 DOI: 10.1111/j.1539-6924.2012.01843.x

Reducing Drinking Water Supply Chemical Contamination: Risks from Underground Storage Tanks

Richard T. Enander,[1,*] R. Choudary Hanumara,[2] Hisanori Kobayashi,[2] Ronald N. Gagnon,[1] Eugene Park,[2] Christopher Vallot,[1] and Richard Genovesi[2]

Drinking water supplies are at risk of contamination from a variety of physical, chemical, and biological sources. Ranked among these threats are hazardous material releases from leaking or improperly managed underground storage tanks located at municipal, commercial, and industrial facilities. To reduce human health and environmental risks associated with the subsurface storage of hazardous materials, government agencies have taken a variety of legislative and regulatory actions—which date back more than 25 years and include the establishment of rigorous equipment/technology/operational requirements and facility-by-facility inspection and enforcement programs. Given a history of more than 470,000 underground storage tank releases nationwide, the U.S. Environmental Protection Agency continues to report that 7,300 new leaks were found in federal fiscal year 2008, while nearly 103,000 old leaks remain to be cleaned up. In this article, we report on an alternate evidence-based intervention approach for reducing potential releases from the storage of petroleum products (gasoline, diesel, kerosene, heating/fuel oil, and waste oil) in underground tanks at commercial facilities located in Rhode Island. The objective of this study was to evaluate whether a new regulatory model can be used as a cost-effective alternative to traditional facility-by-facility inspection and enforcement programs for underground storage tanks. We conclude that the alternative model, using an emphasis on technical assistance tools, can produce measurable improvements in compliance performance, is a cost-effective adjunct to traditional facility-by-facility inspection and enforcement programs, and has the potential to allow regulatory agencies to decrease their frequency of inspections among low risk facilities without sacrificing compliance performance or increasing public health risks.

KEY WORDS: Drinking water; environmental results program; groundwater protection; petroleum products; underground storage tanks

1. INTRODUCTION

Multimedia environmental health risks that arise from commercial and industrial operations can present a variety of regulatory challenges. To address potential threats to air, surface, or ground water, for example, regulatory agencies have employed a number of chemical-specific and sector-based strategies. Driven by science-policy decisions, these strategies have, however, often lacked a statistical foundation for benchmarking compliance performance or for measuring the overall effect size of a given intervention.

In this article, we describe and present the results of an evidence-based, regulatory intervention approach for reducing potential air releases and

[1]Rhode Island Department of Environmental Management, Providence, RI, USA.
[2]University of Rhode Island, Kingston, RI, USA.
*Address correspondence to Richard T. Enander, Rhode Island Department of Environmental Management, 235 Promenade Street, Providence, RI 02908, USA; richard.enander@dem.ri.gov.

incidents of soil and groundwater contamination from underground storage tank (UST) operations at predominantly commercial gasoline stations located in Rhode Island. Historically, groundwater contamination with constituents of fuel such as benzene and methyl tertiary butyl ether (MTBE) has resulted in the impairment of potable water supplies for millions of people nationwide with more than $1 billion per year spent in state and federal funds.[(1)] In Rhode Island, for example, public water drawn from a well field used to service more than 4,000 people in the village of Pascoag was found to be contaminated with MTBE at levels an order of magnitude higher than the drinking water health advisory of 40 ppb.[(2)] Nationally, more than 470,000 UST releases were recorded as of September 30, 2008 (with just over 7,300 new leaks found in FFY'08), while nearly 103,000 old leaks remain to be cleaned up.[(3)]

To prevent leaks and protect groundwater resources, the U.S. Energy Policy Act of 2005 required that state environmental agencies inspect all UST facilities at least once every three years. A number of states, however, either do not have sufficient in-house inspection resources to meet this requirement or are meeting the mandate at significant cost. In 2006, the State of Rhode Island Department of Environmental Management (DEM) received a State Innovation Grant from the U.S. Environmental Protection Agency's (EPA) National Center for Environmental Innovation to help address this issue. The objective of this three-year study was to evaluate an alternative UST inspection model and determine whether it could be a cost-effective alternative to traditional facility-by-facility inspection and enforcement programs.

1.1. Evidence-Based Strategy

The approach used in this study followed the Massachusetts Environmental Results Program (ERP) four-component model[(4)] consisting of regulatory assistance, compliance certification using standardized checklists,[3] independent agency inspections, and statistically-based performance measurement. Initially, randomized field inspections of a representative sample of facilities are conducted by agency staff to establish baseline compliance rate conditions; that is, before any state-led intervention is launched. Once independent baseline field data are in, a detailed plain-English regulatory guidebook and self-audit checklist package is mailed to each facility owner/operator in the regulated universe. Facilities are provided a defined time interval to conduct comprehensive self-evaluations, complete audit checklists, and return all certified results to the state—while agency technical expertise and training are provided, it is not unusual for companies to hire outside third-party inspectors to complete the initial self-audit requirement. At the conclusion of the self-audit time period, an additional set of comprehensive, randomized agency field inspections are carried out to assess whether compliance performance improvements occurred. Criteria established by regulation and beyond compliance measures are then quantitatively assessed using objective onsite field inspection data and accepted statistical techniques. Note that the use of the guidebook and checklist, along with agency expertise, are significant assistance interventions not currently employed in a traditional inspection program.

2. STUDY BACKGROUND

With EPA funding support, Rhode Island became the first state to apply the ERP model to the regulation of USTs. The Rhode Island program is mandatory and requires facilities to self-certify to compliance standards using a comprehensive facility evaluation checklist and workbook[(5,6)] that were developed by EPA, DEM, and a group of external stakeholders.[(7)] At the time of the study, there were 1,910 federally regulated tanks at 608 facilities and approximately 1,500 heating oil tanks, all of which were required to register with DEM.

2.1. Program Elements

The ERP is an integrated approach to improving regulatory compliance and achieving environmental protection. DEM adopted the statewide UST ERP,

[3] A similar checklist intervention approach is recommended by the World Health Organization (WHO) for use in surgical settings. In a Harvard led study published in the *New England Journal of Medicine*, investigators found that the WHO self-administered checklist intervention approach—with baseline/postintervention statistical assessment and technical assistance in the form of written materials, lectures, direct guidance, training, and site visits—was successful at reducing surgical death rates by nearly one-half ($p = 0.003$) and inpatient complications from 11% to 7% ($p = 0.001$). See: World Health Organization. Implementation Manual Surgical Safety Checklist (2008 First Edition): Safe Surgery Saves Lives. Available at: http://www.who.int/patientsafety/ safesurgery/tools_resources/ SSSL_Manual_final Jun08.pdf; Hayes AB *et al.* A Surgical Safety Checklist to Reduce Morbidity and Mortality in a Global Population. *New England Journal of Medicine* 2009; 360(5):491–499.

as it believed that the comprehensive nature of the model would assist UST system owners and operators in their efforts to understand and comply with complex system requirements—thereby reducing potential threats to human health and the environment. The ERP provides relevant information necessary for owners/operators to understand maintenance and operational requirements pertaining to UST systems, while improving accountability to the public for environmental performance.

