COMMONWEALTH OF MASSACHUSETTS
EXECUTIVE OFFICE OF ENVIRONMENTAL AFFAIRS
**DEPARTMENT OF ENVIRONMENTAL PROTECTION**
ONE WINTER STREET, BOSTON, MA 02108 617-292-5500

MITT ROMNEY
Governor

KERRY HEALEY
Lieutenant Governor

ELLEN ROY HERZFELDER
Secretary

ROBERT W. GOLLEDGE, Jr.
Commissioner

# SUMMARY

## OF THE

## MA31 CONVERSION FACTOR ANALYSIS AND

## INTERIM TEST EFFECTIVENESS EVALUATION

## JULY 21, 2003

## ABSTRACT

On November 16, 2000, EPA gave final approval to the current Massachusetts I&M Program. Approval of the I&M State Implementation Plan recognized that the Massachusetts test equipment was less effective than EPA's benchmark test (IM240), established interim "test effectiveness" targets that the Massachusetts test was required to meet, required that Massachusetts conduct a study to verify whether those targets were being reached, and stated that EPA would take action to make up for any shortfall if those targets were not reached.

That study – "MA31 Conversion Factor Analysis and Interim Test Effectiveness Evaluation" - is being submitted to EPA today to fulfill that EPA requirement. In order to ensure that the study met EPA's expectations, EPA and DEP cooperatively designed the study.

The study had three objectives:

1. Evaluate the conversion factors used by the Massachusetts program to express test results in terms that are comparable to IM240, and recommend improvements if needed so that Massachusetts test results more closely approximate results that would be expected from the IM240 test.

2. Evaluate the effectiveness of the MA31 drive trace relative to the IM240 drive trace. "Test Effectiveness" measures how well a state's I&M test equipment and drive trace identify excess emissions from a vehicle fleet compared to EPA's benchmark equipment and drive trace, IM240.

3. Evaluate potential program changes that could increase the MA31 test effectiveness, if needed.

The study found that:

1. The conversion factors needed to be revised. In **July, 2001**, DEP revised the conversion factors to make the test more accurate. That change also had the effect of decreasing failure rates. A final change to the conversion factors will be implemented by DEP after consultation with EPA and discussing the proposed changes with stakeholders. This change, which DEP expects to make by October 2003, will also increase NOx effectiveness.

2. The effectiveness of the Massachusetts test after the change in conversion factors in July 2001 exceeded target levels for HC and CO, and was below the target level for NOx. In **March 2003** DEP made a program change that increased test effectiveness as follows:
    - HC: 91% compared to 85% target
    - CO: 93% compared to 87% target
    - NOx: 75% compared to 85% target

3. DEP should improve test effectiveness for NOx, and should consider implementing a longer test or changing the pass/fail cutpoints to make the Massachusetts test stricter.

# DETERMINATION OF POLLUTANT-SPECIFIC CONVERSION FACTORS FOR MASSACHUSETTS-TO-IM240 CONVERSION OF EMISSIONS TEST RESULTS

## INTRODUCTION

### Why Does Massachusetts Have an I&M Program?

Massachusetts continues to violate minimum federal standards for ground-level ozone pollution. On our "bad air" days there is an increase in asthma attacks and hospitalizations for severe respiratory ailments. In order to reduce the number of "bad air" days and to comply with the US Clean Air Act and EPA regulations, Massachusetts must implement various federally mandated programs. To reduce pollution from motor vehicles, Massachusetts is required to operate an Enhanced Inspection and Maintenance (I&M) program. I&M programs identify onroad gasoline motor vehicles that substantially exceed their designed emissions levels by periodically testing them. Repairs are required for vehicles that fail the emissions test. The U.S. Environmental Protection Agency (EPA) sets minimum standards for I&M programs[1].

States include their I&M programs in a State Implementation Plan (SIP), a federally-enforceable binding commitment from the state to EPA, that lists the programs the state will use to meet federal pollution reduction goals. These programs include reductions from many sources, such as power plants, small businesses, factories, construction equipment and onroad vehicles. For I&M programs, EPA establishes different levels of credit to states for reducing pollution based on the kind of I&M test and program the state implements.

### How Does an I&M Program Work?

Recent – or "Enhanced" - I&M programs require placing vehicles on a dynamometer, a treadmill-like device that puts resistance against the tires to simulate onroad driving. Then vehicles are accelerated and decelerated according to a prescribed driving pattern ("drive trace"), and tailpipe pollutant levels are collected and recorded. Pollutant readings for hydrocarbons (HC), Carbon Monoxide (CO) and Oxides of Nitrogen (NOx) are compared against each pollutant's pass/fail points.

[1] 40 CFR Part 51, Subpart S (§51.350 et seq.).

## What is EPA's Model I&M Program?

EPA developed a "model" enhanced I&M test, called IM240. This model program tests vehicles using laboratory-grade instruments in stand-alone contractor-run buildings. In these centers, which only perform emissions testing (i.e., no safety inspections or repairs), emissions are continually measured for 4 minutes on a dynamometer. The IM240 dynamometer drive trace is derived from the EPA test used to certify new vehicles, and covers almost 2 miles of simulated driving at speeds of up to 57 mph.

Many states believed that this IM240 model was too expensive, was inconvenient for motorists, or did not fit well with the existing businesses that performed emissions tests. In some states, early implementation of the model IM240 program was a failure: long lines, motorist revolts and the canceling of multi-million dollar contracts. In response to this situation, Congress in 1995 passed a law that required EPA to grant states flexibility to select alternative tests and programs.

