Statistical Support Memorandum regarding an ERP
for the Auto Salvage Yard Sector in Orange County, FL

February 15, 2002

Prepared by Michael Crow & Rachel Cleetus of Tellus Institute
Funded by USEPA, OPEI
Contact: Michael Crow, Tellus Institute ▪ Tel: (617) 266-5400 ▪ Fax: (617) 266-8303 ▪ mcrow@tellus.org

Introduction

The term "Environmental Results Program" is used to describe a collection of integrated compliance assurance tools used in conjunction to improve environmental performance in an industrial sector. These include the use of self-certification, random inspections and educational outreach. An integral part of an Environmental Results Program (ERP) is the use of statistical techniques to measure facility performance is because it allows agencies to better understand the effectiveness of ERP in generating real environmental results. Such an approach to performance measurement allows an agency to monitor environmental performance of a large group of facilities (e.g., an entire sector) based on collection of data at a smaller, statistically valid sample of facilities. An initial statistical analysis is typically conducted after an agency has conducted a baseline round of inspections at randomly selected facilities, implemented the compliance assistance/self-certification approach throughout the target universe, and conducted a follow-up round of performance inspections at another set of randomly selected facilities. The statistical analysis usually has two primary purposes: to gauge changes in facility performance over time and to gauge the accuracy and reliability of self-certification forms.[1]

Implementing an ERP in the Auto Salvage Yard Sector of Orange County, Florida:

In 1998, using data obtained from the County Occupational Licensing and the local Yellow Pages, the Florida DEP identified a universe of 21 facilities as potential targets for a compliance assurance initiative targeted at auto salvage yards in the geographic region located between Toll Road 417 in Eastern Orlando and the Orange County Line. This area is entirely within Orange County. Five of the facilities were out of business when visited, leaving a population of 16 facilities. One of the 16 was excluded on the basis that (1) it had been inspected in the prior year by waste management, and, more importantly, (2) while the facility does conduct auto salvage operations, is classified under a different SIC code because it undertakes significant additional scrap metal activities. This latter reason justifies the exclusion of the facility, assuming that no such similar facilities were included previously or will be included in the future targeted population.

DEP conducted full inspections under EPA Compliance Assistance Site Visit criteria at only the remaining 15.[2] These Round 1 inspections uncovered high levels of non-compliance. After DEP followed up with compliance assistance and enforcement action, DEP proceeded to inspect those 15 facilities again (hereafter called Round 2). This follow-up inspection was conducted no more than 120 days after the initial inspection.

---

[1] Unlike in the Massachusetts ERP program, FL will implement a voluntary self-certification program. This fact influences the statistical approach, as described in later text.

[2] DEP believes strongly that these are the only 15 relevant facilities in the geographic area in question.

During these inspections, data on administrative and environmental issues were collected and significant patterns in facility behavior regarding violations and compliance emerged. (Please refer to FL DEP documents for further detail on the results of this study.) The issue at hand is that the FL DEP would now like to expand its compliance assurance efforts to target an ERP-like approach on the entire population of auto salvage facilities in Orange County (consisting of approximately 100 facilities), and the determination must be made if and how the original data collected can serve as a baseline sample.

It is our understanding that the expanded ERP will be conducted the following fashion. For the new study of the larger population (all of Orange County), all facilities, including the original subpopulation, will be offered compliance assistance and the opportunity to voluntarily fill out a self-certification questionnaire, which will contain the same questions as that used by on-site inspectors. Out-of-compliance facilities will be given an opportunity to take return-to-compliance action. Within a 3-4 month period of the deadline for self-certification, a follow-up inspection of a random sample of the facilities will be undertaken (this will be referred to as Round 3). Data collected from this sample will then be compared with Round 1 data (i.e., the data collected from the initial visit to the 21 facilities in Eastern Orange County).

