



# **The SCORE Pilot Project**

## **FINAL REPORT**

**January, 2004**

Hazardous Materials and  
Waste Management Division



---

**Colorado Department  
of Public Health  
and Environment**

---

## Table of Contents

Executive Summary.....	3
1.0 Introduction.....	4
2.0 Project Implementation.....	5
Task 1: Project Design.....	5
Task 2: Selection of Participants.....	7
Task 3: Collection of Self-Certifications from Participants.....	8
Task 4: Inspection of Participants by HMWMD Inspectors.....	8
Task 5: Data Evaluation.....	9
Question 1 Analysis.....	9
Question 2 Analysis.....	12
Question 3 Analysis.....	16
Other Analysis.....	17
3.0 Summary of Findings.....	19
Appendix 1: Correspondence to SCORE Participants	
Appendix 2: List of Environmental Business Practice Indicators (EBPIs)	
Appendix 3: SCORE Data	
Appendix 4: Statistical Analyses of SCORE Data	

## Executive Summary

The Hazardous Materials and Waste Management Division (HMWMD) has conducted a unique environmental compliance project called the SCORE (Self-Certification and Reporting) Pilot Project. This project was designed to answer one fundamental question:

### Is self-certification of hazardous waste compliance an effective, efficient, and accurate method of assessing compliance in the large and diverse universe of Small Quantity Generators in Colorado?

For this pilot project, 76 Small Quantity Generators (SQGs) of hazardous waste were asked to conduct a self-assessment and self-certification of their compliance with hazardous waste regulatory requirements. This self-assessment and certification was accomplished using a checklist of all applicable regulatory requirements specifically prepared for this project. After these self-certifications were received by the HMWMD, hazardous waste inspectors performed follow-up inspections using identical checklists. Data from the self-certifications and the follow-up inspections were compared and evaluated using methodology very similar to that used by Massachusetts in their Environmental Results Program, or ERP.

### Results

The HMWMD has concluded through an evaluation performed on the collected data that self-certification of hazardous waste compliance by SQGs is a very effective method of assessing compliance. This should allow for much more efficient and effective regulation of the SQG universe as compliance problems can be more precisely addressed; widely violated requirements can be targeted for compliance assistance ('who' needs 'what'), and any facilities with extensive violations can be returned to compliance through appropriate means.

Further, the HMWMD has concluded that self-certification is not burdensome to facilities. The checklist was completed by most facilities in less than one hour. In addition, completing the self-certification checklist re-familiarizes facility personnel with hazardous waste requirements, giving them a leg up on hazardous waste training which is a commonly violated regulatory requirement. Self-certification allows for the collection of vital information, such as compliance rates and waste generation rates, which is not now collected.

Because of the success of the SCORE Pilot Program, the HMWMD plans to approach the Colorado Hazardous Waste Commission with these results and propose that new regulations be promulgated requiring each SQG to self-certify their compliance every two years, in a manner similar to the Biennial Reporting System, or BRS, for Large Quantity Generators (LQGs).

## **The SCORE Pilot Project**

### **1.0 Introduction**

In 2001, the Hazardous Waste Compliance Unit of the Hazardous Materials and Waste Management Division (HMWMD) began a unique environmental compliance project known as the SCORE (Self-Certification and Reporting) Pilot Project. This project was designed to answer one fundamental question:

#### **Is self-certification of hazardous waste compliance an effective, efficient, and accurate method of assessing compliance a large and diverse universe of Small Quantity Generators?**

The SCORE Pilot Project was precipitated by 3 things:

- a. Because of a relatively low number of inspectors on staff (4.6 FTE), and the resulting low inspection frequency of the approximately 1100 notified SQGs in Colorado (~10% of SQG universe inspected each year), the SQG universe was not being effectively regulated. Compliance rates were not improving, and no general deterrence, and little specific deterrence, was occurring. Given our available resources, we realized that some alternative was necessary to effectively regulate the SQG universe.
- b. Colorado requires no periodic reporting of any kind by SQGs. Therefore, very little is known about the universe, such as an accurate numbers of SQGs, predominant industry sectors, waste generation rates, and common waste types. We wanted, and needed, to better understand the SQG universe. We also needed to acquire information from the SQG universe that would help us determine environmental outcomes for Colorado, such as the total generation rates for all hazardous wastes, for each hazardous waste, and for each industry sector. This could only be done by collecting information for SQGs similar to that collected by the Biennial Reporting System (BRS) currently in place for Large Quantity Generators (LQGs).
- c. Massachusetts had just completed the first iteration of their 'Environmental Results Program,' or ERP. This program focused on certain small business industry sectors and required self-certification of compliance with water, waste, and air requirements. The ERP has been shown to be very effective. We felt that we could compliment the Massachusetts ERP results if we could show positive results with a universe-wide (non-sector based), hazardous waste only (not multi-media), pilot project evaluated along the same lines as the Massachusetts ERP.

In addition to the fundamental question presented above, the HMWMD hoped that the SCORE Pilot Project would show that 1) self-certification can point out common compliance problems and allow more precise compliance assistance both in terms of 'what' and 'who'; and 2) self-certification can collect a substantial amount of very useful information efficiently and without being burdensome to industry. The results, presented

herein, are very encouraging. In the near future, these results will be taken to the Colorado Hazardous Waste Commission where regulatory changes institutionalizing an SQG self-certification program will be proposed.

## **2.0 Project Implementation**

The SCORE Pilot Project involved the following five principle tasks:

1. Project design
2. Selection of participants
3. Collection of self-certifications from participants
4. Inspection of participating facilities by HMWMD inspectors
5. Data evaluation

The following sections explain how each of these tasks was implemented.

