, OBDII vehicles will not realize this disbenefit since these vehicles will be correctly repaired, in order to deactivate the MIL, regardless of whether the vehicles are sent to centralized or decentralized facilities for inspection.

The CALIMFAC model was further modified to account for OBDII vehicles. The key assumptions used in modifying the model were:

- The OBDII system will restrict the growth of high, very high and super emitters for the first 70,000 miles of in use driving since most major emission control components are warranted for this time and owners are more likely to get their vehicles repaired immediately under warranty.
- The OBDII system will identify 95% [2] of the vehicles in the high, very high and super emission regimes since the system is designed to detect malfunctions or a drift in calibrations that will likely cause an increase in emissions 1.5x the vehicle certification standards.
- Malperforming vehicles identified by the OBDII system will be effectively repaired since the fault code will identify the source of the malfunction and the mechanics must correctly repair the vehicle in order to deactivate the MIL. Within the model, this element is reflected by moving those vehicles in the high, very high and super emission regimes in equal proportion to the normal and moderate regimes.
- After the first 70,000 mile of use, the movement of OBDII equipped vehicles from the high, very high and super emission regimes to the moderate and normal emission regimes occurs regardless of whether the vehicle encounters a centralized or decentralized inspection test. This implies that there is no difference in the repair disbenefit for OBDII vehicles.

In determining the emission benefits from the implementation of OBDII vehicles it is assumed that the OBD system will identify 95% of the vehicles that are high, very high and super emitters. To verify this assumption, four FTP tests were performed on a 1995 OBDII compliant vehicle that was equipped with an 8 cylinder, 4.6 liter displacement engine with 3,124 miles on the odometer. Table 3 shows the FTP composite emissions as a result of various defects (listed below) being induced in the vehicle. Table 3 also lists the fault codes that were stored in the vehicle's OBDII computer after each test.

Defect	HC (g/mi)	CO (g/mi)	NOx (g/mi)	MIL-Status	Codes
Baseline	0.101	1.059	0.089	OFF	None
1	0.439	6.896	0.166	ON	P0125/P0155
2	0.095	0.756	0.188	ON	P1407
3	0.122	1.002	0.171	OFF	None

Table 3 Composite FTP Results

1 Disconnected two oxygen sensors before the catalyst
2 Blocked off the EGR valve
3 Disconnected the purge flow line, coming from the purge valve control solenoid, at the engine

P0125/P0155 Generic codes indicating a heater circuit malfunction in bank one, sensors one and two.
P1407 Manufacturer specific code indicating that there is no EGR flow.

Although, the testing of the vehicle's OBDII system was not exhaustive, the results indicate that the OBD system performed as required by CARB regulation [3] with the exception of the purge flow monitoring system. Further investigation revealed that the manufacturer had certified this vehicle as being deficient with respect to the purge flow monitoring system. The OBDII system on this vehicle contains a continuity check of the purge flow control valve to ensure it's proper operation, however, the flow itself is not monitored. The current OBDII regulation allows manufacturers to certify their vehicles as being OBDII compliant with one deficiency upto the year 1999, thereafter, the vehicles must be fully OBDII compliant.

RESULTS

Figure 2 shows the reduction in HC emissions from the 1990 I&M program, with and without the introduction of OBDII equipped vehicles. The emissions benefit is calculated

Source: ARB (1995) “A Study of the Relative Benefits of On-Board Diagnostics and Inspection and Maintenance in California” by comparing the with I&M (or with OBDII) emission rates to the without I&M and without OBDII emission rates. Figure 2 shows that the projected percentage emission benefits from the current I&M program without OBDII vehicles will increase with time. This increase in the percentage of emissions reduced is misleading in that it does not represent an actual improvement in the current I&M program, rather it reflects the implementation of zero emission vehicles (electric cars). For example, if the emission benefits from the current I&M program is 20% in the 1997 calendar year, then in the year 2000 the gasoline fueled fleet (approximately 98%) will experience a similar reduction on a g/mi basis. Therefore, the emission reduction on a percentage basis increases since only gasoline fueled vehicles see this benefit. The benefits from implementing OBDII vehicles without an I&M program continue to increase beyond the year 2020 due to the dominance of OBDII equipped vehicles in the future calendar years. The OBDII benefits result from owners of malfunctioning vehicles having them repaired immediately under warranty. The main impact due to the introduction of OBDII vehicles is that it will increase the effectiveness of the current I&M program. The benefits of the I&M program with OBDII are from: detecting malfunctions in older technology vehicles, OBDII benefits from the immediate repair of malfunctions during the first 70,000 miles of use, and the OBDII system identifying malfunctions (after 70,000 miles of use) which are then repaired at the vehicles next Smog Check. Figure 2 shows that the incremental HC emission benefit from implementing OBDII vehicles is 7.4%.