Hypothetically (assuming Florida DEP determines that the sample meets the criteria for acceptability detailed throughout the rest of this memo), this Round 1 data would serve as the "baseline," allowing DEP to understand the combined impact of the initial two rounds of inspections and the self-certification/compliance assistance process on the entire population of auto salvage facilities in Orange County.

It is preferable not to use the Round 2 data as a baseline because such data reflects the intervention of a prior, recent inspection. It would be unwise to assume that, at the point in time at which Round 2 data was collected, the subpopulation performance could be considered a representative baseline sample for the rest of the population, which has not been inspected.

It is very important point out that, even though the self-certification process will be voluntary, DEP should draw its Round 3 sample from among all 100 facilities. If it were to sample from only among those facilities that self-certified, the sample is likely to be biased. It is also important to note that this voluntary approach will not allow DEP to draw conclusions from the self-certification data about the entire population. Using the self-certification data, DEP will only be able to draw conclusions about those facilities that chose to self-certify. Nonetheless, even though its measurement impact is reduced, the self-certification process may still have a positive effect on the performance of facilities that choose to use it.

Purpose of Tellus Statistical Support to Florida DEP:
DEP and Tellus agreed that Tellus would examine whether the original pilot project's data (collected from 15 of 15 facilities east of the Greenway Toll Road, a.k.a. Toll Road 417) could be used as the environmental baseline data for the application of an ERP-like approach in among the entire population of auto salvage facilities in Orange County. The analysis was intended to address the following specific issues:
1) Whether a baseline sample consisting of the original set of facilities (i.e., those facilities east of the Greenway Toll Road) presents a selection bias problem -- in other words, whether those facilities can be considered to be a random sample of the population of facilities in all of Orange County. The primary intent of the analysis will be to lay out criteria upon which Florida DEP can evaluate whether a selection bias problem may exist.
2) Whether the number of facilities from which data were collected in Round 1 allow for statistical inferences with a high confidence level. (In addition, Tellus briefly examined issues related to what was originally Objective #2 in the proposal, regarding confidence level/interval options and sample

size options for Round 3. Tellus did so because the analysis of Question #2 in this list would be incomplete without at least mentioning the ability to compare Round 3 and Round 1.)

3) Whether any data previously collected present obvious problems for making statistical inferences (e.g., through problematic phrasing of questions).

Keeping these objectives in mind, this memorandum is intended to help FL DEP better understand statistical issues presented by conducting and measuring the results of an Orange County-wide, ERP-like program for auto salvage yards, as a follow-up to a more targeted auto salvage yard enforcement project. This memo has been written with the expectation that Florida DEP is familiar with the statistical issues associated with ERP, as presented EPA's draft "Generic ERP Statistical Methodology."

To address the questions above, Tellus consulted background materials provided by the FL DEP and also conducted a conference call with FL DEP personnel.[3,4] All analysis is based upon the written and oral information provided by FL DEP. Tellus did not have resources available to conduct independent investigation or verification of this information.

Conceptual Approach to the Issues

If the Round 1 data is to serve as a baseline sample, it is crucial that it satisfy the requirement of still being considered a random sample, even with the new enlarged population. From a statistical standpoint, the typical **protocol for choosing a good random sample** in advance is clear:

1. **Compile a complete list of facilities** – The first step of generating the sample is ensuring that reasonable efforts have been made to generate a complete, accurate and up-to-date list of all facilities in a carefully delineated population.

2. **Decide on desired statistical properties –** Decisions regarding desired confidence level, confidence interval and sample stratification should be made. More detail on all these issues is contained in the draft U.S. EPA "Generic ERP Statistical Methodology."

3. **Draw a random sample** – The population list (and subgroups if necessary) is then completely randomized and a sample of the appropriate size for the desired confidence level and confidence interval is then drawn from it.[5] Alternatively, each facility could be assigned a number and then the sample could be drawn using some random number generating protocol or a random number table. This procedure ensures that each facility would have an equal probability of being chosen and thus the resulting sample will not contain any bias in terms of facility characteristics.