EPA's IM240 guidance proposes two different pass/fail points (cutpoints). EPA allows states to choose either the less stringent "start-up" cutpoints or the more stringent "final". Both sets of cutpoints are designed to identify vehicles with malfunctioning emission control systems that pollute much more than they were designed to. EPA final cutpoints are stricter and obtain more pollution reductions. EPA Start-up cutpoints are more lenient and obtain fewer reductions. Massachusetts uses the more lenient cutpoints because its air quality plans (the SIP) did not require greater air pollution reductions from motorists.

## How are alternative I&M programs different than EPA's IM240 model?

One difference in the programs is where the testing is located. EPA envisioned the creation of a relatively small number of stand-alone emissions test centers and requiring all motorists to get tested there. That approval would have taken the emissions testing business out of local, conveniently located inspection shops and forced motorists to drive to one of only 60 locations statewide for testing. Additionally, the question of how to accommodate safety inspections with an IM240 emissions program was complicated. Safety inspections would either be taken away from existing businesses or motorists would have to drive to two different locations for the tests.

Another difference is the effectiveness of alternative tests. EPA considers these alternative programs to be less effective than the IM240 model because the drive traces tend to be shorter and at lower speeds, the testing equipment is usually cheaper and less accurate, and because it believes that tests are delivered more accurately in test-only businesses compared to businesses that also repair vehicles. Because EPA considers these alternative tests to be less effective than IM240, EPA assigns less emissions credit to states with alternative I&M tests.

## Why Didn't Massachusetts used the EPA Model Test, IM240?

The Department of Environmental Protection (DEP) and the Registry of Motor Vehicles (RMV) are jointly responsible for implementation of the Massachusetts Enhanced Emission and Safety

Test. Their goals are to provide a comprehensive test that is convenient to motorists, works well in local inspection shops, provides the emission reductions needed for the SIP, and ensures vehicle safety.

In the mid-1990s, after Maine, Pennsylvania, and Texas had abandoned their centralized IM240 programs, the Administration directed the DEP and the RMV to examine the possibility of implementing an alternative test in Massachusetts.

After conducting an "I&M Summit" with over 30 state officials from 11 different state agencies, and holding a workshop with I&M business and equipment officials, national trade associations, and other stakeholders, the DEP and RMV decided that the Massachusetts I&M program would have to balance as well as possible the achievement of three different goals:

- It had to be convenient in terms of price and location for Massachusetts motorists.
- It had to fit well with the business plans of the inspection and repair industries, ensuring that the private sector could still deliver this program.
- It had to achieve significant pollution reductions.

Development of an alternative test and a compatible contract model took over 2 years. It included technical research by DEP and RMV staff, thoughtful review and consideration by a large and diverse I&M Advisory Committee, publication of a report highlighting the kind of choices that Massachusetts was considering, and passage of legislation in late 1997 that permitted DEP and RMV to proceed.

Implementing IM240 in Massachusetts would have cost inspection stations significantly more to buy or lease equipment ($55,000 to $65,000 per test lane compared to the current $36,000). If the IM240 test were kept in local shops, the inspection fee would have cost up to $50; it is currently $29. If the IM240 test were instead conducted in stand-alone test centers, motorists would have had to drive much greater distances for the test. Safety inspections would either be taken from local business and placed in the test centers, or motorist would have had to drive to different locations to obtain the two tests.

It takes 4 minutes to measure emissions on the IM240 test at speeds up to 57 mph. It takes from 1 1/2 to 3 minutes to measure emissions on the MA 31 test at speeds limited to 30 miles per hour. While the IM240 test would have provided a higher level of accuracy and emission reductions, it did not meet the goals that stakeholders had developed for Massachusetts.

## What Advantages Did the Alternative Test Offer Massachusetts?

The use of an alternative test helped optimize customer convenience in two ways. First, it offered a faster test that would reduce wait times for motorists. The difference in test times between the MA31 test and the EPA IM240 as actually delivered in the field – including set up and measuring emissions - could range from 3 ½ to 5 minutes. While that is a small increment for a single vehicle, the cumulative effect on wait times for a line of cars can be significant for people at the end of the line

Second, the less expensive equipment helped keep the motorists' fee below many other states. With many states I&M fees ranging from $40 to $50, the cumulative savings over our $29 fee for the estimated 4.2 million vehicles is significant.

The use of an alternative test also fit better with the business plans of the Massachusetts inspection industry. Operating the dynamometers at lower speeds (31 miles per hour vs. the IM240's top speed of 57 miles per hour) addressed industry concerns about noise and safety in a small shop environment. The cost of equipment was of great concern to the inspection industry. In addition, every minute saved on an emissions test was worth about $1 to the inspection stations. At 2-3 emissions tests per hour, stations viewed the savings as significant.

EPA – which approved the use of the Massachusetts, New York and Rhode Island alternative tests – believed that the equipment and trace were capable of achieving the level of pollution reductions described in Massachusetts' pollution reduction plans. Similar equipment is currently being used in New York and Rhode Island.

## What Alternative Test Did Massachusetts Develop?

To maximize customer convenience, DEP and the RMV decided to keep emissions and safety testing together, and to keep the combined test in local inspection stations, convenient to where people live and work. There are about 1600 public testing stations available now, instead of the 60 or so we would have had under EPA's plan.