Key elements of the Rhode Island ERP are:

- Certification Workbook for UST Facilities which specifies compliance requirements and best management practices,
- Compliance Certification Checklist and Forms Booklet that consists of a Self-Audit Checklist (yes/no questions), Certification Statement and Return to Compliance Plan forms required to be completed by the owner/operator. On the Certification Statement form (text shown later), the UST system owner/operator must certify the current compliance status of the facility and acknowledge that the facility must comply with all applicable environmental laws. The Return to Compliance Plan form is used for compliance problems (identified in the process of filling out the Compliance Certification Checklist) that cannot be corrected before submittal of the certification forms. The Return to Compliance Plan describes what steps the facility will take to meet its requirements and when it will return to full compliance,
- Randomized Independent State Agency Facility Inspections for performance measurement and to confirm the accuracy of self-reported audit results and compliance with the UST system regulations, and
- Workshops and Technical Assistance to inform owners and operators of their responsibilities under ERP.

The submittal of self-audit data to DEM is accompanied by a "Certification Statement," which the owner/operator must sign under penalty of law. The certification language is presented below:

> "I_______________, as the UST owner(s) attest,
>
> 1) That I/we have personally examined and am/are familiar with the information contained in this submittal, including any and all documents accompanying this certification statement;
>
> 2) That, based on my/our inquiry of those individuals responsible for obtaining the information, the information contained in this submittal is, to the best of my/our knowledge, true, accurate and complete;
>
> 3) That I/we am/are fully authorized to make this attestation on behalf of this facility;
>
> 4) That _______________ is/are the Operator(s) of this facility. I have discussed the division of duties with the operator(s). I understand that the Department of Environmental Management may pursue either the owner, operator or both for any violations of the Rules and Regulations For Underground Storage Facilities Used For Petroleum Products and Hazardous Materials, where owner/operator is mentioned.
>
> 5) I/we am/are aware that there are significant penalties for submitting false information.
>
> *If owner and operator are separate individuals, Operator must also sign:*
>
> I/we as the operator(s) of the Facility attest that I/we am/are fully authorized by the Facility owner(s) to sign this certification statement. I acknowledge that I am the operator of this facility. I have discussed the division of duties with the owner(s) and clearly understand my/our responsibilities. I/we understand that the Department of Environmental Management may pursue either the owner, operator or both for any violations of the Rules and Regulations For Underground Storage Facilities Used For Petroleum Products and Hazardous Materials, where owner/operator is mentioned. I/we am aware that there are significant penalties for submitting false information, including the possibility of fine and imprisonment for knowing violations."[6]

2.2. Facility Compliance

To be in compliance a facility participating in ERP must meet minimum DEM requirements for its UST system. Facilities must meet all environmental requirements for each regulated UST system. The UST system requirements include spill, overfill, corrosion protection; release detection; financial responsibility; proper installation, correct operation, maintenance, repair, and testing of each system; controlling, reporting, and remediating releases; reporting and record keeping; and temporary and permanent closure. Under state law and in accordance with ERP provisions, owners and operators of one or more UST systems are responsible for preventing and quickly detecting UST system releases—facilities are also responsible for reporting and cleaning up any releases that occur and are held accountable for UST system(s) leaks. The DEM ERP is designed to help ensure that regulatory requirements are routinely met and that releases do not occur.

2.3. Agency Compliance and Inspection Program

Rhode Island UST system owners/operators are required to comply with the Rules and Regulations For Underground Storage Facilities Used For Petroleum Products and Hazardous Materials, December 2008.[8] Before enactment of the U.S. Energy Policy Act of 2005, RI General Law 46–12-30.2 required facility inspections to be conducted once every two years. To meet this inspection mandate, DEM made a policy decision to supplement the onsite inspection programs with the sector-wide ERP self-certification model; due to historic resource constraints, comprehensive onsite inspections of each regulated facility could only be conducted just once every six years. Today, statewide compliance efforts and agency facility inspections, conducted in accordance with the 2005 U.S. Energy Policy Act requirements, are supported by the RI UST ERP which requires all registered facilities to complete and submit self-audit/certification checklists and return-to-compliance plan forms on a three-year cycle.

3. METHODS

The Rhode Island ERP collects binary (yes/no), count and descriptive data submitted in the form of self-certification checklist responses and Return-to-Compliance (RTC) plans.[9] Field data are also collected during the course of independent baseline and postintervention facility audits conducted by agency staff. For each registered UST system, operators were required to complete and submit a Compliance Certification Checklist and Certification Statement (and, if required, an RTC form) to DEM by June 30, 2005. The Compliance Certification Checklist, Statement, and RTC Plan forms were included in a Forms Booklet that was mailed to all 608 Rhode Island registered facilities. To determine whether compliance improvements occurred sector-wide, data collected from a statistically predetermined number of randomly selected facilities, inspected at baseline and postcertification, were subjected to statistical analysis

3.1. Field Data Collection

UST baseline ($n = 96$) inspections were conducted April–September 2004 by four DEM regulatory staff using the standardized self-certification ERP checklist. In December 2004, ERP self-certification workbooks and checklists were mailed to each registered UST facility ($n = 608$). Six outreach workshops were held in February and March 2005 with 297 people attending. A total of 462 ERP self-certification booklets were returned to DEM on time by June 30, 2005. On July 28, 2005, a Notice of Intent to Enforce (NOI) was sent from the UST program to 146 facilities that had not submitted. The NOI stated that facilities had until August 31, 2005 to submit their self-certification checklists and RTC plans; as a result, 131 additional self-certification booklets were submitted to DEM. In total, 283 facilities ultimately submitted 1,291 RTCs during the first round of the UST ERP. More than 30 formal enforcement actions, 100 violations, and $300,000 in penalties were assessed for noncompliance with ERP requirements (based on failures to submit self-certification statements and violations discovered during postintervention inspections).

Postintervention facility inspections ($n = 93$), for statistical comparison with baseline data, were conducted May 2007 through August 2008. Following training meetings with project interns, baseline and postintervention field data were entered into ten Excel worksheets. In total, more than 45,000 data points were entered into spreadsheets by student interns; data were crosschecked for accuracy and completeness. Baseline and postintervention sample size was determined as shown in Appendix Section A.1.

3.2. Data Evaluation and Grouping

Baseline and postintervention inspections were conducted by DEM regulatory staff using a *Facility Profile and UST Facility Inspection Report* compliance inspection checklist. The checklist contained 118 Y/N regulatory compliance questions (i.e., requiring a RTC plan for "N" responses) addressing tank corrosion protection, tank leak detection, piping corrosion protection, piping leak detection, spill prevention and overfill protection, spill containment, and groundwater monitoring.[5] Data from each of the 118 environmental compliance indicator questions were evaluated and then grouped according to their expected utility for measuring UST sector performance improvements in future years.