### **Task 1: Project Design**

The HMWMD wanted the SCORE Pilot Project to have the following characteristics:

- a. Consistency and comparability to the Massachusetts ERP: Partly because we knew the ERP was successful, and partly because we wanted results that would compliment the ERP, we designed a program that was very similar to Massachusetts' program. To that end, we designed SCORE with a 2-part inspection format – the first inspection was performed by facility staff in order to complete a self-certification checklist and the second inspection was completed by HMWMD hazardous waste inspectors who completed an identical hazardous waste compliance checklist to the self-certification checklist. We then performed similar analysis on the collected data to that conducted by Massachusetts, which is presented later in the section covering Task 5: Data Evaluation.

We had also spoken with Massachusetts' staff and had learned about several improvements to the ERP that they were making for their second ERP iteration. We incorporated these improvements as well, which were: 1) for better comparability, performance of the follow-up inspection within 90 days of the original self-certification; and 2) completion of an identical compliance checklist by HMWMD inspectors to the self-certification checklist with direct equivalency of the questions on the checklists.

- b. A 12 to 14 month project duration: This was meant to be a limited duration project. Based on our inspection schedule, we knew we could complete the program within a year if it was well designed and implemented. We were confident that the pilot would be successful, so we wanted to aggressively implement the program and be able to quickly analyze the results.

- c. A statistically significant sample set: HMWMD determined the necessary sample-set size in two ways. First, our staff had been performing approximately 100 SQG inspections per year, so we knew that a 100-inspection program could be accomplished in the desired one-year timeframe. Second, sample-size calculations were performed using standard texts and specialized literature on survey analysis to determine the minimum sample-size necessary to characterize the entire SQG universe with adequate statistical power. This analysis revealed that a random sample size of 78 facilities would give us a 95% probability of correctly characterizing the true compliance rate within  $\pm 9\%$ . A sample size of 108 facilities would achieve the same 95% confidence level, but would reduce the margin of error to  $\pm 7\%$ . HMWMD determined, based on these results, that a sample size of between 75 and 100 facilities would be adequate for the SCORE Pilot Program.
- d. A 2-inspection format: As described previously, to understand the accuracy of the self-certifications, we knew that we needed to conduct follow-up inspections. Each facility's checklist answers could then be compared to the inspector's answers to determine the level of agreement. This agreement/non-agreement evaluation constitutes a large portion of the data analysis presented in the Data Evaluation section of this report.
- e. The 2 inspections should be conducted as close together as possible: Massachusetts had some trouble with data comparability between the self-certification evaluations and the follow-up site inspections due to large amounts of time passing between the evaluations. We realized that the more time that passed between the self-certification and the inspection, the more facility characteristics might change, including manufacturing processes, amounts of waste on-site, knowledgeable employees leaving the company, etc. This would have made the conditions at the facility on the follow-up inspection less comparable to those that existed at the time the facility performed the self-certification evaluation. We wanted to maximize comparability.
- f. Participating facilities would enjoy enforcement amnesty for the pilot period, subject to certain conditions: Enforcement amnesty for participants was important for two reasons. First, because the self-certification evaluation met all of the requirements for a self-audit under Colorado's Self-Audit Statute and because facilities were voluntarily participating, we were required to offer enforcement immunity. Second, and more importantly, we felt that offering enforcement immunity for participants was an important gesture by which we could encourage participation and ensure accurate results. The conditions under which enforcement immunity would not apply were: 1) failure of the facility to promptly return to compliance; 2) failure of the facility to cooperate with the HMWMD in investigating any self-disclosed violations; and 3) an imminent and substantial endangerment was presented by the discovered violations.

- g. The self-certification checklist and the inspector checklist were constructed with an identical set of questions: This allowed for direct comparison of the data collected from the facility and by the inspectors. In addition, to the maximum extent possible, the questions were structured such that a ‘yes’ response indicated compliance and a ‘no’ response indicated non-compliance.

The project was designed to answer the one fundamental question presented at the beginning of this report, but to do that, data analysis was broken down to answer the following 3 questions:

Question 1: For all facility and inspector responses to all checklist questions, considered in aggregate, is there agreement between how the facilities assessed their compliance and how the HMWMD inspectors assessed compliance?

Question 2: For each checklist question, is there agreement between how facilities assessed their compliance and how the HMWMD inspectors assessed their compliance?

Question 3: For each facility, is there agreement between how the facility assessed their compliance and how the HMWMD inspector assessed their compliance?

### **Task 2: Selection of Participating Facilities**

For reasons that we did not anticipate, selecting participants proved to one of the hardest and most time-consuming tasks of the project. As mentioned previously, we could not require a randomly selected sample set of SQGs to participate. We had to solicit ‘random volunteers’ – or volunteers that were as randomly selected as we could make them. To do this, we took the notified universe of ~1100 SQGs and broke them into 3 random groups. We did this by numbering an alphabetical listing of the SQGs 1,2,3,1,2,3,1,2,3, etc. We then chose the group comprising all facilities designated by a ‘1’ and called that Group 1.