To establish the pass/fail points for the program, Massachusetts decided to use EPA's less-stringent IM240 start-up cutpoints. Although these less-stringent cutpoints would result in fewer high-emitting vehicles being failed than the IM240 test, the emission reductions from the I&M program would still be sufficient to meet Massachusetts' SIP requirements.[2]

DEP and the RMV worked through an active I&M Advisory Committee composed of inspectors, educators, environmentalists, EPA and repairers to identify and develop a program and a test that best met the Commonwealth's goals and that the stakeholders believed would work well in Massachusetts.

The equipment that Massachusetts chose for its alternative test had less costly and less accurate pollution measuring equipment than the more costly laboratory grade IM240 equipment. This equipment, known as MASS99, was the same equipment selected for New York's I&M test, and is currently used in NY and RI.

[2] States typically use much looser cutpoints during program implementation to accommodate any bugs in the testing system or in the I&M program. As the program is implemented, states typically change their cutpoints in a series of steps until the program's final cutpoints are reached.

The drive trace that Massachusetts selected is a shorter, lower-speed 31 second dynamometer drive trace (MA31), with a top speed of 30 mph. The trace is the same as the "BAR31" alternative drive trace developed by California's Bureau of Automotive Repair (BAR), and is also used by Oregon and Rhode Island.

These two test features – less expensive repair-grade equipment and a shorter, lower speed drive trace - are together called the "MA31 test".

## How Does the I&M Test Work?

When a vehicle arrives for an enhanced I&M test, the inspector enters vehicle information into the test's computer system. Two key test features hinge on proper vehicle identification:

1. Dynamometer loading – for an accurate test, it is necessary for the dynamometer to place the proper resistance against vehicle's tires. The dynamometer loading calculations come from the EPA test used to originally certify that the vehicle meets "new car" emission standards. They vary by type and weight of vehicle.

2. Cutpoints (pass/fail points) – to determine whether a vehicle has passed or failed, it is necessary that test results be compared to the correct cutpoint. Cutpoints vary by vehicle type [car vs. truck], model year, and, for trucks, by weight category.

Once the vehicle is ready for testing, an inspector "drives" the vehicle on the dynamometer while the emissions and exhaust flow are collected and analyzed. At the conclusion of the test, the test system corrects for any differences from the IM240 equipment and drive trace, calculates the emissions and the distance driven on the dynamometer, and reports emissions in EPA IM240 standard grams per mile. The grams per mile results are then compared to the cutpoints (which are also expressed in EPA IM240 grams per mile), and a pass/fail determination is made.

## How Do Results From MA31 and IM240 Compare?

When a vehicle is tested using IM240 test equipment and the IM240 dynamometer drive trace (trace), emissions for each pollutant are measured and reported in EPA IM240 standard grams per mile for each pollutant. The emissions are compared to IM240 cutpoints to make a pass/fail determination.

When a vehicle is tested using the MA31 test (using MASS99 test equipment and the MA31 trace), emissions are first measured as a concentration (parts per million) and are then converted to a mass measurement (grams per mile). Then the software converts the amounts for each pollutant into EPA IM240 standard grams per mile before being compared to IM240 start-up cutpoints for a pass/fail determination. The initial reporting of emissions as a concentration and subsequent conversion to grams per mile is a byproduct of using the less expensive analytical equipment used in Massachusetts, Rhode Island and New York.

The software conversion of “raw” grams per mile readings into EPA IM240 standard grams per mile readings is necessary for two reasons. First, the MASS99 equipment has different measurement characteristics than IM240 equipment. Second, vehicles tested using the MA 31 drive trace are being driven on a different “road” than the IM240 drive trace. Just as driving on a different road will yield different “miles per gallon” figures, using a different drive trace yields different emissions readings. Since the Massachusetts pass/fail points are expressed in EPA IM240 standard grams per mile, the readings for this alternative drive trace must be converted to EPA IM240 standard grams per mile.

## How Are MA31 Test Results Converted to IM240 Standard Grams per Mile?

Almost all states that use an alternative I&M test convert their emission test results to EPA IM240 standard grams per mile for each pollutant. The vehicles emissions are then compared to the EPA cutpoints (also expressed in EPA IM240 standard grams per mile), and a pass/fail determination is made. The conversion can account for either a different drive trace, different equipment or both. Massachusetts uses both a different trace and different equipment than IM240.

Some states choose not to express emissions readings in EPA IM240 grams per mile. Rhode Island, for example, instead changes its cutpoints to account for its differences in drive trace and equipment.

The only software-based conversion factors available from other states at the beginning of the Massachusetts I&M program corrected for different equipment. DEP used equipment correction factors from the New York Department of Environmental Conservation, which uses the same equipment as Massachusetts.

Since there were no software-based conversion factors available to account for the differences in the MA31 drive trace, Massachusetts made three revisions: 1) adjusted its start-up cutpoints to accommodate the different trace; 2) added the NY correction factors to accommodate the different equipment; and 3) planned to conduct a study in the future that would determine a combined software correction factors for both the trace and the equipment.

## Why Change the Conversion Factors?

EPA conditionally approved the use of the conversion factors at the start-up of the Massachusetts I&M program, and stipulated that DEP must develop Massachusetts-specific conversion factors. In response to that requirement, DEP and EPA agreed that Massachusetts would conduct a study with two purposes: Massachusetts would be able to increase the accuracy of its test, and EPA would obtain information that it could use to establish final test effectiveness numbers for the Massachusetts test. In order to ensure that both needs were met, DEP sought EPA’s assistance during the development of the study. EPA reviewed and accepted the study design.