First, all potentially measurable indicators were grouped together in one key table. It was determined that responses to 41 of the original 118 UST compliance checklist questions could be used for future performance measurement purposes—that is, trial calculations (combining baseline field data with "assumed/projected" postintervention performance improvement proportions) showed that it was mathematically possible to measure statistically significant improvements (if observed) in these variables over time.[4] Rather than winnowing down the list of potentially measurable indicators into a smaller number or subset of Environmental Business Practice Indicators,[5] all 41 variables were carried forward in the performance measurement calculations. Using readily available software, this approach assured that all areas of noncompliance were identified and statistically characterized. Because indicator variables were not screened out at the beginning of the process, all statistically significant changes in performance could be monitored over time; adjustments for multiple comparisons were made during the analysis.

A second table was then created for 19 of the 118 indicators that showed 96% to 100% compliance at baseline. Due to high performance at baseline, these indicators could not be used to demonstrate statistically significant *improvements* in future years—they could be used, however, to evaluate future trends (i.e., whether observed compliance rates were being sustained over time). Further, since most of the identified "performance trend" indicators were technology/equipment driven (i.e., compliant equipment—such as, fill pipes equipped with drop tubes; drop tubes end within 6" of bottom; spill buckets have a minimum of 3 gal. capacity; and CARB hoses certified—were installed before baseline audits), statistically significant compliance rate decreases in performance are unlikely to be found in future years.

Of the remaining indicators, 24 were not suitable for measurement purposes either due to small sample size (2/24 indicators) or because the indicators (22/24 indicators) were not applicable to any of the facilities in the random sample at baseline; 34 additional indicators, related to Stage I/II vapor recovery, could not be evaluated further since data were not collected during the postintervention site inspection phase (even if postintervention data had been collected, Stage II facility vapor recovery indicators would not have been measurable after 2013 as vehicles will be equipped with onboard vapor recovery equipment). Therefore, only 41 of 118 original checklist questions/indicators (or 35%) had the potential to show statistically significant improvements over time.

3.3. Indicator Hypothesis Testing

The objective was to determine whether "improvement" over baseline conditions occurred in the postintervention setting for each of 41 potentially measurable compliance indicators. For this purpose, it is appropriate to use the Fisher's exact probability test when sample sizes are small which is the case for a number of indicators. Hence a one-sided test was chosen and study hypotheses were defined as H_o = no difference in the proportion of facilities in compliance at baseline and postintervention, and H_a = improvement in compliance postintervention. Thus, p values for each indicator variable were calculated and compared to the nominal level of significance of 0.05 to determine whether statistically significant improvements occurred over time (description of statistical test for multiple comparisons given in Appendix Section A.2.).

3.4. Negative Binomial Regression Analysis

To address the contribution of both positive and negative values to overall compliance performance, a negative binomial regression analysis was performed on grouped data. The selected 41 indicators were grouped into five compliance categories: tank corrosion protection (2 indicators, Category B), tank leak detection (12 indicators, Category C), piping leak detection (11 indicators, Category E), spill prevention and overfill protection (11 indicators, Category F), and groundwater monitoring and tank pad observation wells (5 indicators, Category I). With this analysis we sought to understand whether overall statements could be made regarding compliance improvement across broad regulatory categories rather than on an individual, indicator-by-indicator basis as in section 3.3 above (details provided in Appendix Section A.3.).

[4]For example, of nine facilities to which the indicator "E.22 System calibrated and inspected last year" applied, 44% ($p_1 = 0.44$) were found to be in compliance with the identified regulatory standard at baseline. Trial runs of the Fisher's Exact test using an online statistical program (http://www.quantitativeskills.com/sisa/statistics/fisher.htm), showed that if a second round sample of facility inspections (postintervention) were to result in a performance improvement of 47% over baseline, for example, then this difference (i.e., increase in compliance) would be statistically significant at the $p < 0.04$ level.

[5]Environmental Business Practice Indicators are "industry-specific performance measures that provide a snapshot of a facility's environmental performance." They can be viewed as a subset of measures that have been selected from a larger pool of potentially measurable environmental performance indicators. Source: About the Massachusetts Environmental Results Program, http://www.mass.gov/dep/service/ about11.htm.

3.5. Significant Operational Compliance Data

Twice each year, states submit comprehensive UST inspection compliance data to EPA's Office of Underground Storage Tanks (OUST). The percentage of facilities in Significant Operational Compliance (SOC) with release prevention, release detection, and release detection and prevention regulations are reported to OUST at mid-year and year end. OUST defines SOC to mean that the UST systems at a facility have proper equipment/procedures in place and are being properly operated and maintained in order to detect a release. The OUST "End of Year Activity Reports" describe the reported data sets as follows:

> Compliance rates are based on the percentage of inspected facilities found to be in significant operational compliance with federal UST requirements over the course of a federal fiscal year. In accordance with EPA guidelines, states are allowed to report based on requirements more stringent than the federal SOC requirements. Furthermore, states have different approaches to targeting inspections. For example, some states focus inspections on suspected noncompliant facilities, while other states conduct random inspections.[10]

On the basis of how the data were collected, DEM found that it was not possible to make statistically valid comparisons of compliance performance over time using OUSTs "aggregated" SOC data sets since Rhode Island's data, like most other states, largely represented changing proportions of nonrandom samples over time—that is, with the exception of ERP data, Rhode Island's inspections were largely targeted inspections representing a small proportion of the regulated universe in any given year. The% SOC data, therefore, represented those facilities that comprised a target group in each end of year activity report and were not reflective of the universe of facilities as a whole. Because direct comparisons between these year end data sets were not possible, DEM reviewed historic field inspection data—for the 96 baseline and 93 postintervention ERP "randomly selected" facilities—and calculated aggregate %SOC scores for each of the three standard EPA OUST compliance categories: release prevention, release detection, and release detection and prevention. Statistical comparisons were then made on the three sets of grouped data (see Results section 4.4, Table III for detail).

3.6. Economic Analysis

A first order analysis of the costs associated with the traditional RI UST inspection program was performed and compared with the costs needed to support the alternate ERP approach. Various ERP models were assessed where both sample size (100 or 250 inspections) and frequency (every one to three years) were combined in five different scenarios. Because of the fewer inspections required for ERP, costs associated with inspections would be reduced for each scenario. Additional expenses to support ERP-related activities (workshops, data gathering, statistical analysis, oversight) and "payback" or time to recover ERP start-up costs and realize savings were also evaluated.

All analyses were performed on an annual cost basis. Costs associated with data gathering and analysis occurs only once per cycle, so these costs were spread out over the time period of the cycle. Fixed cost items that had to be evaluated included a staff person responsible for collecting and organizing self-certification forms as well as other ERP administrative duties throughout the entire cycle ($20,000/year). When ERP was programmed to include inspections every two or three years, costs associated with data gathering and analysis in each round were divided accordingly to show costs on a yearly basis and then added to the fixed staff cost of $20,000. For example, if 100 random inspections were conducted every year, the projected cost to support ERP activities would be calculated as $20,000 (staff person) + $$X$ (ERP Manager) + $$Y$ (Intern) + $$Z$ (mailing costs). If the ERP cycle were extended to once every two years, the work required to support ERP would cost $40,000 (two years of staff person) + $$X$ (ERP Manager) + $$Y$ (Intern) + $$Z$ (mailing costs).