All 400 facilities in Group 1 were sent a letter (Appendix 1) explaining the SCORE Pilot Project and the enforcement immunity for participants. This same letter provided an open invitation to any of them to participate, but clearly explained that we would only accept the first 100 to sign up. We anticipated that we would easily get 100 volunteers and would then need to turn facilities away. This was not the case. In the end, we were only able to get about 45 of the Group 1 facilities to participate, and this only after personally calling every facility on the list and asking them to sign up. 45 facilities were not enough for the SCORE Pilot Project, so we extended the invitation to all Group 2 facilities and, again, only about 40 facilities decided to participate even after we called each facility. This left us with a sample size of 85 facilities, which met our sample size criteria. A sample of 85 facilities is sufficient to estimate the true rate of compliance in the SQG universe with a 95% confidence and a margin of error between 8% and 9%.

It is acknowledged that the 85-facilities selected do not, technically speaking, constitute a statistically ‘random’ sample, which is an assumption built into the statistical analyses described in this report. We do not think we could have gotten a better sample, and believe that the facilities chosen for this pilot study are sufficiently representative of the universe of SQGs so that sampling bias can be ignored. It took enormous effort to put together the sample set we used. Why we had such a problem getting facilities to participate, we do not know for certain. Based on the feedback received during the phone calls made to each facility, it seemed that there were two commonly stated issues: 1) an attitude that ‘HMWMD’s priority on this project was not their priority’; and 2) participation would take too much of the facility’s staff time. In addition, it is possible that our commitment to give participants enforcement immunity was not trusted or not compelling.

### **Task 3: Collection of Self-Certifications from Participants**

After the 85 participants were determined, we sent them each a packet of information. The packet included a cover letter thanking them for their participation and explaining what the project entailed, a Self-Certification checklist with a signature block committing facility management to the accuracy of the submitted information, and a guidebook explaining the regulatory basis for each checklist question and how compliance with that regulatory requirement could be achieved. The cover letter stated that we wanted to receive the checklists back within 30 days. These documents are all presented in Appendix 1.

Two problems began to emerge. First, though most checklists were received back from the facilities promptly, there was a significant percentage which were not received on time. We began to call these facilities and follow-up. This was an unexpected and sizeable amount of work. Second, another problem was the relatively large percentage of SQGs that were improperly notified and really operating as Conditionally Exempt SQGs (the class of very small quantity generators of hazardous waste). We anticipated this problem because we had seen it during implementation of another program entitled “The COMPASS Project” the previous year. When we encountered a miss-notified facility, the facility had to be removed from the SCORE Pilot Project since most of the checklist questions would not apply to that facility.

In the end, we received 76 self-certifications back from the original 85 participants. This was at the very low end, but still met our required sample size criteria.

### **Task 4: Inspection of Participating facilities by HMWMD inspectors**

As we received self-certifications back from facilities, we began to schedule the follow-up inspections to be performed by HMWMD inspectors. All of the inspectors had been trained regarding conduct the inspections and what data needed to be collected. Most of the inspections were performed within 90 days of receiving the self-certification. In some cases, the 90 day requirement could not be met due to scheduling problems, travel distance, etc.



During each inspection, the inspector completed a checklist identical to the one completed by the facility. In addition, the inspector spent as much time as was necessary explaining noted compliance problems to the facility and what steps they should take to return to compliance. There were no cases where enforcement was needed due to immunity requirements not being met.

### **Task 5: Data Evaluation**

All data collected for the SCORE Pilot Project is presented in Appendix 3. All statistical evaluations performed on the SCORE data are reported in Appendix 4.

Please refer back to the fundamental question and the three data analysis questions presented in the section of this report explaining Task 1 on page 7.

The basic approach to evaluating the collected data was to reproduce the Massachusetts data evaluation and augment that with additional important analyses. The following sections describe analyses done to answer each of the key questions presented under Task 1.

### **Question 1 Analysis**

**Question 1:** For all facility and inspector responses to all checklist questions, considered in aggregate, is there agreement between how the facilities assessed their compliance and how the HMWMD inspectors assessed compliance?

*H<sub>0</sub>: (Null hypothesis) The frequency of compliance as determined by facilities is less than or equal to the frequency of compliance as determined by the inspectors*

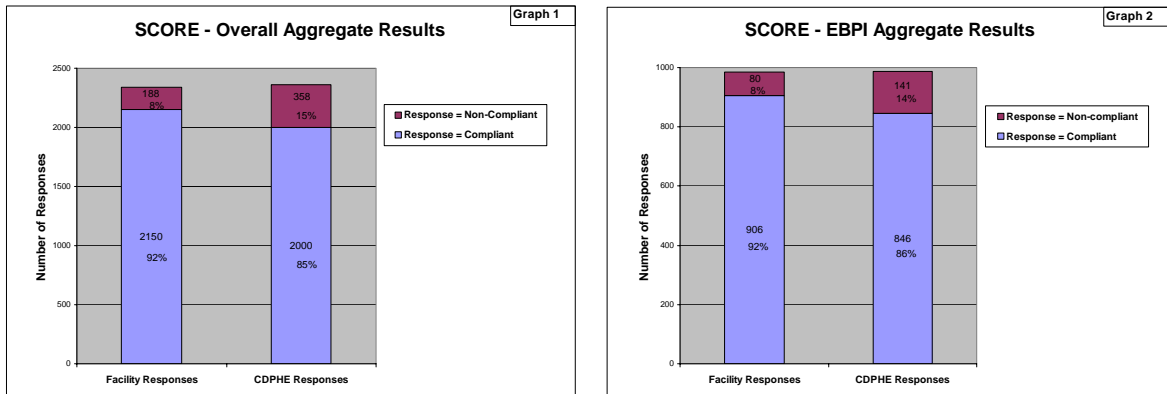
*H<sub>A</sub>: (Alternative hypothesis) The frequency of compliance as determined by facilities is greater than the frequency of compliance as determined by the inspectors*

The answer to Question 1 gives a strong indication as to whether self-certification is a viable tool to assess compliance in a large diverse under-inspected universe like the SQG universe.