## CONVERSION FACTOR STUDY

In response to EPA's requirement, and to ensure that there was one software-based conversion factor for each pollutant that accommodated differences in both test equipment and drive trace, DEP initiated a study to compare the emission results from the IM240 and MA31 tests. The study and many of the methodological technical details were designed in cooperation with EPA staff. The results were designed to provide one conversion factor for each pollutant that would convert emissions gathered from the MA31 test into EPA IM240 standard grams per mile.

### How Was the Study Conducted?

In October 2000 DEP contracted with Sierra Research of Sacramento, California (Sierra) to perform the study. Sierra subcontracted with Gordon-Darby, Inc of Louisville, KY (Gordon-Darby) to test vehicles in Arizona using the IM240 and MA31 tests. Gordon-Darby operates the emissions testing program for the greater Phoenix, Arizona area, widely considered to best represent IM240. Agbar Technologies, the contractor administering the Massachusetts I&M program, was responsible for providing and maintaining a MASS99 test equipment for the study.

Vehicles were tested on the IM240 equipment first using an IM240 trace. This was followed by a test on the IM240 equipment using the MA31 trace. Vehicles were then moved to the MASS99 test equipment where they were tested using the MA31 trace, followed by an IM240 trace. This gave each test vehicle every possible combination of test equipment and trace.

### How Were Vehicles Selected for the Study?

Vehicles that showed up for their routine Arizona emissions inspection were candidates for study participation. To provide a diverse enough set of data to evaluate the conversion factors, it was necessary to test a variety of vehicle types and model years with various emission rates, including older dirtier vehicles with high emissions. The study design, which had been reviewed by EPA, established the goal of recruiting equal numbers of vehicles across model years, vehicle types, and level of emissions.

Testing for the study was performed from December 2000 to August 2001. There were 612 vehicles tested that met selection criteria and yielded valid tests.

# CONVERSION FACTOR ANALYSIS

## How Were the New Conversion Factors Calculated?

The emission results from the IM240 and MA31 tests are compared using a linear regression. This type of analysis plots the emission results and then forms a line that best fits the data points so that all the points are as close to the line as possible. The mathematical slope of the line becomes the conversion factor.

This type of analysis is necessary to match the MASS99 equipment's emission calculations, which are designed for a linear conversion of emission scores. The linear equation used in the MASS99 system is:

IM240 Equivalent Emissions = Raw MA31 Emissions X Conversion Factor

Since the main purpose of the conversion factors is to make accurate pass/fail determinations, it is critical that the linear regressions generating the conversion factors for each pollutant are as accurate as possible when emissions are near the cutpoints.

Test accuracy is much more important for vehicles that pollute at levels near the cutpoints. For those vehicles, normal variations in accuracy for the test coupled with normal variations in the amount of pollution emitted by any one vehicle can be enough to cause the vehicle to pass a test once, then fail a second time.

In contrast, normal variances in test accuracy for clean vehicles - for example, those vehicles that emit 10 times below the cutpoints - is much less important because those clean cars will still pass. Likewise, normal variances in test accuracy from gross polluters – for example, those vehicles that pollute 10 times above the cutpoints - is also much less important because those extremely dirty vehicles will still fail.

To ensure that the correction factor analysis was not biased by extremely high emitters, the study excluded data from the extremely high emitting vehicles for some pollutants.

## Why Were Interim Conversion Factors Calculated?

At the beginning of the program, DEP installed conversion factors in the software that only corrected for differences in the equipment, and adjusted the start-up cutpoints to correct for differences in the trace. Since DEP's final cutpoints did not include any correction for the drive trace, and since DEP moved to its final cutpoints in April 2001, DEP decided to analyze a preliminary dataset of 341 vehicles.

## What Were the Outcomes of the Interim Conversion Factor Analysis?

Table 1 compares the interim conversion factors to those initially used in the software. As can be seen, there was a substantial disparity between the two sets of conversion factors. The disparity was causing the reported emissions from vehicles *to be higher* than EPA IM240 standard grams per mile, which would cause motorists' emission readings to be inflated. This overestimation would also cause motorists to fail when they should have passed.

Following this analysis, on July 11, 2001, the conversion factors were changed to the interim values calculated from the preliminary 341-vehicle dataset. The change was made to increase the accuracy of the test and reduce false failures. It also had the effect of lowering the failure rates

**TABLE 1**
**COMPARISON OF INITIAL AND INTERIM CONVERSION FACTORS**

| | HC | CO | NOx |
|---|---|---|---|
| **Massachusetts I&M Program – Initial** | 1.50 | 0.86 | 0.86 |
| **Interim – based on AZ 341 Sample Dataset** | 0.98 | 0.57 | 0.56 |

There were three outcomes from this change:

1. MA31 test accuracy was improved;
2. emissions readings for vehicles were lower after the change; and
3. the emissions failure rate declined from a range of 7%-9% to 4% - 6%.

## Why Were the Conversion Factors So Different?

The initial software conversion factors accommodated only differences in the equipment. The drive trace differences were accommodated by adjustments to the initial cutpoints. As of April 2001, however, with the implementation of the final cutpoints, the cutpoints no longer included any correction for the drive trace. The interim conversion calculations accommodate both differences in equipment and differences in trace.