4. **Ensure a minimum sample size** – To achieve a high level of statistical significance, a random sample of at least 30 facilities is typically required. Very small sample sizes do not allow one to carry out standard hypothesis tests. Samples of less than 30 may be taken, but statistical confidence will suffer.

With these matters in mind, we can proceed to answering the questions presented in the proposal.

Question 1

Does this sample of 15 facilities differ in any systematic way from the entire population of facilities in Orange County? That is, can this sample of facilities be expected to present different levels of environmental performance from what would be observed if one had sampled, in 1998, 15 facilities randomly from among the total population of 100? If the subpopulation and the overall population differ significantly in any way, this would be a major impediment to assuming Round 1 as a baseline.

---

[3] Materials Provided by Lu Burson of FL DEP, in an e-mail of December 10, 2001. Materials include the following: "Environmental Results for the Auto Salvage Sector: a Central District Pilot Project," "Team Scrappers," "Scrappers FINALCHART Legend," "Possible Action Steps for Other Yards" (Graphic), "Salvage Yards Multimedia Inspection Worksheet," and the summary inspection data sheet for Round 1 and Round 2.

[4] Conference call with Lu Burson, Steve Adams and Dave Herbster on January 25, 2002.

[5] If the agency is stratifying its sample, it should generate randomized lists of facilities within each stratification category.

We asked FL DEP about several kinds of facility characteristics that would appear to be indicators of expected differences in environmental performance. These include facility size, trade association membership, terrain differences, and presence of a stormwater permit. A discussion of each of these characteristics is presented below, based upon information provided by FL DEP.[6]

- **Facility Size** – According to the best estimation of Florida DEP, there is no difference in the proportion of small and large facilities between the Round 1 facilities and the total county population of facilities.[7] This is important because the original study found that smaller facilities tended to be in violation more frequently than larger operations. Any difference in the proportions between the sub-population and the greater population would indicate a biased sample. The larger the difference, the larger the bias.

- **Trade association membership**– DEP has indicated that it believes trade association membership can be considered an indicator of greater likelihood of being in compliance. We recommend that DEP compare the proportion of trade association membership of the Round 1 facilities to the total Orange County population to understand if there are more than minor differences. This examination should consider membership status of all facilities at the point in time in which Round 1 facilities were inspected. If the proportion of membership is substantially different, a significant bias exists.

- **Terrain Differences**–The problem of contaminated soils (usually used oil contamination) is a common issue for auto salvage yards. Contaminated soils, in turn, lead to contaminated groundwater. This is a cross-cutting problem, regardless of the specific job specialty of the yard. Variation in type of terrain in which auto salvage yards are sited can be expected to influence both the occurrence of groundwater contamination, and the visual observation of soil contamination (which leads to the discovery of groundwater contamination). For instance, in Orange County, the extreme eastern part of the county (where the original sample was collected) is more moist and green compared with the western part of the county (which is now to be included in the study). The scrubland/sandhill terrain of the western part means that there is a higher possibility of groundwater contamination.

  It is important to confirm that DEP does not expect soil/groundwater contamination, or *the observation* of soil/groundwater contamination, to be qualitatively or quantitatively different between the Round 1 sample and the total Orange County population. DEP has made clear that observation of soil contamination is primarily based upon process-oriented questions during the inspection. For example, Round 1 facilities were specifically asked about their used oil disposal practices in a series of 25 questions. The answers to these questions can be expected to remain comparable in the greater population. Therefore, presumably, one would expect the incidence of the discovery of soil contamination to not be impacted by the terrain. However, the higher possibility of groundwater contamination in the western part of the county introduces a bias if DEP attempts to show results related to the incidence of groundwater contamination.