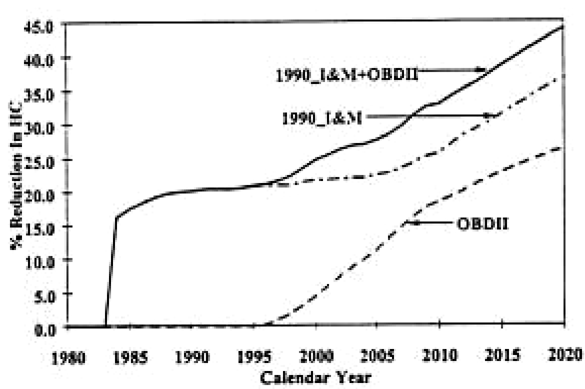


Figure 2 Reduction In HC Emissions From The 1990 I&M Program, With And Without OBDII Vehicles

Figure 3 shows the calendar year specific FTP composite HC emission (g/mi) rates if there is no I&M program, with the implementation of OBDII vehicles but no I&M program, 1990 I&M program and 1990 I&M program with the implementation of OBDII vehicles. This figure shows that the general trend is towards lower calendar year specific HC emission rates due to the phase-in of newer and cleaner (Low Emission Vehicles) technology vehicles. The highest emission rate will result from not having an I&M program and not implementing OBDII vehicles. The lowest emission rate will result from implementing both of these programs.

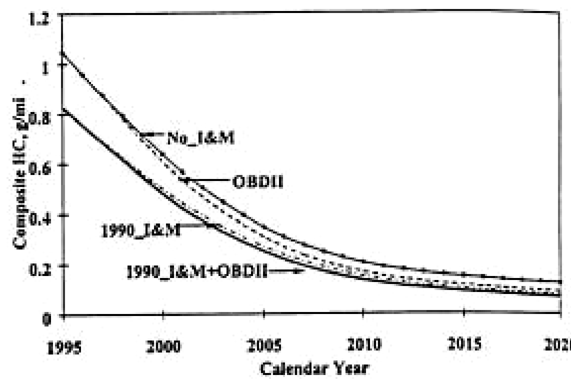


Figure 3 Comparison Of Fleet Average HC Emissions For No_I&M, 1990_I&M, OBDII_Only & 1990 I&M Program With OBDII

Figure 4 shows the reduction in CO emissions from the 1990 I&M program, with and without OBDII. The percent reduction in CO follows trends similar to that observed for HC, however, the incremental benefit from implementing OBDII is less than 1%. This does not imply that the OBDII system is less effective in identifying CO failures rather it reflects how OBDII is currently modeled in CALIMFAC. Within the model it is assumed that the OBDII system will increase the probability of identifying a malperforming vehicle. Rather than increasing the probability of identification it would be more correct to fail a vehicle based on the OBDII fault code which will increase the effectiveness of OBDII.