## Final Conversion Factors

Following completion of this study in August 2001, conversion factors for each pollutant were calculated for the 612 vehicle sample dataset. While the original study called for 1000 vehicles to be tested, a variety of factors, which are discussed in the technical report, led to a smaller sample size. That reduction was discussed and accepted by EPA. To improve the accuracy of emission calculations for vehicles emitting near the cutpoints, the analysis of the complete data set excluded vehicles with raw MA31 emissions more than 1.5 times each pollutant's maximum cutpoint, as was done with the interim conversion factors. Emissions were high enough to fail both tests.

Table 2 compares conversion factors developed from both the 612 and 341 vehicle datasets to the initial factors used in the program.

**TABLE 2**
**COMPARISON OF CONVERSION FACTORS**

| | HC | CO | NOx |
|---|---|---|---|
| **Massachusetts I&M Program Initial** | 1.50 | 0.86 | 0.86 |
| **Interim, Implemented July 2001** (based on AZ 341 Sample Dataset) | 0.98 | 0.57 | 0.56 |
| **Final** (based on AZ 612 Sample Dataset) | 0.87 | 0.53 | 0.60 |

As Table 2 shows, the final conversion factors developed from the 612 vehicle dataset are close to the interim conversion factors. The final conversion factors for HC and CO are slightly lower, meaning the interim conversion factors still cause emissions results reported by MA31 to be slightly higher than EPA IM240 standard grams per mile, although they represent a substantial improvement to test accuracy compared to the initial conversion factors.

For NOx, the final conversion factor is slightly higher than the interim conversion factor, meaning the interim conversion factor results in NOx emissions being slightly lower than EPA IM240 standard grams per mile. Similar to HC and CO, the interim NOx conversion factor still represents a substantial improvement in test accuracy compared to the initial conversion factor.

## When Will the Final Conversion Factors Be Implemented?

The final conversion factors will be implemented after final quality assurance of the data and the conversion calculations is complete and stakeholders are provided advance notice of these changes. Changing the conversion factors also affects other components of the program, such as the recently implemented dilution check.

In addition, other software upgrades have been in progress. Attempting too many simultaneous changes constitutes poor network management, increasing the potential for network disruption. Once the final conversion factors are verified and stakeholders consulted, DEP will implement the changes at a time that minimizes potential for network disruption.

## INTERIM TEST EFFECTIVENESS EVALUATION

### What Is Test Effectiveness?

"Test effectiveness" is a measure of how well a state's I&M test equipment and drive trace identify excess emissions from all the states' vehicles compared to EPA's "gold standard" test equipment and drive trace, the IM240.

Test effectiveness numbers are derived from studies that compare the state's equipment and drive trace under ideal, controlled conditions to the IM240 equipment and drive trace under ideal, controlled conditions.

These studies then calculate a test effectiveness percentage for each of the three tested pollutants: hydrocarbons (HC), oxides of Nitrogen (NOx), and carbon monoxide (CO). "100% test effectiveness" would match the test effectiveness of EPA's standard, the IM240 test and equipment.

This test effectiveness comparison is deliberately done under ideal conditions in order to separately analyze how well the equipment and the trace perform. In contrast to "test effectiveness", "program effectiveness" examines how well the equipment and trace perform as they are used in the field, and is usually analyzed as a part of a full program evaluation[3].

"Excess emissions" is the difference between how much a failed vehicle pollutes when tested and how much it pollutes after repairs are done. Typically, repairs on a failed vehicle will reduce emissions well below the cutpoints. That is because most vehicles in good repair will pass the test by a large margin, and most vehicles with malfunctioning emissions control systems will fail the test by a large margin. Repairs made before the test occurs are not counted as excess emissions, because the test did not identify them. In that case, the air is cleaner, but the I&M program does not get credit for that reduction.

When analyzing test effectiveness – how well the equipment and drive trace perform in controlled circumstances – EPA assumes vehicle pollution is reduced only to the cutpoints. When analyzing program effectiveness – how well the program works in the field – EPA analyzes data that show the actual reductions achieved from repairs.

Test effectiveness numbers are used by EPA and states as one of the many inputs in EPA's computerized models for estimating air pollution ("models"). There are dozens of different inputs that describe, for example, ambient monthly temperatures, average vehicle miles traveled, average age and profile of the vehicles in the state, and so on. These "models" are used to create a system of air pollution "credits" linked to reductions of pollution from air pollution sources (e.g., vehicles, power plants, other businesses). EPA and states then use these "models" to

[3] A program evaluation analyzes how well the entire program works in the field, including factors like in-use accuracy of the MASS99 equipment and software, amount of permitted variance in the driving trace, inspector's ability to match the driving trace, amount of pre-test repairs, the effect of improper test administration by the inspector, the effect of motorist non-compliance, et cetera.

establish state pollution reduction targets, and to develop estimates of "credits" that come from each I&M program.

EPA periodically changes its models to reflect greater knowledge about how vehicles pollute. For example, EPA's MOBILE 6 model, released in 2001, revised and lowered estimates of the benefit from I&M programs to reflect the fact that the emissions control systems of vehicles manufactured in the 1990s tended to perform better than those manufactured in the 1980s (data from 1980s vehicles was the basis for EPA's MOBILE 5 model). In other words, after the Massachusetts I&M program started, EPA lowered its estimate of the effectiveness of the program because vehicles were in general running cleaner. While the I&M program stayed the same, EPA's new model gives it less credit.