  We suggest that DEP carefully consider whether the Round 1 facilities are more or less likely than the greater population to be *found* in violation of soil contamination/groundwater contamination regulations due to the terrain upon which they are located. If they are approximately the same, this issue does not present a bias problem. (In a call based upon the first

---

[6] The budget and time-frame did not allow for any independent research by Tellus on background issues related to the earlier study conducted by FL DEP (the Central District Pilot Project). We have relied exclusively on information provided by FL DEP – consisting of written materials and phone calls – for this.

[7] As noted previously, we relied on FL DEP for this information. We have not been provided with data on the new population of facilities to make an independent judgment of this fact.

draft of this memo, DEP indicated that it collected no information on groundwater contamination in the first round of data. Therefore, this issue cannot be examined by comparing Round 3 and Round 1 data, even if such information is collected in Round 3. In further rounds, however, such data could be collected to allow comparison -- bearing in mind the cautions above.)

- **Stormwater permits as a predictor for overall compliance** –Based on the results of the previous study, the understanding was that the presence of a stormwater permit was a good indicator of whether a facility would be found to be in overall compliance. (Stormwater permits are required when there is a point-source discharge to surface water or municipal stormwater drains/pipes.) Is there any indication that this hypothesis could still be relevant for the greater population? If so, DEP should examine the issue more carefully, as described in the next two paragraphs.

  Tellus perceives two important considerations. First, are all facilities equally likely to be required to obtain stormwater permits? As mentioned above, the soil and water characteristics of the old and new populations do differ. However, it was confirmed[8] that stormwater permits must be obtained, when there is a point-source discharge to surface water or municipal stormwater drains/pipes, regardless of whether it enters the groundwater or not. In this case, the eastern and western parts of the county would be comparable, and this issue does not appear to create a bias problem (assuming there are no other factors that would create a difference in likelihood to obtain or need a stormwater permit between the subpopulation and the overall population).

  Second, at the time of the Round 1 inspections, was the proportion of Round 1 facilities holding stormwater permits closely similar to the proportion of permit holders in the total Orange County auto salvage yard population? If not, and if holding a stormwater permit is an indicator of good environmental performance, it is unlikely that the Round 1 data presents a representative picture of baseline performance for the total Orange County population.

Our limited review so far does not suggest sufficient bias to disallow carefully use of the Round 1 data as baseline data. Nonetheless, we suggest that FL DEP carefully reconsider the four characteristics above to better understand whether the Round 1 data can serve as a baseline. In addition, we suggest that FL DEP reconsider whether there are any other characteristics of performance that could be expected to be indicators, and if so, examine those carefully. Other potential indicators mentioned to us include the level of automation at facilities and the presence of computerized inventory -- both of which DEP found to be correlated to facility performance. Because both of these characteristics appear to be functions of facility size, and because DEP is unlikely to have information on these characteristics before conducting inspections in the total population, these characteristics were not considered above.

Question 2: Sample Size & Statistical Confidence Associated With Using Round 1 Data
Will the number of facilities from which data were collected in Round 1 allow for statistical inferences with a high confidence level? (In addition, Tellus briefly examined issues related to what was originally Objective #2 in the proposal, regarding confidence level/interval options and sample size options for Round 3. Tellus did so because the analysis of Question #2 in this list would be incomplete without at least mentioning the ability to compare Round 3 and Round 1.)

To address this question, we first assume the facilities inspected in Round 1 do not present a sampling bias, as discussed above. In examining whether Round 1 data can serve as a baseline for the complete Orange County population of auto salvage yards, we must also examine the issue of whether it provides sufficient statistical confidence.

---

[8] By Lu Burson, FL DEP.

When the study population was only the 15 relevant facilities east of Toll Road 417, it was advisable to simply census the entire population -- as is standard practice for populations numbering fewer than 30. DEP did census the population, inspecting all 15 relevant facilities.[9] Therefore, the results represent the ideal baseline if the study population is confined to those facilities east of Toll Road 417.