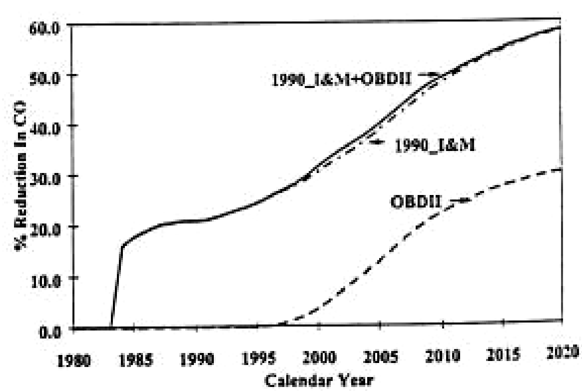


Figure 4 Reduction In CO Emissions From The 1990 I&M Program, With And Without OBDII

Figure 5 shows the reduction in NOx from the 1990 I&M program, with and without OBDII. The NOx emission reduction trends are similar to that observed for HC. The

Sierra Research on OBD

OBDIII AS REPLACEMENT TO I/M

? Earlier this year Sierra completed a successful field demonstration of this technology in Sacramento.

? OBDIII is more effective in detecting emissions-related defects than the best enhanced program and identifies problems much earlier.

? At less than $100/new vehicle, it is much cheaper than I/M.

? It will make it totally unnecessary for millions of passing vehicles to be subject to I/M tests.

? It will largely eliminate current concerns about I/M test equipment, inspectors, and contractors.

- Slide from Sierra presentation at 1999 Mobile Sources Clean Air Conference

Real Effectiveness Questionable

- Results so far:
  - High “false failure” rate
  - Some evidence that motorists are learning to ignore MIL
- Past experience suggests that OBD, like centralized IM240 testing, will not ensure motorists behave the way regulators desire

Inappropriate Problem Definition

- Sierra and ARB assume that if the *technology* performs as expected, a program based on OBD emissions detection will be effective
  - Underplays the human variables, dynamic factors, etc. that we’ve been detailing

Air Planning Process Exacerbates Analytical Problems

- Focused on inputs rather than results
  - EPA prescribes program details rather than required outcomes
  - SIP credits, rather than real reductions, drive decision making by both EPA and states
- Focused on models rather than real world
- Prospective "credit" rather than practical results

Creates Wrong Incentives

- States have incentive only to do things that receive credit
  - Often different from what would be effective
- Inertia against policies that don’t fit predetermined paradigm
- Evaluation focus is on how to show the things that get credited rather than real-world outcomes
- Low analytical standards probably both cause and effect

The NRC's Role

- The problem:
  - How to design and operate effective and cost-effective I/M programs
  - How to design institutions that result in good programs
- No one really knows what would be effective due to lack of systematic effort to ensure that most high emitters receive substantive repairs
- Urgent need for *institutional* changes that would be more likely to ensure that states *discover and implement* program elements that will reduce real-world emissions from the on-road fleet

Recommendations NRC Should Consider

- Means to elevate quality of EPA’s and states’ science
- EPA should focus on outcomes (emission reductions in this case) not means
- NRC-supervised peer-review of I/M science and program management
- “Plan B” for incremental change

Elevate I/M Science Quality

- System must ensure adequate appraisals of real-world effectiveness and cost effectiveness.
- Outside peer-review of states' and EPA's science is essential
  – Peer review should be prospective and/or coincident rather than months or years after the fact
  – Peer-review panels should be selected by independent organizations such as NAS

EPA Should Prescribe Goals, Not Means

- States should have real flexibility in meeting goals
- Goals should be transparent
- Means of assessment should be clear and scientifically valid
- Sanctions for failure should be rapid, progressive, and automatic
- Would require Clean Air Act changes

Assessment of I/M Science and Management

- Avoid trying to design the “best” system
  - We don’t know what that would be and it likely varies from state to state
- Say what we know and don’t know about vehicle emissions and I/M effectiveness
- Suggest policy options for states to consider in trying to improve effectiveness and cost effectiveness
- Recommend minimum quality standards for mobile source research and evaluation

Changes Will Be Difficult

- Major culture change at EPA and state agencies required
- Clean Air Act Amendments might not be possible
- NRC therefore needs “Plan B” for incremental improvements to system
  - Should recommend specific changes that:
    - Are well supported by available data and analysis
    - Don’t involve major new capital investments or expensive long-term commitments
    - Reduce program costs
    - Increase follow-through on substantive repair
    - Ensure independent, real-world assessment of effectiveness