To answer Question 1, the first evaluation done to the SCORE data was a simple aggregation of all facility responses and all HMWMD inspector responses. This data is presented below in Graphs 1 and 2. Graph 1 shows the overall total aggregate response comparison. Facilities reported that they were in compliance with 92% of all applicable requirements, HMWMD found a compliance rate of 85%. The statistical evaluation of this data showed that the frequency of compliance as reported by facilities is statistically higher than the frequency of compliance reported by HMWMD inspectors. This statistical analysis is discussed more completely below.

This simple data analysis is compelling in that it shows a remarkably small (7% delta) difference between the aggregate of facilities’ measurement of their compliance and the HMWMD’s measurement of compliance.

Graph 2 shows the same data presentation, but for a specific subset of checklist questions that we have designated key Environmental Business Practice Indicators (EBPIs). These questions were chosen specifically because non-compliance would indicate a business practice resulting in higher risk to public health and/or the environment. The complete list of EBPIs can be found in Appendix 2, but includes such items as obtaining



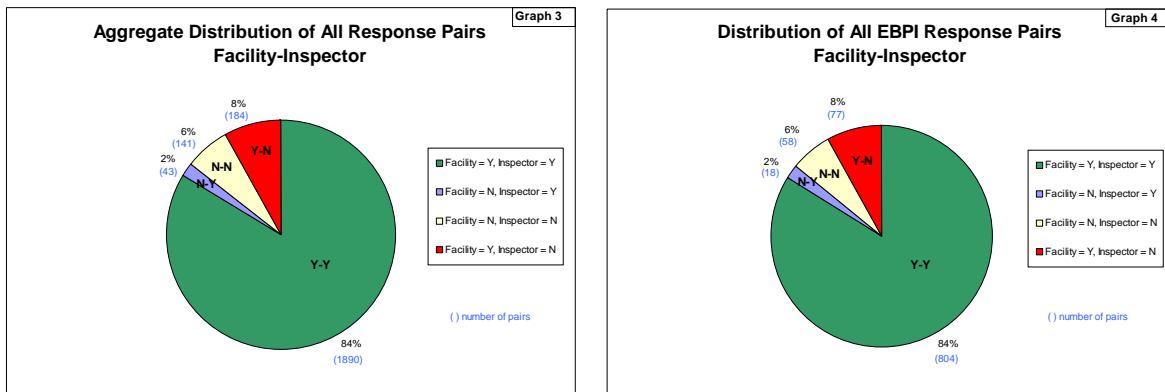
appropriate hazardous waste training for employees, not storing incompatible wastes together, ensuring that waste containers are kept closed, etc. The data in Graph 2 is remarkably consistent with Graph 1. Facilities reported that they were in compliance with 92% of EBPI requirements, HMWMD found a compliance rate of 86%. Again, this data shows a statistical difference between the facility data and the HMWMD data, but also shows a small (in this case, 6% delta) difference between facility and inspector results. Importantly, Graph 2 also demonstrates that the aggregated compliance rate for key EBPI indicators is equivalent, and no worse than, general compliance.

To completely answer the question of whether there was agreement between the facility and inspector responses, it is necessary to look at the ‘data pairs.’ Each question at each facility was first answered by the facility during their self-certification. Then each identical question was answered again by the HMWMD inspector during the site inspection. These 2 responses to each question are what we are calling ‘data pairs.’ Evaluation of the data pairs becomes the heart of the SCORE Pilot Project evaluation and we used it to answer Questions 2 and 3 as well. Data pairs with no response or a not applicable response were excluded such that only four possible data pair responses were considered:

- Facility/Inspector : Yes/Yes (Y/Y)
- Facility/Inspector : Yes/No (Y/N)
- Facility/Inspector : No/Yes (N/Y)
- Facility/Inspector : No/No (N/N)

Facility/Inspector Y/Y and N/N pairs are cases of obvious agreement. The Y/Y pairs are what we hoped to see because they indicate agreement on compliance. The N/N pairs also indicate agreement between the facility and the inspector, but indicate non-compliance. The N/Y pairs are interesting because they indicate that the facility thought they were not in compliance, but the inspector thought they were. These results are rare in the data set, but could indicate targets for better compliance assistance outreach efforts. The Y/N pairs are the most important and most problematic pairs because they indicate that the facility reported compliance, but the inspector found non-compliance. This could be due to 1) the facility purposely miss-reporting (which is what we tried to ascertain in Question 4), or 2) the facility not understanding compliance, which we can address through compliance assistance.

Graphs 3 and 4 present a pie-chart aggregate distribution for the various pair types – Graph 3 presents all data and Graph 4 is only EBPI data. Clearly, the vast majority of pairs in both cases are the Y/Y pairs, which indicate agreement on compliance and high overall compliance rates. When the Y/Y and N/N pairs are added together, since both show agreement, there is a 92% agreement rate between facility responses and HMWMD inspector responses. Again, we think this result is very compelling. Both graphs show an identical distribution of the pair types, again indicating the EBPI question responses are not indicating an inordinately higher risk to health and the environment from increased violation rates. Both graphs show that HMWMD found a 14% total non-compliance rate (N/N + Y/N), while the facilities reported an 8% non-compliance rate (N/N only,



excluding N/Y as assumed facility error).