However, 15 facilities is a rather small number on which to base statistical results, when we attempt to consider this a sample of the entire Orange County population, which we assume to be comprised of 100 facilities. As mentioned previously, to achieve any meaningful level of statistical significance, a random sample of at least 30 facilities is required. As part of the ERP, it is necessary to compare data from random samples drawn at different points in time; thus it is important to have collected enough data to identify trends and recognize similarities. As a very rough rule of thumb, large differences between these groups can be identified if there are about 20 facilities in each group, moderate-sized differences between groups can be identified with about 50 facilities, and small differences between groups with about 200 facilities from each group[10]. The exact sample size required changes based on statistic used, desired confidence level and desired power. Increasing the sample size will increase the likelihood of finding smaller differences or trends in the data. Based on the original pilot study, which found a high degree of non-compliance initially (above 60% on many criteria) and then a large improvement at the time of follow-up (violations reduced by 50%), we can expect that, at least initially, an ERP in the auto salvage sector in Orange County would most probably involve analyzing large changes.

The determination of the number of inspections required to make statistically valid inferences from this sample of the county-wide population is based on standard statistical theory.[11] If the original sample of 15 facilities is used as a random baseline sample for the entire population of 100 facilities, any results determined from this data through one-sample tests (where results are being tabulated within a single round of data) are subject to a margin of error of +/-24% at a 95% confidence level (or +/-20% at a 90% confidence level). This is a *very large* margin of error. For example, if 50% of facilities were found a non-compliant on a particular question, DEP can have 95% confidence that the actual level of non-compliance for that question among the total population would be between 26% and 74%. (During a call to discuss the first draft of this memo, DEP requested another specific example, which is presented as Attachment 1.)

Our recommendation for the sample size required for one-sample tests for Round 3 is a minimum of 63 facilities, based on a total population size of 100 facilities, a confidence level of 95% and a margin of error of +/- 7.5%[12] If a 90% confidence level is acceptable, a sample size of 55 is sufficient. Even higher numbers of inspections are desirable if DEP wishes to improve the ability to compare Round 3 data with Round 1 data, as discussed in the next paragraph.

The choice of sample size for Round 3 also affects the level of precision achievable with two-sample tests[13] (where comparisons between Round 1 and Round 3 are being made and thus the effect of the margin of error in measuring results in each round is compounded). For purposes of illustration, let us

---

[9] The 16th facility is not considered a relevant facility in the population because it conducts very different additional activities and is categorized under a different SIC code than the other facilities.

[10] See, for example http://www.ncs.com/research-notes/glossary.htm#power and Statistical Rules of Thumb by Gerald van Belle and Steve Millard.

[11] For a more comprehensive treatment of this issue, please consult our Generic Statistical Methodology. We will restrict ourselves here to a discussion of the particular case presented by the FL DEP.

[12] All sample sizes have been calculated using the on-line sample size calculator at http://www.polarismr.com

[13] The calculation of the required sample sizes for two sample tests is based on the expected proportions of the two samples, their standards errors, and the confidence interval and margin of error that is acceptable.

assume that the DEP is trying to test whether there are differences in the proportion of non-compliant facilities in Round 1 and Round 3. Say in Round 1 the DEP found that 75% of the facilities were non-compliant and they expect that they will find an improvement in compliance in Round 3 such that now only 37.5% facilities are out of compliance. The original sample in Round 1 was 15 facilities and we assume that the Round 3 sample is set at 63 facilities. Using a Z-statistic we can test whether this improvement is statistically significant at the 95% level. And, in fact, it is significant (Z=3.03, far above the critical value of 1.96). However, if non-compliance went down from 75% to 51%, this would not be a statistically significant improvement (Z=1.914). Note that if a 90% confidence level is used a drop from 75% to 54% is statistically significant, but a difference smaller than that would not be statistically significant.