4. RESULTS

4.1. Indicator-by-Indicator Analysis

Data resulting from the 96 randomized baseline and 93 postintervention field inspections were organized into two tables as discussed earlier. The 41 measurable indicators, from the original list of 118 checklist questions, were the primary focus of the statistical analysis component of this study. The Fisher's exact probability test was used to test for performance improvements among the 41 indicators—on an indicator-by-indicator basis. The objective was to determine whether "improvement" over baseline conditions occurred in the postintervention setting for each compliance variable.

Table I. UST Facility Baseline (96 Random Inspections, '04) and Postintervention (93 Random Inspections, '07) Performance Comparisons

	Baseline		Post intervention		Statistical Comparison	
Measurable Indicators	Sample Size (n_1)	Proportion (p_1)	Sample Size (n_2)	Proportion (p_2)	Percentage[a] Change (95% CI)[b]	P[c]
1 E.16 Tightness tests annually+	6	0.00	9	0.22	22	0.343
2 E.17 Passing results for each reqd. year	6	0.00	9	0.33	33	0.185
3 B.21 Is system tested every 3 years + w/in 6 months of repair	7	0.14	12	0.50	36	0.144
4 I.4B/I.8P records of GW monitoring well checks	55	0.18	42	0.60	42 (24, 60)	<0.001**
5 C.28 W/ ATG, >20 years: tightness test passing results, 2 years.	17	0.41	15	0.73	32	0.07
6 B.25 Records of all repairs/test results	7	0.43	12	0.83	40	0.095
7 E.22 System calibrated and inspected last year	9	0.44	11	0.91	47 (5, 75)d	0.038*
8 F.3 Inspect spill buckets daily	94	0.52	93	0.40	−12	
9 E.4 Records of LLD tests for last 3 years.	81	0.58	69	0.68	10	0.135
10 F.11 Sumps free of water/debris/product	81	0.62	78	0.76	14 (0, 28)	0.043*
11 E.21 Records of system checks/repairs	10	0.60	15	0.93	33	0.064
12 E.12 System calibrated/inspected last year	65	0.66	60	0.80	14	0.062
13 C.20 Monitoring system been calibrated/inspected past year	55	0.67	62	0.79	12	0.110
14 E.20 Continuously use interstitial monitoring	12	0.67	17	0.94	27	0.078
15 I.5B/I.6P Well caps closed tightly and locked	92	0.67	85	0.95	28 (23, 33)	<0.001**
16 F.2 Tank have operational spill containment device	96	0.68	93	0.69	1	0.497
17 C.14 ATG sys calibrated and inspected last year	80	0.69	70	0.81	12	0.055
18 E.11 Records of system checks/repairs	67	0.73	61	0.75	2	0.464
19 I.2B/I.4P Wells equipped w/road box and lock cap	92	0.73	85	0.96	23 (13, 33)	<0.001**
20 C.31 Records of inventory control	94	0.74	81	0.70	−4	
21 E.7 Conducted tightness test w/in past year	17	0.76	10	0.90	14	0.371
22 C.13 Records of last 36 months ATG sys checks	78	0.77	70	0.74	−3	
23 C.19 Records of monthly sys checks for past 36 months	56	0.79	66	0.73	−6	
24 C.10 Use ATG to conduct leak rate tests	82	0.79	71	0.85	6	0.267
25 C.11 Recent ATG leak rate tests pass	63	0.79	61	0.62	−17	

(Continued)

Table I. Continued

Measurable Indicators	Baseline Sample Size (n_1)	Baseline Proportion (p_1)	Post intervention Sample Size (n_2)	Post intervention Proportion (p_2)	Statistical Comparison Percentage[a] Change (95% CI)[b]	P[c]
26 F.13 Sensors upright and at correct height	73	0.79	76	0.96	17 (7, 27)	0.002**
27 F.8 Containment sump present	96	0.80	93	0.81	1	0.602
28 C.12 Records of last 36 months leak test	67	0.81	60	0.95	14 (3, 25)	0.013*
29 F.15 Sensors mounted properly	73	0.81	76	0.96	15 (5, 25)	0.003**
30 C.30 Perform inventory control properly	91	0.81	81	0.77	−4	
31 F.12 Sumps have sensors	82	0.82	78	0.97	15 (6, 24)	0.001**
32 F.19 Qualified UST contractor check device	87	0.84	90	0.98	14 (6, 22)	0.001**
33 I.3B/I.5P Wells equippedw/ pipe not screened at top	91	0.85	85	0.95	10 (1, 19)	0.017**
34 C.26 W/ ATG, <20 yrs: tightness test passing results	22	0.86	14	1.00	14	0.216
35 E.1 Leak detection method in place for each run	93	0.91	85	0.98	7	0.067
36 C.7 Leak detection system operating properly	93	0.92	90	0.94	2	0.406
37 F.14 Sensors functioning properly	72	0.93	76	0.95	2	0.466
38 F.6 Fill pipes/box covers labeled/marked	96	0.94	93	0.94	0	
39 E.10 Interstitial monitoring for leaks	71	0.94	61	0.97	3	0.415
40 F.17 Secondary piping test boot disconnected	75	0.95	75	0.95	0	
41 I.6B/I.7P Are any well caps submerged under water	91	0.95	85	1.00	5 (3, 7)	0.035**

Notes: CI = confidence interval; n = number of facilities in sample; p = postintervention; p_1 = number of facilities in compliance at baseline/number of facilities in the sample; p_2 = number of facilities in compliance postintervention/number of facilities in the sample; For an indicator, a facility was counted if one or more tank "Y's" or "No's" were recorded. The facility was in compliance only if one or more. "Y's" were recorded and no "N's".

[a]Calculated as $100(p_2 - p_1)$. Sample calculation for indicator No. 33 (I.3B/1.5P Wells equipped w/ pipe not screened at top): 100 (0.95 – 0.85) = 10.

[b]95% CIs calculated for indicators showing statistical significance at $\alpha = 0.05$; 95% CIs calculated as $(p_2 - p_1) \pm 1.96 \times$ square root $[p_1(1.00 - p_1) / n_1 + p_2(1.00 - p_2) / n_2]$. Sample calculation for indicator No. 33 (I.3B/1.5P Wells equipped w/ pipe not screened at top): CI = (0.95 − 0.85) ± 1.96 × square root [0.85(1.00 − 0.85)/91 + 0.95(1.00-0.95)/85] = 0.1 ± 1.96 sq. root (0.0443) = 0.1 ± 0.09.

[c]P-values were calculated with the Fisher's exact test online, available at http://www.quantitativeskills.com/sisa/statistics/fisher.htm

[d]Due to small sample size, computation of the confidence interval on the difference of proportions for E.22 followed Agresti A and Caffo B (2000), American Statistician, pages 280-288.

P-Values calculated only for performance indicators showing improvement (1-tailed test).

*Statistically significant at the 0.05 (95%) confidence level without an adjustment for multiple comparisons.

**P Value = Holm's-modified Bonferroni adjustment for multiple comparisons calculated on a category by category basis (i.e., B, C, E, F, and I).