**Statistical Analysis:** A statistical evaluation of this data was performed comparing facility responses with inspector responses. This evaluation was based on comparing the frequency or proportion of compliant responses reported by the facilities versus the inspectors (i.e., the sum of all “Yes” responses for all questions for both the facility and inspector populations divided by the total number of responses). The result of the statistical analysis (Table X of Appendix 4) indicates that there is less than 0.1% confidence that the null hypothesis is true (the frequency of compliance as determined by facilities would be less than or equal to the frequency of compliance as determined by the inspectors) and a 99.9% confidence that the alternative hypothesis is true (the frequency of compliance as determined by facilities is greater than the frequency of compliance as

determined by the inspectors). Therefore, we reject the null hypothesis in favor of the alternative hypothesis. In other words, we can be 99.9% confident that, when all facilities are considered in aggregate, facilities will slightly over-estimate and over-report their compliance. The details of this analysis can be found in Appendix 4.

In many ways, this result should be expected. How the facilities completed the checklist may contribute to over-reporting of compliance: 1) facilities ignorant of the regulatory requirements assume they are in compliance, but are not; 2) facilities more knowledgeable about the requirements have taken certain actions to comply, but these actions are insufficient; 3) facilities are not in compliance, but are purposely reporting compliance, and/or 4) facilities return to compliance as the checklist is completed, but were not in compliance before completing the checklist. Statistically, as the data for all checklist questions and all facilities is aggregated, it becomes a very large dataset; as the data set becomes large, the statistical tests become very sensitive, indicating differences as statistically significant that are, in fact, very minor. When the SCORE Program is rolled out on a broader scale, these factors will have to be further studied.

In summary, based on this evaluation, while these results show that facilities tend to over-estimate their compliance in a statistically significant manner, the overwhelming level of agreement between the response sets has caused us to determine that the answer to Question 1 is **YES! For all facility and inspector responses to all indicators, there is general agreement between how the facilities assessed their compliance and how the HMWMD inspectors assessed compliance.**

### **Question 2 Analysis**

Question 2: For each compliance indicator, is there agreement between how facilities assessed their compliance and how the HMWMD inspectors assessed their compliance?

*H<sub>0</sub>: (Null hypothesis) The frequency of compliance as determined by facilities is less than or equal to the frequency of compliance as determined by the inspectors*

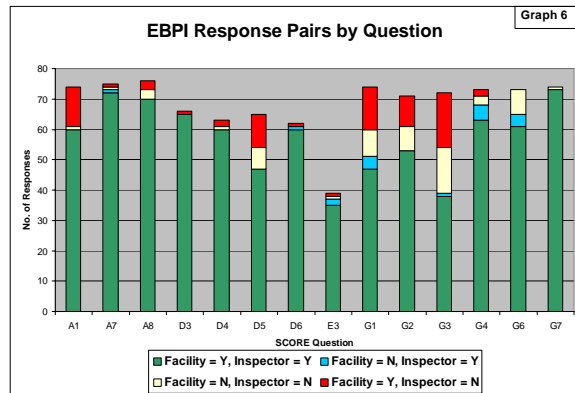
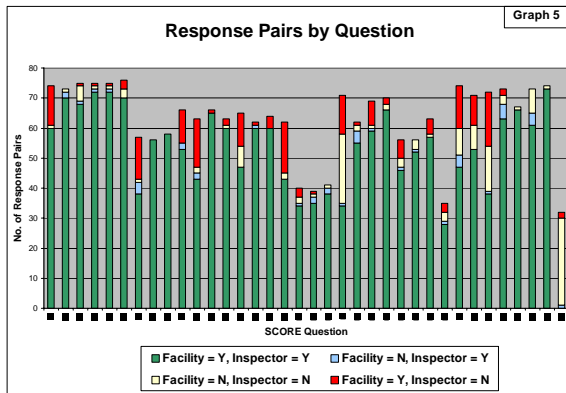
*H<sub>A</sub>: (Alternative hypothesis) The frequency of compliance as determined by facilities is greater than the frequency of compliance as determined by the inspectors*

Question 2 is important because the answer to this question could indicate several things. First, when there is agreement within data pairs, it indicates a requirement that is well understood. In the Y/Y case, it also indicates a high compliance rate. Second, when there is disagreement within the data pairs, it could indicate a bad checklist question that was poorly understood or difficult to interpret. This would be important to understand for subsequent self-certification and checklist-driven initiatives. More importantly, disagreement may indicate compliance requirements that are poorly understood and not well implemented by the regulated universe. Understanding where these problems lie heavily affects delivery of compliance assistance efforts – particularly ‘what’ is

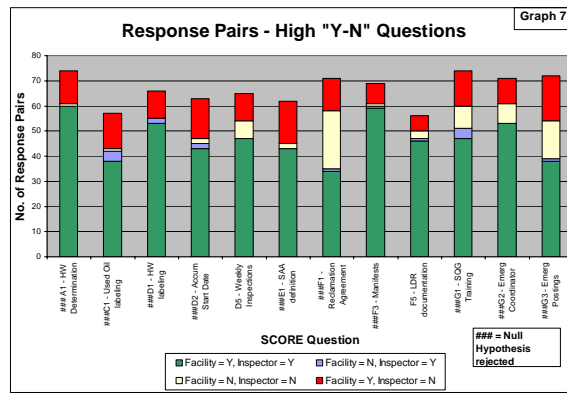
delivered. Lastly, response pair disagreement could indicate dishonest responses from a facility indicating compliance when non-compliance was actually present.