## What Test Effectiveness Percentages Did EPA initially assign to the MA 31 Test?

When EPA approved the Massachusetts I&M program, it included the assignment of test effectiveness percentages as follows:

- HC (hydrocarbons): target is 85% as effective as IM240
- CO (carbon monoxide): target is 87% as effective as IM240
- NOx (oxides of nitrogen): target is 85% as effective as IM240

In other words, EPA predicted that the shorter MA31 drive trace and less expensive equipment would, under ideal conditions, be less effective at identifying vehicles with broken emission control systems than EPA's gold standard IM240 trace and equipment.

EPA's initial calculation was based in part on New York data and in part on professional judgment. Since the New York data was from a small number of vehicles, and since the NY I&M test used the same equipment but a different drive trace than Massachusetts, EPA required Massachusetts to conduct a test effectiveness study to develop more accurate test effectiveness percentages for MA 31.

## How Was Test Effectiveness Analyzed?

DEP estimated excess emissions identified by the MA31 test and the IM240 test, compared those results, used that comparison to estimate how well MA31 performed in comparison to IM 240. Then DEP compared the resulting test effectiveness percentages for the MA31 test to Massachusetts' targets. The correlation factor study described earlier in this summary was also used as part of this Test Effectiveness work.

In order to do this, DEP obtained vehicle data from an actual working I&M program that used the IM240 test, and applied appropriate statistical methods to provide results more representative of Massachusetts vehicles, as follows:

- A random sample of IM240 tests from Arizona's ongoing I&M program was selected;
- The IM240 records from the Arizona vehicles were obtained;
- MA31 results for those vehicles were estimated based upon calculations derived from the 612 car dataset;
- a statistical method ("Monte Carlo") was used to simulate the varied kinds of test results found in operating I&M programs; and
- the excess emissions identified in the IM240 and MA 31 datasets were compared.

DEP believes that this analysis is sound and that it is sufficient for its intended purpose: to provide an estimate of MA31 test effectiveness and to suggest ways to improve effectiveness. Details about this methodology are contained in the "MA31 Conversion Factor Analysis and Interim Test Effectiveness Evaluation" (the Technical Report). On July 21, 2003, DEP submitted the Technical Report to EPA and requested that EPA review DEP's work and provide suggestions to improve test effectiveness. DEP will also discuss the Technical Report with I&M program stakeholders.

**What Are the Major Limitations of the Study?**

All studies have limitations. Both the 2%AZ sample and the 612 vehicle dataset varied somewhat from the actual Massachusetts fleet. For example, the distribution of vehicles in the Arizona fleet contained more older vehicles than are tested in Massachusetts. In addition, the 612 vehicle data set study was designed to obtain relatively equal numbers of vehicles across model years, vehicle types and rates of emissions. It was not designed to be representative of the actual distribution of Massachusetts vehicles by model year.

Appropriate statistical methods were used in an effort to make the results more representative. Since EPA's original selection of numbers for test effectiveness was an estimate, this study's results very likely yield more accurate test effectiveness figures.

A second limitation –differences in results seen when vehicles are tested multiple times on the same equipment – affects all I&M programs. This "test-to-test" variance comes from the vehicle being tested (vehicle variance) and from inherent limitations of the test.

A single motor vehicle tested multiple times, driven precisely the same way each time, on equipment that functioned perfectly each time, would yield varying pollution readings. Those differences appear to be caused by the vehicle itself producing different amounts of pollution, even when driven in exactly the same manner.

The test itself also introduces additional test to test variations. There is always some difference in how the driving trace is conducted; precise duplication is impossible with human drivers. In addition, the equipment that analyzes the emission gases has margins of error. For example, the less expensive chemi-luminescense equipment used to measure NOx in Massachusetts has a greater margin of error than the Flame Ionization Detector used in the more expensive IM240 equipment.

Add these variances together, and it is normal for I&M test results to vary for the same vehicle tested on the same equipment.

To offer an analogy, it's similar to taking one's pulse. If someone were to sit still in a chair and have their heart rate counted for one minute, they would get a specific pulse rate. Conducting that test many times would yield different results, since hearts sometimes beat slower and sometimes faster. Each result is accurate at the time it is taken; but each yields different results because the heart's pulse rate varies.

Now, instead of counting the pulse for one full minute, what if someone counted the number of beats for six seconds and multiplied the result by ten to get a pulse rate (in other words, use a shorter, faster, and more convenient test). This quicker test introduces even more variability into the measurement of an accurate pulse rate.

That kind of faster, more convenient pulse taking is common, because it counts well enough to identify major pulse rate problems (for example, an extremely fast heartbeat).

I&M tests only need to be good enough to identify major pollution problems: the dirty, broken vehicles that are polluting much more than they were designed to.

If a clean car, polluting a $1/10^{th}$ of the pass/fail points, has varied results from test to test, the vehicle is still clean and still passes. If a grossly polluting vehicle, emitting at 10 times the pass/fail points, has varied results from test to test, the vehicle is still dirty and still fails.

However, a vehicle that pollutes right at the pass/fail points is a different matter. That vehicle has a malfunctioning emission control system, and its own "vehicle variance" alone will cause it to sometimes pass and sometimes fail. Add in the variance from the test itself, and it's expected that some vehicles will be clean enough to pass one time, then dirty enough to fail the next.

This "vehicle variance" is common in every I&M program across the country, including EPA's benchmark IM240 test. That factor alone can cause the vehicles that pollute right at the cutpoints to fail an I&M test one day and pass the next.