4.2. Measurable Indicators and Hypothesis Testing

Table I is comprised of eight columns and lists data for the 41 indicators determined to be potentially measurable. The first two columns provide the indicator number and checklist question descriptor (the full checklist question, to which the descriptor relates, can be found at: http://www.dem.ri.gov/programs/benviron/assist/usterp/pdf/ustcl04.pdf). The next four columns list sample size and compliance proportion data for sampled facilities at baseline (n_1, p_1) and postintervention (n_2, p_2). Data provided in column "n_1" under the "Baseline" header, for example, show the total number of UST facilities inspected, and to which the specified regulatory requirement was found to apply, before the ERP initiative was launched; corresponding proportions (p_1) for those facilities found to be in compliance at baseline with the relevant indicator is also shown. Though 96 facilities were inspected at baseline, n_1 is shown to range from 6 to 96 as regulatory criteria did not uniformly apply to all facilities within the baseline sample—that is, the applicability of any given regulatory requirement varied according to technology/operations present. More than 75% of the indicators applied to $\geq$55 of the sampled facilities; in general, as sample size increases, smaller improvements in compliance performance can be detected.

The "Percentage Change" column shows the difference in compliance rate proportions for each indicator at baseline and postintervention levels; the 95% confidence interval for those indicators showing statistically significant improvements—at the 0.05 level of significance, P value in the final column—is also shown. Baseline compliance at the sampled facilities ranged from 0% to 95% for the 41 indicators listed in Table I; by comparison, post-ERP intervention compliance ranged from 22% to 100%. In Table I, the median compliance rate proportion at baseline (column p_1) is 76%, indicator number 21; whereas the median for all post-ERP intervention data was 85% (column p_2). A comparison between indicators below the median, containing the lowest compliance rate proportions, to those above the median showed that the indicators in the lower half of the data set related predominantly to periodic equipment testing/calibration, record keeping, and inspection requirements, whereas those in the upper half are weighted more toward technology and operational requirements, along with some record keeping/inspection requirements—compliance rate proportions for these indicators range from 77% to 95%.

Statistically significant improvements in performance among the 41 ungrouped "individual compliance indicators" were found subsequent to ERP implementation, as follows: (1) at the 95% confidence level, 12 of 41 (29%) compliance indicators showed statistically significant improvements using the Fisher's exact test, (2) after applying the Holm's modified Bonferroni adjustment for multiple comparisons, 3 of the 12 indicators with $p < 0.05$ were no longer considered significant, (3) at the 90% confidence level, 19/41 (46%) indicators showed significant improvement (Table I). For 6 of the 41 measurable indicators, the change was negative, as reported in Table I. Of these six, only two (F3 and C11) showed a large negative change—that is, >10%; small changes of 4% or 6% are to be expected. Because our objective in the UST ERP intervention study was to test for improvement in the postintervention setting—that is, improvement over baseline conditions—we conducted a one-sided statistical test. In a two-sided test, investigators look for changes in any direction which was not our objective in the original study design and therefore the magnitude of the two negative changes were not evaluated. To address the contribution of both positive and negative values to overall compliance performance and to address category-wide compliance, a negative binomial regression analysis was performed.

Finally, it should be noted that all of the performance improvement measures (p_2, p_1) listed in Table I are observed values. The calculation of a 95% Upper and Lower Confidence Limit (UCL, LCL), therefore, provides an interval within which one would expect the true difference (p_2, p_1) in industry performance to lie (i.e., with only a 1 in 20 chance that the true difference would be found outside this range). Thus, confidence intervals make it easier to recognize the differences that are too small to be of statistical significance. Confidence limits for all statistically significant results are given.

4.3. Negative Binomial Regression Analysis

Negative binomial regression analysis was used to assess compliance performance over time by grouping 41 measurable indicators into five "categories"—that is, tank corrosion protection, tank leak detection, piping leak detection, spill

Table II. Comparison of Compliance Performance at Baseline and Postintervention

Categories with Measurable Indicators	Average (%) Improvement (n)[a]	95% CI	P[b]
B. Tank corrosion protection Repair/test records kept, three-year system tests	38 (2)	−0.011, 0.773	0.078**
C. Tank leak detection Proper operation, leak rate tests, systems calibration, tightness tests, record keeping, inventory control	4 (12)	−0.037, 0.111	0.703
E. Piping leak detection System check/repairs record keeping, leak detection, interstitial monitoring, tightness tests, system calibration	9 (11)	0.004, 0.170	0.016*
F. Spill prevention and overfill protection Labeling/marking, spill containment, sensors functioning, inspections/record keeping, equipment in good condition	6 (11)	0.008, 0.118	0.012*
I. Groundwater monitoring wells and tank pad observation wells Equipped w/ road box, locked caps, not screened, caps not submerged/closed tightly, record keeping	21 (5)	0.164, 0.268	<0.001*

Notes: [a]Average performance improvement for measurable indicators by compliance category; n = total number of measurable indicators in each category; CI = confidence interval; Average % improvement calculated as the difference between means postintervention (m_2) and at baseline (m_1) * 100; Baseline/postintervention means = the proportion of all indicators in compliance (x's) at each facility, summed across all facilities within a given category and divided by the number of facilities in that category at baseline (n_1) and postintervention (n_2); sample variance (s) = ((x1-xbar)**2 + ··· +(xn-xbar)**2)/(n-1); 95% CI's on the difference between means = (m_2-m_1)±1.96*((s1/n_1)+(s2/n_2))**0.5, t value for B.=2.11, 17 degrees of freedom.

[b]P calculated using negative binomial regression analysis on categorical data; one-sided p- values were: 0.039 (B), 0.352 (C), 0.008 (E), 0.006 (F), and 0.001 (I).

*P ≤ 0.05, statistically significant at the 95% confidence level.

**P ≤ 0.10, statistically significant at the 90% confidence level.

prevention and overfill protection, and groundwater monitoring and tank pad observation wells. Comparisons of overall industry compliance at baseline and post-ERP intervention found statistically significant improvements in compliance by "category" as shown in Table II. In category B, the Negative Binomial model showed the best fit based on the Akaike Information Criterion (AIC) fit criteria (AIC = 45.12) among the different models (i.e., Poisson, Poisson with deviance adjusted, Negative Binomial); the parameter estimate of the explanatory variable (indicating the difference in compliance from post to baseline) showed significance ($\beta = +0.698$, Wald $\chi^2 = 3.11$, $p = 0.0776$) at the 0.1 level. Similar results were obtained for categories E (AIC = 636.56, $\beta = +0.132$, Wald $\chi^2 = 5.76$, $p = 0.0164$); F (AIC = 707.40, $\beta = +0.0617$, Wald $\chi^2 = 6.27$, $p = 0.0123$); and I (AIC = 462.29, $\beta = +0.263$, Wald $\chi^2 = 53.81$, $p < 0.0001$). In category C, the Negative Binomial model also showed the best fit based on the AIC fit criteria (AIC = 832.10, $\beta = +0.017$) among the three models, however the parameter estimate of the explanatory variable was not significant ($p = 0.7030$). Overall, significant improvements in compliance were found for four of five compliance categories ($p < 0.05$ for three categories); the average observed compliance performance improvement for these four categories ranged from 6% to 38%.