Question 2 requires an evaluation of pairs for each indicator, or checklist question. Some of the checklist questions were informational and not measures of compliance. For this evaluation, the informational questions have been excluded.

Graphs 5, 6, and 7 present bar-chart distributions of data pair types using the same color scheme as Graphs 3 and 4 presented previously. Graph 5 presents the results of all the compliance-related checklist questions, Graph 6 presents the EBPI checklist questions, and Graph 7 presents checklist questions with relatively high levels of Y/N data pairs.



These 3 graphs show that all questions have a high rate of agreement: high Y/Y and/or a high rate of occurrence. This requirements are well compliance facilities is very good. are certain checklist 7) which show a discrepancy between assessed their compliance and how the HMWMD inspector assessed compliance.



that all questions response pair pair occurrence Y/Y + N/N pair indicates that the understood and assessment by the However, there questions (Graph significant how the facilities

**Statistical Analysis:** Table 1, presented below, shows a summary of the statistical evaluation of checklist question data. The results reported here were based on testing the null hypothesis (*the frequency of compliance as determined by facilities is less than or equal to the frequency of compliance as determined by the inspectors*). From the table, only 11 of the 52 questions had facility compliance greater than inspector compliance at a level where the null hypothesis had to be rejected. Therefore, for 41 of the 52 questions (79% of the questions), the null hypothesis was accepted. In other words, for 41 questions, the difference between the facility responses and the inspector responses was not different enough to say that the facilities were over-reporting compliance. Details of this analysis can be found in Table Y of Appendix 4.

**TABLE 1:**

Statistical Outcome	Question 2		
	Number of checklist questions	Checklist Questions	% of questions
Facility Compliance > Inspector Compliance = <i>Statistically Significant (H<sub>0</sub> rejected)</i>	11	A1, C1, D1, D2, E1, F1, F3, G1, G2, G3, H2	19%
Facility Compliance > Inspector Compliance = <i>Not Statistically Significant (H<sub>0</sub> not rejected)</i>	17		35%
Facility Compliance = Inspector Compliance ( <i>H<sub>0</sub> not rejected</i> )	11		21%
Facility Compliance < Inspector Compliance ( <i>H<sub>0</sub> not rejected</i> )	13		25%
Totals	52		100%

Why do some of the checklist questions have significant differences between responses, but most do not? It is interesting to compare the list of questions on the first row of Table 1 to Graph 7. Graph 7 includes 10 of the 11 questions for which the null hypothesis was rejected (see footnote in bottom right corner of Graph 7) and describes the type of compliance requirement involved. It is striking which checklist questions show the significant differences. When compared to other compliance data evaluations performed by HMWMD outside of the SCORE Project, the checklist questions with significant difference show remarkable correlation to the most commonly violated SQG regulatory requirements.

Table 2 (next page) compares the list of most commonly violated SQG requirements to the list of SCORE checklist questions that showed significant difference.

**TABLE 2:**

<b>Regulatory Violation</b>	<b>Regulation with historically high non-compliance in SQG universe**</b>	<b>Regulation with high Y/N data pair frequency in SCORE Pilot Project^^</b>
inadequate container labeling	1	√ (null hypothesis rejected)
inadequate or no training	2	√ (null hypothesis rejected)
inadequate hazardous waste determination	3	√ (null hypothesis rejected)
inadequate management of used oil	4	√ (null hypothesis rejected)
<b>Regulatory Violation</b>	<b>Regulation with historically high non-compliance in SQG universe**</b>	<b>Regulation with high Y/N data pair frequency in SCORE Pilot Project^^</b>
containers left open	5	
inadequate emergency preparedness	6	√ (null hypothesis rejected)
failure to conduct weekly container inspections	7	√
poor recordkeeping	8	

\*\* Presented by rank: 1 = most commonly violated regulation; 2 = second-most commonly violated regulation; etc.

^^ Refer to Graph 7 and Table 1

We believe this result to be extremely significant because, in certain cases, the poor compliance rate combined with the high Y/N pair occurrence may indicate that these regulatory requirements are poorly understood (making an adequate hazardous waste determination, training, emergency preparedness). In other cases, this result may indicate that rapidly changing conditions at a facility, such as drum accumulation, drum shipments, staffing changes, etc., require constant facility attention to remain in compliance (container labeling, weekly inspections). In still other cases, it probably indicates simple facility oversight and lack of attention (open containers, recordkeeping). Most importantly, however, this information indicates that facilities are not lying or purposely reporting false compliance. If that was occurring, we do not believe there would be such close alignment between requirements that have very high historical non-compliance and SCORE questions with high Y/N pair occurrence. We also believe we would be seeing more questions for which the null hypothesis needed to be rejected.

Based on this evaluation, we have determined that the answer to Question 2 is **YES! For most compliance indicators, there is agreement between how facilities assessed their**

compliance and how the HMWMD inspectors assessed compliance. **HOWEVER, there are some compliance indicators which show a statistically significant difference between facility-assessed and inspector-assessed compliance, but this is due to a lack of knowledge by the facilities, not due to false reporting.**

**Question 3 Analysis**

Question 3: For each facility, is there agreement between how the facility assessed their compliance and how the HMWMD inspector assessed their compliance?