When developing and implementing I&M programs, states balance the added cost to motorists and inspection stations of the laboratory grade equipment that has some variance with less expensive equipment that has somewhat more variance. Some states chose the relatively more accurate and more expensive equipment. New York, Rhode Island, Massachusetts and other states chose less expensive equipment.

## **What Were the Results of the Interim Test Effectiveness Analysis?**

The test effectiveness numbers predicted for the MA31 test and equipment after the correction of conversion factors in **July 2001** were:

- HC: 87% as effective as IM240 compared to the 85% target
- CO: 90% as effective as IM240 compared to the 87% target
- NOx: 69% as effective as IM240 compared to the 85% target.

In other words, the Massachusetts test was predicted to exceed its test effectiveness targets for HC and CO and to obtain about 8/10ths of the target for NOx.

## **What Were The Interim Report's Recommendations?**

For the short term, the recommendation was to delete "Fast Pass" from the MA31 test. "Fast Pass" was implemented to reduce the amount of test time for motorists and inspectors. It allowed a vehicle to pass if its emissions results were below the pass/fail points on any one of six times the 31 second test was given. In other words, if a vehicle with "true" emissions results above the pass/fail points dipped below those points on any one of up to six chances to pass – due to vehicle variance - the vehicle was passed. This lowered test effectiveness because it tended to allow more polluting cars to pass.

The analysis predicted that substituting "Fast Pass" (six chances to pass) with a more protective test (two chances to pass) would produce these test effectiveness results:

- HC: 91% as effective as IM240 compared to the 85% target
- CO: 93% as effective as IM240 compared to the 87% target
- NOx: 75% as effective as IM240 compared to the 85% target

In other words, by eliminating "Fast Pass" the Massachusetts test was predicted to exceed its targets for 2 out of 3 pollutants, and to obtain about 9/10ths of the target for NOx. This change was made in **March 2003**. Implementing the final change to the conversion factors will further improve test effectiveness for NOx and slightly reduce test effectiveness for HC and CO.

For the long term, the recommendation is to evaluate and select methods that will increase NOx effectiveness. These actions will increase the program's failure rate. Possibilities include:

- Switch from the short MA31 test to MA147 (a longer, higher speed test) for all vehicles that are tested on the dynamometer
- Use a hybrid: the shorter MA31, with a longer MA147 as a "second chance to pass"

- Lower pass/fail points for the MA31 test (requires a regulatory change)

**What Actions Were Taken?**

In **March 2003** DEP deleted fast pass in order to increase predicted test effectiveness for NOx from 69% of IM240 to 75% of IM240.

**What Are the Next Steps to Improve Test Effectiveness?**

DEP will discuss this Interim Effectiveness Report with EPA and other stakeholders to obtain their input and recommendations.

DEP will propose, discuss and implement improvements to the test designed to raise test effectiveness for NOx to the 85% target. While the schedule will be dependent upon the method selected, DEP intends to select a method in summer 2003 and implement the improvement before Spring 2004.

**Do the NOx Test Effectiveness Findings Have an Impact on Massachusetts' Ability to Meet its Clean Air Goals?**

Massachusetts does not attain the federal standards for ozone, and has established (and obtained EPA approval) of a State Implementation Plan (SIP) that commits to implementation of strategies for reducing ozone pollution and attaining the federal standards. NOx is a significant contributor to ozone pollution.

The major elements of Massachusetts' strategy for reducing NOx pollution are the summertime NOx cap and trade program for power plants (which provides about half of the needed reductions), the "Low Emission Vehicle" Program in which new vehicles purchased in Massachusetts are required to meet California's standards for clean engines (which provides about a quarter of the needed reductions), and the Enhanced Emission Test Program (which is expected to provide the remaining quarter of needed NOx reductions).

This study demonstrates that Massachusetts equipment and drive trace currently obtains about 90% of the test effectiveness for NOx that is expected from the IM test. The study also identified several possible modifications of the Massachusetts emissions test that would bring the test closer to the test effectiveness targets established by EPA in its approval of the Massachusetts' IM SIP.

Massachusetts will discuss with EPA the report findings, what effect the current NOx test effectiveness percentages may have on our pollution reduction plans, and what actions may be needed to meet the SIP requirements.

# I&M GLOSSARY

## Measuring Emissions

**“Conversion Factors” –** Adjustments made to convert raw emissions results from an emissions test to EPA IM240 standard gpm. Conversion factors compensate for equipment or a drive trace that is different than IM240. Most states with emissions test different than IM240 use software-based conversion factors to convert raw emissions results into EPA IM240 standard gpm. Using conversion factors to convert raw emissions readings to EPA IM240 standard grams per mile is similar to scoring standardized tests, like the MCAS and the SAT. In those tests, a raw score is obtained, the raw score is converted into a standardized score, and the standardized score is reported as the result of the test.

**“EPA IM240 standard gpm”** - Grams per mile of pollution emitted by a vehicle measured by IM240 equipment while running an IM240 drive trace. Almost all states with different equipment and drive traces scale (convert) the raw emissions to EPA standard IM240 gpm.

**“Drive Trace” –** The prescribed mix of acceleration, cruising, braking and speed that a vehicle is driven through when conducting an emissions test. Drive traces can be lower or higher speed, can be longer or shorter and can include variations in speed or constant speeds. Each different test has its own different drive trace.