4.4. Significant Operational Compliance

To compare Rhode Island baseline and postintervention compliance rates, DEM regulatory staff reviewed historic field inspection checklist data for the 96 baseline and 93 postintervention ERP "random" inspections and calculated aggregate %SOC values for the three standard EPA OUST compliance categories: release prevention, release detection, and release detection and prevention. A comparison of facility compliance, using OUST SOC reporting metrics, at baseline to that post-ERP intervention showed statistically significant improvements in performance over time with point estimates

Table III. UST Facility Significant Operational Compliance Comparisons

	Baseline ('04)		Post intervention ('07/'08)		Statistical Comparison	
	Sample Size (n_1)	Proportion (p_1)	Sample Size (n_2)	Proportion (p_2) (95% CI)[a]	Percentage Change[b] (95% CI)[c]	P[d]
Random inspection data						
% in SOC w/ release prevention	96	0.75	93	0.94 (0.89, 0.99)	19 (9, 29)	<0.001*
% in SOC w/ release detection	96	0.53	93	0.75 (0.67, 0.84)	22 (9, 35)	0.001*
% in SOC w/ release detection and prevention	96	0.51	93	0.72 (0.63, 0.81)	21 (8, 35)	0.002*

Notes: n = number of facilities in sample; p = proportion = no. of facilities in compliance/number of facilities in the sample or assumed universe of facilities; CI = confidence interval.
[a]Wald CI calculated using online program at: http://www.measuringusability.com/wald.htm.
[b]Calculated as $100(p_2 - p_1)$.
[c]c95% CIs calculated for indicators showing statistical significance at $\alpha = 0.05$; 95% CIs calculated as $(p_2 - p_1) \pm 1.96 \times$ square root $[p_1(1.00 - p_1) / n_1 + p_2(1.00 - p_2) / n_2]$.
[d]P-values calculated with the Fisher's exact test online, available at http://www.quantitativeskills.com/sisa/statistics/fisher.htm.
*P ≤ 0.05, statistically significant at the 95% confidence level.

ranging from 19% to 22%, Table III—which was in agreement with previous findings.

4.5. Economic Analysis

Table IV presents the results of a first order comparative analysis of costs. As shown, there was only one fixed cost item included in the calculations—a staff person responsible for collecting and organizing self-certification forms as well as other ERP administrative duties throughout the entire cycle ($20,000/year). When ERP was programmed to include inspections every two or three years, the other costs associated with data gathering and analysis in each round were divided accordingly to show costs on a yearly basis and then added to the fixed staff cost of $20,000. For example:

- On Line 5, if 100 random inspections were conducted every year, then the projected cost to support ERP activities is shown to be $20,000 (staff person) + $12,000 (ERP Manager) + $1500 (Intern) + $2000 (mailing costs) = $35,500.
- If the ERP cycle is extended to once every two years, the work required to support ERP would cost $40,000 (two years of staff person) +$12,000 (ERP Manager) + $1500 (Intern) + $2000 (mailing costs) = $55,500 every two years or $55,500/2 = $27,750 per year.
- If the ERP cycle is extended to once every three years, the work required to support ERP would cost $60,000 (three years of staff person) +$12,000 (ERP Manager) + $1500 (Intern) + $2000 (mailing costs) = $75,500 every three years or $75,500/3 = $25,167 per year.

For sample sizes of 250, one year cycles were not applicable because 250 inspections/year would be the current "traditional program" so analyses for 2 and three-year cycles are presented using the same approach. The costs to start-up ERP are also presented in line 8 of Table IV and includes workbook development, workshops, etc. Annualized inspection costs are presented in Lines 1–4 for the different ERP scenarios. The payback is calculated by dividing the net savings of ERP (Line 7) into start-up costs (Line 8) and is shown to range from 0.65 to 1.22 years.

5. DISCUSSION AND CONCLUSIONS

Application of the ERP model to the Rhode Island UST sector produced positive, measurable results. Before conducting the 96 randomized baseline facility inspections in 2004, DEM did not have a clear or reliable picture of historic compliance levels for the industry sector. Inspections conducted in prior years were targeted and, therefore, not statistically representative of the entire regulated universe. By applying ERP methodology, 2004 baseline levels of compliance performance for measurable indicators were found to range from 0% to 95%, with a median of 76%. By itself, the use of this ERP

Table IV. Rhode Island Underground Storage Tank Inspection Program Economic Analysis

		I. Traditional Inspection		II. ERP (100 Sample Size)			III. ERP (250 Sample Size)	
	Scenario	~250/year		100/year	100/two years	100/three years	250/two years	250/three years
Line 1	Personnel required	40% Inspector1	$36,964	$14,786	$7,393	$4,929	$18,482	$12,321
		20% Inspector2	$16,698	$6,679	$3,340	$2,226	$8,349	$5,566
		90% Inspector3	$95,566	$38,226	$19,113	$12,742	$47,783	$31,855
		5% Inspector4	$5,878	$2,351	$1,176	$784	$2,939	$1,959
		10% Insp. supervisor[a]	$13,676	$13,676	$13,676	$13,676	$13,676	$13,676
Line 2	Annual personnel cost (salary, benefits, overtime)	$168,782		$75,718	$44,697	$34,357	$91,229	$65,378
Line 3	Travel expense 30 miles ave./trip, @ $.48/mile	$3,600		$1,440	$720	$480	$1,440	$960
Line 4	TOTAL ANNUAL INSPECTION OPERATING COSTS	$172,382A		$77,158	$45,417	$34,837	$92,669	$66,338
Line 5	ERP Annual operating costs[b]			$35,500	$27,750	$25,167	$32,000	$28,000
Line 6	Total annual inspection and ERP costs (4 + 5)			$112,658	$73,167	$60,004	$124,669	$94,338
Line 7	Net Annual Savings From Traditional Program (4A-6)			$59,724	$99,215	$112,378	$47,713	$78,044
Line 8	ERP start-up costs[c]			$73,000	$73,000	$73,000	$73,000	$73,000
Line 9	Years to recover ERP start-up costs (payback) 8 divided by 7			1.22	0.74	0.65	1.53	0.94

[a]Supervisor time/cost same for whatever program exists; other inspector costs linearly related to number of inspections.

[b]Annual operating costs include ERP Manager ($12,000 for 100; 18,000 for 250), staff person ($20,000 per year, fixed cost), data entry (intern—$1,500 for 100, $4,000 for 250), mailing costs for 750 facilities ($2,000). For sample size of 100: ERP Supervisor (10% of time of $120,000 annual salary/benefits), QA/QC data analysis, calculate post-ERP inspection proportions, Fisher/Bonferonni analyses, report writing and review; Data Entry—three weeks work, $1500.

For sample size of 250: ERP Supervisor (15% of time of $120,000 annual salary/benefits); Data Entry—eight weeks work, $4000 mailing costs of workbook and checklist to all 750 facilities, $2000 staff person used to organize and track all self-certification forms returned to DEM (20% of time of $80,000 annual salary/benefits).