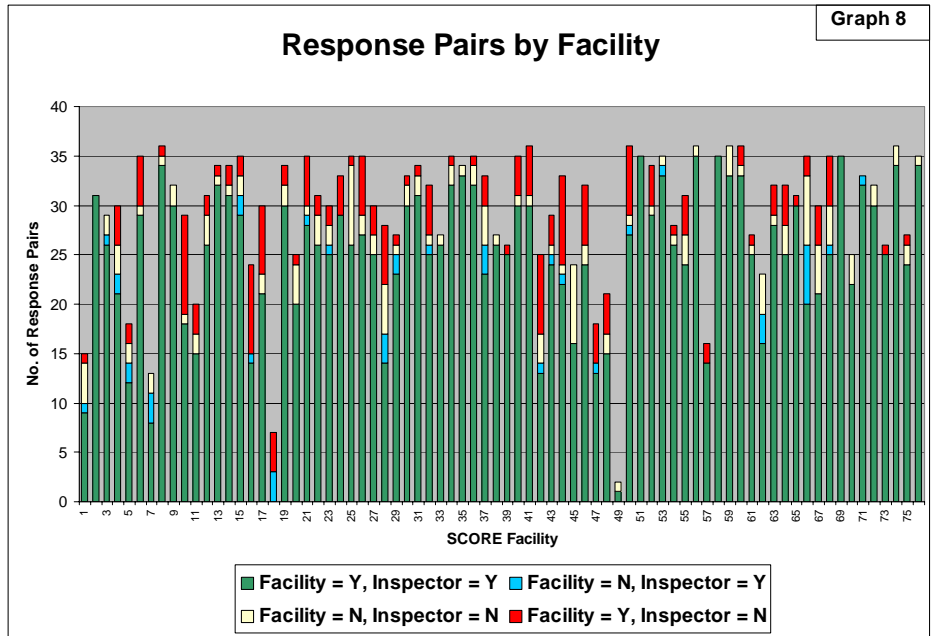
*H<sub>0</sub>: (Null hypothesis) The frequency of compliance as determined by facilities is less than or equal to the frequency of compliance as determined by the inspectors*

*H<sub>A</sub>: (Alternative hypothesis) The frequency of compliance as determined by facilities is greater than the frequency of compliance as determined by the inspectors*

Question 3 is important because the answer indicates who is, and who is not, having trouble with compliance. It could also indicate who might be falsely reporting compliance. Are the facilities that are having problems widely distributed within the SQG universe, or are they concentrated in certain industry sectors? This again drastically affects delivery of compliance assistance efforts – particularly ‘who’ is targeted.

Question 3 requires an evaluation of pairs for each facility. Again for this evaluation, the informational questions have been excluded.

Graph 8 presents a bar-chart distribution of all data pairs for each facility using the same color code as previous graphs.





The facilities showing a significant difference between their assessment of compliance and the inspector’s assessment are not concentrated in any particular sector, geographic region, or by any other characteristic we could determine.

Statistical Analysis: Table 3, presented below, shows a summary of the statistical evaluation of facility-specific data. The results reported here were based on testing the null hypothesis (*the frequency of compliance as determined by facilities is less than or equal to the frequency of compliance as determined by the inspectors*). From the table, only 3 of the 76 facilities reported compliance greater than inspector compliance at a level where the null hypothesis had to be rejected. Therefore, for 73 of the 76 facilities (96% of the questions), the null hypothesis was accepted. In other words, for 73 facilities, the difference between the facility responses and the inspector responses was not different enough to say that the facilities were over-reporting compliance. This is a very high rate of correlation. It again indicates that facilities were good at assessing and reporting their compliance relative to how the inspectors assessed compliance. This further supports the conclusion that facilities were not falsely reporting compliance. Details of this analysis can be found in Table Z of Appendix 4.

**TABLE 3:**

Statistical Outcome	Question 3	
	Number of facilities	% of Facilities
Facility Compliance > Inspector Compliance = <i>Statistically Significant</i>	3	4%
Facility Compliance > Inspector Compliance = <i>Not Statistically Significant</i>	47	62%
Facility Compliance = Inspector Compliance	16	21%
Facility Compliance < Inspector Compliance	10	13%
Totals	76	100%

Based on this evaluation, we have determined that the answer to Question 3 is **YES! For most facilities, there is agreement between how facilities assessed their compliance and how the HMWMD inspectors assessed compliance. HOWEVER, there are a few facilities which show a statistically significant difference between facility-assessed and inspector-assessed compliance.**

**Other data evaluations:**

It was of interest to HMWMD to evaluate whether there was any evidence to suggest that facilities deliberately gave incorrect or fraudulent responses in the self-certifications. This is important because indications of purposefully incorrect responses could

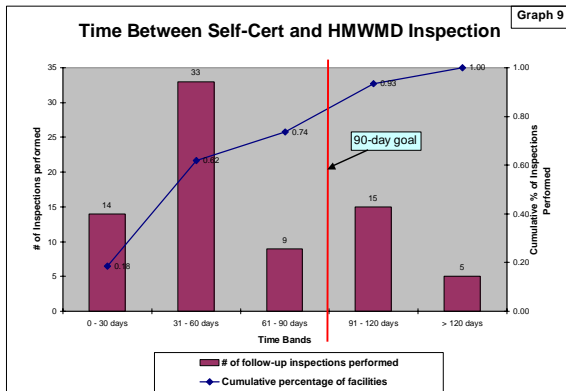
compromise self-certification as a useful tool to assess compliance in a large under-inspected universe.

The data pair of importance for addressing this issue is the Y/N pair, where the facility reported compliance but the inspector reported non-compliance.