**“SIP” –** The State Implementation Plan, or SIP, is a document a state submits to EPA which contains an inventory of the amount of pollution in a state, a description of the air pollution reductions needed to meet federal standards, and a list of the pollution reductions programs that a state will use to meet those federal standards. A SIP can describe a single program (e.g., I&M SIP) or describe a comprehensive plan to meet a federal standard (e.g., attainment SIP).

## Test Reliability

**"Errors of Commission" –** Occurs when a vehicle failed a specific test, but should have passed its reference test. The reference test for MA31 is the IM240 equipment and drive trace. Therefore, Errors of Commission in Massachusetts refers to the failure of a vehicle using the MA31 equipment and drive trace but should have passed if tested on the IM240 equipment and drive trace.

**"False failure"** - Occurs when a vehicle failed a specific test but should not have. Every test has false failures. In Massachusetts, a “false failure” occurs when a vehicle fails MA31, but should have passed. Contributing factors can include vehicle variance, how the test is performed, or the equipment not working as designed.

**"Vehicle variance" –** The tendency of a motor vehicle to produce different amounts of pollution even when driven in exactly the same way. In I&M tests, vehicle variance is one of the reasons why the same vehicle can yield different test results from multiple tests.

Vehicle variance of clean cars tends not to affect test results, because the range of pollution emitted in typically well below the cutpoints. Vehicle variance of grossly polluting cars also tends not to affect test results, because the range of variability is typically well above the cutpoints.

However, vehicle variance from cars that pollute around the cutpoints tends to influence test results: they can pass one day and fail the next, even if two tests are performed in exactly the same way.

(NB: Vehicles that pollute around the cutpoints and the grossly polluting vehicles both have malfunctioning emissions control systems that cause them to pollute much more than they were designed to. Repairs to these vehicles are the way that I&M programs reduce pollution).

## Test Effectiveness

**"Excess emissions"** - Pollution reductions that result from repairs of vehicles which fail an I&M test. Repairs made before the test are not counted as excess emissions, because the test did not identify them. As False Failures and Errors of Commission decrease, excess emissions also decrease, since 1) fewer vehicles are being failed and repaired; and 2) typically the vehicles identified in those categories show significant decreases in emissions after repairs are done. When analyzing test effectiveness – how well the equipment and drive trace perform in controlled circumstances – EPA assumes that vehicle pollution is reduced only to the cutpoints. When analyzing program effectiveness – how well the program works in the field – EPA analyzes data that show the actual reduction achieved from repairs.

**"Program effectiveness" -** Examines how well the equipment and trace perform as they are used in the field, and is usually analyzed as a part of a full program evaluation.

**"Test effectiveness"** – An EPA-required calculation expressed in percentages (e.g, 85%) that indicates how well a state's I&M test equipment and drive trace identify excess emissions from a vehicle fleet compared to EPA's "gold standard" test equipment and drive trace, the IM240. Analysis of test effectiveness is done under controlled conditions in order to isolate the performance of the equipment and the trace perform from the way they are used in an actual testing environment.

## Descriptions of I&M Tests

**Cutpoints –** The pass/fail level for I&M tests. Vehicles which pollute above the cutpoints at the time they are tested fail. Vehicles which pollute below the cutpoints at the time they are tested pass. Different cutpoints are assigned for different pollutants.

**IM240 test** - EPA's "gold standard" equipment and drive trace for I&M programs. The "240" refers to the 240 second-by-second measurement of a vehicle emissions taken over the 239 second test. IM240 was derived from a small portion of the FTP (Federal Test Procedure), a two-day comprehensive test used by the federal government to certify emission levels from new vehicles.

**IM240 start-up cutpoints** – EPA-assigned pass/fail levels that are more lenient than final cutpoints. The cutpoints vary based on the pollutant being measured, the classification of the motor vehicle (e.g., passenger car vs. light duty truck) and amount of pollution the vehicle was allowed to pollute when it was new (the age of the vehicle). The cutpoints are expressed in EPA IM240 standard gpm.

**IM 240 final cutpoints –** EPA assigned pass/fail levels that are stricter than start-up cutpoints. States can choose whether to use the EPA start-up or the EPA final cutpoints. The cutpoints vary based on the pollutant being measured, the classification of the motor vehicle (e.g., passenger car vs. light duty truck) and amount of pollution the vehicle was allowed to pollute when it was new (the age of the vehicle). The cutpoints are expressed in EPA standard IM240 gpm.

**MA31 test –** The Massachusetts I&M test, consisting of the MASS99 equipment and the MA 31 drive trace. The MA31 has less expensive, less accurate equipment and a different and shorter drive trace. These factors cause the MA test to be less accurate than its reference test, IM240

**MA31 start-up cutpoints -** DEP assigned pass/fail levels that were used when the Enhanced Emissions and Safety Test began in October 1999. These lenient cutpoints varied based on the pollutant being measured, the classification of the motor vehicle (e.g., passenger car vs. light duty truck) and amount of pollution the vehicle was allowed to pollute when it was new (the age of the vehicle). The cutpoints were expressed in EPA IM240 standard gpm.

**MA 31 final cutpoints –** DEP assigned pass/fail levels in use since April 2001. DEP lowered the program's cutpoints from start-up to final phase in a series of steps between October 1999 and April 2001. **MA31** final cutpoints are expressed in EPA IM240 standard gpms and are contained in the I&M program regulations (310 CMR 60.02). MA 31 final cutpoints match EPA's start-up cutpoints.