[c]Funds equal to EPA grant received by RI to develop workbook, checklist and run six workshops ($67,000); 5% of ERP Supervisor time for baseline analysis ($6000)–will vary by state.

benchmarking component allowed investigators to obtain a statistically valid understanding of facility compliance levels prior to implementation of the sector-wide ERP or the U.S. Energy Policy Act of 2005—that is, these data were reflective of compliance levels (across the industry) achieved using traditional regulatory methods and a standard facility-by-facility, inspection-enforcement approach.

ERP technical assistance interventions consisting of stakeholder involvement, self-administered checklists, certified reporting of inspection results, and training and outreach appeared to have been effective in producing statistically significant performance improvements above baseline compliance levels. (Note: The self-education process that derives from reading workbook tutorials and conducting "detailed" walkthrough facility assessments have been found to result in significant performance improvements in a variety of environmental health settings [see footnotes 1–3]) Postintervention assessments conducted among UST facilities drawn at random from the original pool of regulated facilities, resulted in a separate and distinct set of compliance data that were reflective of sector-wide conditions three-years after baseline inspections were performed. The change or observed improvement in compliance among the postintervention facilities for 41 measurable indicators was found to range from 22% to 100%, with a median of 85%. The use of EPA OUST SOC metrics, standard ERP indicator analysis methods, and negative binomial regression analysis all produced the same result—statistically significant improvements in compliance performance post-ERP intervention.

A first order analysis of the costs associated with the traditional UST inspection program in RI was also performed and compared with the costs needed to support the alternate ERP approach. Several ERP implementation models were analyzed where both sample size and inspection frequency were varied and combined into five different scenarios. Because fewer inspections are required for ERP, costs associated with inspections were shown to be reduced for each scenario. Considering expenses to sustain ERP-related activities (workshops, data gathering, statistical analysis, and oversight), overall costs were still found to be lower than that for the more labor-intensive facility-by-facility inspection and enforcement approach mandated by the Energy Policy Act.

On the basis of Rhode Island's experience, it is anticipated that states having difficulty (due to limited funding or field resources, for example[6]) in meeting the U.S. Energy Policy Act three-year inspection requirement would particularly benefit from the cost efficiencies and statistical strengths of the ERP approach. In addition to adopting the ERP model, the integration of a risk-based approach (similar to that recommended by the U.K. Environment Agency in its report *Delivering for the environment: a 21st Century approach to regulation*[7]) where high risk facilities requiring more intervention are inspected more frequently and where the ERP model is applied to lower risk facilities, for example, might also be considered. Indeed, exploratory work for a risk-based approach was undertaken by EPA Region 3 in *The Mid-Atlantic States' MTBE Pilot Project* which developed "a GIS application and a site ranking software tool that can be used by EPA or states to consider relative potential risk when prioritizing UST facility inspections or Leaking Underground Storage Tank (LUST) site corrective action oversight"[8]; additional relevant information and data also exists.[9]

Regarding study limitations, some bias in the results may exist, as the timeframe for the study bracketed the 2005 Energy Act requirements. Observed performance improvements, however, were believed to be largely attributable to ERP-related activities as data trends could not be explained solely by the targeted inspections of neglected facilities conducted pursuant to the US Energy Policy Act of 2005 (i.e., the Act required facilities that had not been inspected since December 22, 1998, to undergo an onsite inspection by August 8, 2007, and subsequently once every three years thereafter). Supporting evidence included: (1) only 10% to 23% of all UST facilities (comprised entirely of neglected sites that had not been inspected for five to seven years) were subjected to targeted inspections in any given year (for the years 2005–2008)—which was consistent with a historical inspection rate of 17% (or once every six years) in Rhode Island, (2) comprehensive ERP interventions took place in 2005 with six statewide workshops attracting 297 attendees (no mention of the Energy Act requirements was made), telephone technical assistance, and self-audit workbook/checklist mailings to all federally regulated facilities, (3) strong enforcement follow-up of non-ERP certifiers/incomplete plan submittals was undertaken in 2005 and 2006 (i.e., 146 Notices

[6]In his article *A Long View: Parting Thoughts from a Veteran Tank Regulator*, Marshall Mott-Smith, a 31 year Florida Department of Environmental Protection veteran (with 22 years as administrator of the Storage Tank Regulation Section) wrote in 2008 that the future of groundwater protection "will not be easy, as many state resources are strained with lower tax revenues, and state legislators are faced with difficult funding decisions. Travel, training, and expense budgets have been slashed, and managers must prioritize their efforts on inspecting those facilities that pose the greatest risk to groundwater." *LUSTLine Bulletin* 58, September 2008, pp. 8,9.

[7]In the U.K. model, compliance and enforcement resources are concentrated "where the risks are highest, including the highest hazards and the poorest performing operators," with less intervention for low risk facilities. The report finds that: "Traditional regulation has achieved much. But the nature of regulation has to change to keep pace with changes in the economy and society. The Environment Agency is further developing its approach to regulation... This approach is outcome-focused and risk-based, clearly communicated and is delivered in a consistent manner. We call this modern regulation... Modern regulation aims to find the right balance—a proportionate, risk-based response, that will drive environmental improvements, reward good performance, but still provide the ultimate reassurance that tough action will be taken on those who fail to meet acceptable standards." Available at: http://www.environment-agency.gov.uk/static/documents/Business/delivering_1906007.pdf

[8]Report on *The Mid-Atlantic States' MTBE Pilot Project*, Public Water Supply Sampling, GIS Plotting of UST & LUST sites and Public Drinking Water Wells, and Ranking Tool Development. U.S. Environmental Protection Agency, Mid-Atlantic Region, Philadelphia, PA. May 2004. Available at: http://www.epa.gov/reg3wcmd/pdf/MTBE.pdf

[9]For example: *Florida's Cause of Leak Study*. Available at: http://www.neiwpcc.org/neiwpcc_docs/ustlust_shows/mott_smith.pdf; *Assessing the Risk of Groundwater Contamination from Petroleum Product Storage*, University of Missouri. Available at: http://extension.missouri.edu/publications/DisplayPub.aspx?P=WQ654

of Intent to enforce were sent to noncertifiers and more than 30 formal enforcement actions, 100 violations and $300,000 in penalties were assessed for noncompliance with ERP requirements), and (4) 1,291 RTC plans were submitted to DEM with commitment dates for returning to compliance.

Taken together, Rhode Island field data and project information indicate that the first round of Rhode Island's mandatory UST ERP was successful at producing statistically significant improvements in industry-wide compliance. Both categorical (for tank corrosion protection, piping leak detection, spill prevention and overfill protection, and groundwater monitoring wells and tank pad observation wells) and individual compliance inspection checklist indicator performance improvements were observed over time. Though these data represent 1st round results only, the weight-of-evidence—from prior ERP implementation experience in Rhode Island[10] and the positive results obtained by other states in other industry sectors[11]—shows that the ERP model compliments existing regulatory structures and can provide a sound alternative to traditional inspection programs. Due to the comprehensive and iterative nature of the UST ERP, it is believed that statewide compliance levels will increase over time as facilities become accustomed to the self-audit, certification, and disclosure/reporting components of the program.