Ideally, in each round the same sample size of facilities would be collected. If DEP chose to achieve, for one-sample tests, a minimum confidence level of 95% and confidence interval of +/- 7.5%, the number sampled in each round would be 63 facilities. In such a case, it would be possible to detect as small a drop in non-compliance as 75% to 65% with a 95% confidence level (or a drop of 75% to 66% at a 90% level of confidence). In both cases, the compounded margin of error equals approximately +/-10%[14], which would likely be quite sufficient to measure with statistical confidence the large changes in non-compliance that DEP may expect to result from this program in the beginning. However, DEP should consider whether to decrease the confidence interval further in future years, when one would expect that the improvements in compliance might get smaller simply because there is less room for improvement due to past successes.

The following two paragraphs offer a bit more guidance on approaching the issues of confidence level/interval, and stratification. We recommend that DEP also review the upcoming revision of EPA's "Generic ERP Statistical Methodology," which will provide additional guidance on the implications of one-sample and two-sample testing for statistical confidence and required sample sizes.

Acceptable Confidence Interval and Confidence Level – In general, it is advisable to aim for at least a 95% confidence level (i.e., 95% confidence that the sampling error will not exceed the specified interval) to ensure the precision of results and the credibility of the analysis. This is especially important when one is trying to measure a statistically discernable improvement in compliance from one year to another. A confidence level of 90% is sometimes used, but is less desirable. Lower confidence levels are not considered acceptable in standard statistical practice.

Stratification Issues – In designing the sample, care should be taken to determine whether sample stratification is necessary. If it is expected that facilities differ from one another in a systematic way based on some objective and relevant criterion (such as size, type of ownership, geographic location, etc.), then the sample should be stratified according to that criterion. The simplest way of doing this is to decide the overall sample size based on a level of statistical significance for the entire population, separate the population into the subgroups, and then draw a proportional random sample from each subgroup that adds up to the overall sample size.[15] This method is called representative sampling and it allows one to make

---

[14] The combined margin of error is determined by formulae for pooled sample variance. These formulae change depending on the individual sample variances and sample sizes.

[15] I.e., each subgroup's proportion within the sample should be equal to each subgroup's proportion within the population. For example, suppose and agency has targeted a population of 2500 facilities, with 60% in the northern part of the state. The agency wants to draw a representative sample, so that the northern and southern halves of the state are proportionally represented in the sample. As Figure 1 indicates, the agency should draw a total sample of 160 facilities statewide to be able to draw conclusions about the population as a whole at a 95% confidence level and a confidence interval of (+/-) 7.5 percentage points. To do a proper North/South representative sample, it should draw a random sample of ninety-six facilities from the northern population of facilities, and a random sample of sixty-four facilities from the southern population.

statistically valid conclusions for the whole population.[16] In the case of the auto salvage yard sector in Orange County, it would be advisable to draw a proportional sample from the eastern (original population) and western parts (additional geographic area in the new population) of the county.[17] Disproportionate sampling could be expected to create a bias, because the Round 1 facilities have already been inspected twice and could be expected to have substantially different performance results than other facilities.

Question #3: Quality of Questionnaire.
Will any data previously collected present obvious problems for making statistical inferences (e.g., through problematic phrasing of questions)?

Based on an examination of the inspection questionnaire previously used, we believe that the questionnaire was, for the most part, well constructed. However, a few issues are apparent and worth mentioning:

- *Use of "not applicable."* Answers in which N/A was selected create two potential problems. First, analysts may have difficulty understanding the precise reason for the non-applicability. Second, the effective sample size for a question is, in many cases, reduced by the number of N/A responses -- e.g., if a question is only applicable to a certain class of hazardous waste generator, and generators not in those classes are allowed to answer N/A. Therefore, DEP will have the lower statistical confidence in the results for such questions. (DEP should take special care in considering whether to revise the language of questionnaire to disallow N/A responses. Doing so could make it difficult to compare responses from different rounds, if facilities responded N/A in earlier rounds.)
- *Collection of new data.* It is possible that new data might be desired that was not previously collected. (Although the specifics of this were not indicated to us.) Our recommendation is that to the extent possible all old questions should be retained without any change to phrasing. This is the only way to ensure direct comparability between the old and new rounds. If new questions are added, it is with the understanding that no data on these is available for comparisons to the old round of inspections. The only analysis of the answers to these questions will be within-round analysis (e.g. extent of compliance, rather than improvement in compliance.) If the program remains in place for a long enough time, then the new questions could then form a baseline for future years.
- *Lack of Quantitative Data Collected.* Very little quantitative data is collected. DEP may wish to consider adding such questions, which provide a different sense of performance change (particularly regarding source reduction) over time compared to yes/no questions.
- *Open-Ended Questions.* The use of open-ended questions in certain circumstances (e.g., page 4, #9) may not allow easy data analysis, and may not be interpretable later if the type of response varies significantly.

Conclusion
Based upon the information provided by Florida DEP, we conclude that the sample of facilities previously inspected can likely be used at least in a limited, cautious fashion as a baseline for the ERP

---

[16] Representative sampling can also be used to control the number of inspections that any individual district or region with in an agency's jurisdiction might have to conduct, even if it is not believed that geographic location of facilities is correlated to their compliance response. Many agencies divide inspection resources among different geographic jurisdictions within the state. Without representative sampling, the possibility exists that one or more geographic jurisdictions would face a disproportionate burden of inspections.

[17] However, if one wants to make statistically significant conclusions for any one subgroup (or for comparisons among subgroups) then one must draw statistically significant samples for each subgroup. This greatly increases the total sample size and thus the resources required for the analysis.

that is to be undertaken on a county-wide basis. However, due regard must be taken of the potential problems we have highlighted regarding potential bias and small sample size. This is particularly important when considering conducting one-sample tests on the baseline data, because of the large margin of error associated with it. Even with two-sample tests, a large margin of error will exist, but can be ameliorated somewhat if DEP conducts a sufficiently large number of inspections in Round 3. While the number of inspections required to maintain a meaningful confidence interval for two sample tests may seem at first glance quite high, one would expect that DEP would have to expend even greater resources if it were to initiate new baseline data collection at a higher level of statistical confidence, as well as conducting follow-up data collection. DEP is faced with a very complex problem, and we suggest that DEP carefully reconsider all the information presented in this memo and the underlying assumptions and data provided to Tellus in the development of this memo. Finally, we recommend that DEP consult a qualified statistician -- either internally or externally -- in making final determinations about how to proceed with the next round of data collection.

Attachment 1

During the call to discuss the first draft of this memo, Florida DEP specifically requested that Tellus examine the statistical confidence that would be associated with a question for which 13.3% (2/15) of Round 1 facilities were in compliance. The following items respond to that question, with the assumption that DEP does determine that the Round 1 facilities do not represent a biased sample, and can serve as a random sample of the entire population of 100 Orange County facilities.

1. *Within Round 1, what is the margin of error when the population proportion is assumed to be 13.3% instead of 50%? (One-sample test)*

    Given: The population is 100, sample is 15.

    At the 95% confidence level, this implies a margin of error of +/-16%.
    At the 90% confidence level, this implies a margin of error of +/-13.5%.

2. *If we expect to see an increase in proportion from 13.3% in Round 1 to 26.7% in Round 3 (based upon observations from Round 2), what sample size is necessary in Round 3 to be able to measure that change with statistical significance?*

    Given: Population is 100, sample size in Round 1 is 15. Assume that the true population proportion in each round mirrors the sample proportions.

    Even at a 90% confidence level and even if DEP were to census the population in Round 3, it is impossible to measure this change with statistical significance. Note that 26.7% is within the margin of error for even a one-sample test of Round 1 data (given the small sample). A two-sample test increases the confidence interval further.[18]

---

[18] All sample size calculations have been done using the on-line statistical calculator at http://www.polarismr.com/visitors/stats.htm. The calculations assume a finite population of 100 (i.e., not an infinite population).