Based on our evaluations, there is no evidence to support, or cause suspicions of, facilities purposefully reporting compliance when they were not in compliance. Rather, we believe, the data indicates that Y/N pairs were caused by facilities either not knowing the regulatory requirements or not understanding how to physically operate in compliance. This conclusion is particularly supported by the evaluation presented under Question 2 where the checklist questions with the highest percentage of Y/N pairs were also the regulatory requirements that are most frequently violated by SQGs across Colorado.

While we believe the SCORE Pilot Project data shows no sign of purposefully incorrect reporting of compliance, this may be due to the ‘voluntary’ aspects of pilot project participation. As self-certification is rolled out to the entire SQG universe as a requirement, vigilance for falsely reported compliance should be maintained.

Timely follow-up inspections:



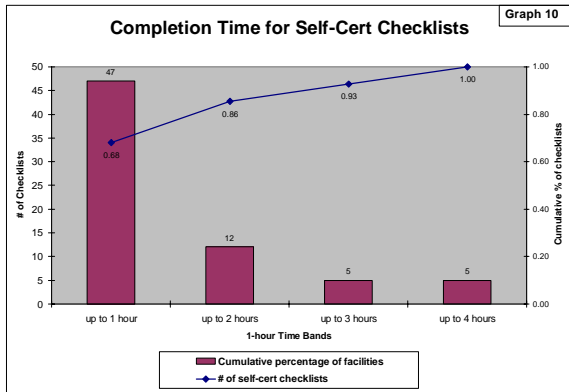
As explained earlier in the report, we endeavored to conduct the follow-up inspections as soon as possible after the facility returned their completed self-certification checklist. Our goal was to accomplish the inspections within 90 days of self-certification receipt.

Graph 9 presents the timeliness results. This graph shows that 56 out of the 76 SCORE facilities, or 74%, were inspected

within the 90-day goal. Another 15, or a cumulative percentage of 93%, were inspected within 120 days. While this is somewhat disappointing, HMWMD inspectors did the best that could have been done. Delays were usually caused by scheduling difficulties or travel distance to the facility. There is no indication that time delays caused any compromised data.

Reporting burden on participating facilities:

In implementing the SCORE Pilot Project, one of the things we wanted to understand was whether self-certification turned out to be an excessive burden on the participants in terms of staff time and cost. To get a handle on this aspect of the project, HMWMD inspectors asked each facility how long the self-certification checklist took to complete. Graph 10 presents this information.



Out of 76 participating facilities, 69 reported how much time it took to complete the checklist. Graph 10 shows that 68% of reporting facilities completed the checklist in 1 hour or less; 86% of reporting facilities completed the checklist in 2 hours or less, and no facilities took longer than 4 hours. We believe this data clearly shows that completing the checklist was not an undue burden to facilities. This is particularly true when the benefits are

considered: 1) staff that completes the checklist is re-trained and re-acquainted with the regulatory requirements through checklist completion; 2) compliance problems are identified so that they can be fixed; and 3) the liability of being found out of compliance is lessened.

### General feedback from participants:

The SQGs that participated in the SCORE Pilot Project really liked the program. We got a lot of very positive feedback. Now, participating in the SCORE Pilot Project was voluntary and protected participants from enforcement on non-compliance and this may well have contributed to the enthusiasm. However, we believe that the feedback was valuable. Generally, comments made to the inspectors fell into the following categories:

- a. checklist provided great training tool and material for staff;
- b. checklist provided great reminder and refresher of requirements;
- c. the on-site visit and face-to-face feedback was very helpful in getting specific questions answered; and
- d. facilities commented that they upgraded some aspect of their hazardous waste management as a result of SCORE participation.

### 3.0 Summary of Findings

1. Self-certification works. **Self-certification of hazardous waste compliance by SQGs is a very effective and accurate method of assessing their compliance.** There is a high degree of correlation between facility assessment and inspector assessment of compliance. This is true at the universe-wide level, the facility-specific level, and the regulation-specific level. This does not mean the self-certification can supplant inspections. In fact, inspections must continue to deter fraudulent reporting and to verify compliance rates within the SQG universe.
2. Self-certification is not burdensome to facilities. Our evaluation shows that facilities did not spend an inordinate amount of time completing the self-certification. In fact, many facilities appreciated the reminder the checklist

provided regarding many lesser known, but applicable, regulatory requirements.

3. Self-certification re-familiarizes facility personnel with hazardous waste requirements and would at least partially satisfy the regulatory requirement for ensuring that all employees are thoroughly familiar with waste handling and emergency procedures. The training requirement for SQGs is one of the most difficult for us to enforce and one of the most difficult for facilities against which to ascertain their compliance. The reason for this is because this regulation is a performance-based requirement. If a facility is in operational compliance, they are assumed to be adequately trained; if their operations are not in compliance, they are assumed to also be out of compliance with training requirements. Based on feedback we received from SCORE participants, it is our belief that, for a facility with staff that are initially well-trained, periodically completing the self-certification checklist will probably be adequate to remain operationally compliant.
4. Self-certification collects vital information not currently collected. Receiving regular data from SQGs is, and will continue to be, very important to the Hazardous Waste Compliance Program for:
  - a. Targeting compliance assistance – what we deliver; and who we deliver it to;
  - b. Performance measurement –
    - i. compliance rates for SQG facilities and industry sectors;
    - ii. compliance rates for individual SQG regulatory requirements; and
    - iii. measurement of waste minimization and waste reduction; and
  - c. Continued regulatory program improvement
    - i. finding and fixing the real problems; and
    - ii. effectively regulating each universe through improving compliance and providing effective deterrence.

(Appendices available separately - email [comments.hmwmd@state.co.us](mailto:comments.hmwmd@state.co.us))