We conclude that the ERP model—coupled with the evolution of improved technology requirements and new federal personnel training/operational standards—can produce measurable performance improvements, representing a cost-effective adjunct to traditional facility-by-facility inspection and enforcement programs for UST facilities. Application of the ERP model to the sizable universe of nonfederally regulated tanks would also be expected to result in improved compliance over time, while implementation among federally regulated facilities (the subject of this study) has the potential to allow regulatory agencies the flexibility to decrease the frequency of inspections among low risk facilities to less than once every three years (as currently required by the U.S. Energy Policy Act of 2005) without sacrificing compliance performance or increasing public health risks. Given "the compression of government funding in the United States" and the resulting potential negative impacts on human health and environmental protection—as recently highlighted by Editor-in-Chief Michael Greenberg in the March 2011 issue of *Risk Analysis*(11)—we believe that this method and type of approach is going to become necessary to manage dwindling state inspection resources.

Finally, the success of the Rhode Island UST ERP initiative may be attributed to a number of factors, especially the (1) collaborative efforts between EPA Headquarters/Region I and DEM's Offices of Waste Management (within which the UST regulatory program resides) and Customer & Technical Assistance (which championed ERP), (2) early stakeholder involvement through face-to-face meetings and information exchanges, (3) provision of technical assistance through the development of industry guidebooks and checklists followed by statewide industry training and outreach, and (4) research partnership with the University of Rhode Island Department's of Chemical Engineering and Computer Science and Statistics. Our analysis also found that the SOC results reported by the EPA OUST can be misleading. Before 2005, state compliance information was based on a relatively small number of targeted inspections and, therefore, was not statistically representative of national compliance rates as reported. Even with the increased inspection frequency mandated by the 2005 Energy Act, annual aggregated enforcement data reports may not accurately reflect the national rate of operational compliance.

[10]See: Enander RT, Gagnon RN, Hanumara RC, Park E, Armstrong T, Gute DM. Environmental Health Practice: Statistically Based Performance Measurement. *American Journal of Public Health,* 2007; 97(5):819–824; Dymova N, Hanumara RC, Enander RT, Gagnon RN. Statistically Speaking: Use of the Global Test Statistic as a Performance Measurement in a Reanalysis of Environmental Health Data. *American Journal of Public Health,* 2009; 99(10):1739–1741; Auto Salvage, Auto Body, and Exterior Lead Paint ERP descriptions—Available at: http://www.dem.ri.gov/programs/benviron/assist/index.htm

[11]See: ERP States Produce Results 2007 Report: States' Experience Implementing the Environmental Results Program. U.S. Environmental Protection Agency, National Center for Environmental Innovation. Available at: www.epa.gov/innovation. The increasing interest and growing body of ERP knowledge has led to the creation of the *States Environmental Results Program Consortium* in October 2006, which now includes a third of all States [Source: States ERP Consortium Fact Sheet, May 2008. Available at: www.erpstates.org].

ACKNOWLEDGMENTS

This work was funded by the U.S. Environmental Protection Agency's National Center for Environmental Innovation through State Innovation Grant number EI-97150001–0, "Underground Storage Tanks—Alternative Inspection Programs

and the U.S. Energy Policy Act of 2005." The findings and conclusions in this article are those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency.

APPENDIX: SAMPLE SIZE DETERMINATION, TEST FOR MULTIPLE COMPARISONS, AND NEGATIVE BINOMIAL REGRESSION

This appendix provides detailed descriptions of the methodologies used to derive sample size, test for multiple comparisons and negative binomial regression model fitting.

A.1. Sample Size

Frequently, a question arises relating to how many baseline and postintervention field audits are needed to statistically compare compliance proportions for a choice of alpha level and power. This was determined with the following equation, as described by Enander *et al.*[9]

$$n = 2\left[\frac{Z_{1-\beta}\sqrt{P_1(1-P_1)+P_2(1-P_2)}+Z_{1-\alpha}\sqrt{2\overline{P}(1-\overline{P})}}{\Delta}\right]^2,$$

where: p_1 = assumed baseline UST compliance rate proportion
p_2 = assumed post intervention UST compliance rate proportion,
$\Delta = |p_1 - p_2|$,
$Z_{1-\alpha}$ = significance level test statistic,
$Z_{1-\beta}$ = power test statistic
$\overline{P} = (P_1 + P_2)/2$.

(Note: Various sample size formulas exist for constructing confidence intervals with specified margins of error at a given confidence level or for testing hypotheses at given alpha and beta levels. Results for variations of related sample size formulas—all based on normal approximation—are typically reasonably close. The formula we chose is commonly used for the case of selecting equal sizes in the pre and post settings with stated alpha and beta levels for postintervention improvement hypothesis testing.)

Considering the above equation, available inspection resources, and potential study outcomes, we chose a target sample size of $n = 100$ facilities to be inspected at baseline and postintervention. Using an alpha level of 5% ($\alpha = 0.05$, $Z_{0.95} = 1.645$) and power of 80% ($Z_{0.80} = 0.842$), in agreement with conventional practice, our goal was to measure a minimum difference (Δ) of an approximate 15% performance improvement postintervention. Alternatively, the EPA *ERP Results Analyzer* demonstrated that for a difference in proportions of 15%, two samples of size $n = 100$ would be sufficient; also, EPA's *ERP Sample Planner* for estimating "sample sizes required for a specified margin of error for a difference in proportions (two samples)" showed that for a desired margin of error of ± 12–15% (at the 95% confidence level) a sample size of ~100 in each round of inspections was required.[12] Ultimately, we conducted 96 baseline and 93 postintervention inspections.

A.2. Test for Multiple Comparisons

In addition to checking each of the *p*-values against the nominal level of significance of 0.05 to determine which indicators showed improvement, a modified Bonferroni approach was used. The Holm's modified Bonferroni adjustment[13] for multiple comparisons was applied to indicators showing significance at the 0.05 level. This procedure avoids inflation of the Type I error rate and was carried out in two steps:

(1) *p* values were ordered from smallest to largest, and
(2) The smallest *p* value was compared against 0.05/*k* followed by comparison of the next smallest *p* value against 0.05/(*k*–1) and so on, until the largest value was compared to 0.05 (where *k* = total number of tests).

A.3. Model Fitting: Negative Binomial Regression Analysis

41 indicators were grouped into five compliance categories: tank corrosion protection (2 indicators, Category B), tank leak detection (12 indicators, Category C), piping leak detection (11 indicators, Category E), spill prevention and overfill protection (11 indicators, Category F), and groundwater monitoring and tank pad observation wells (5 indicators, Category I). To compare the overall compliance performance at baseline and postintervention in each of the five categories, the compliance score for facility and category was calculated. For example,

in the category "piping leak detection," the number of measurable indicators (a1) were at most 11. A facility was designated to be in compliance for a specific indicator if each of the tanks within that facility were in compliance; facilities were in noncompliance if one or more of tanks were noncompliant. Thus, the compliance score (a2) for a facility was "0" (noncompliant on all indicators) or "a1" (compliant on all a1 indicators).

The analysis of all data on a1 and a2 at baseline and postintervention was performed using two methods. Because the compliance score on each facility was count data, Poisson regression and negative binomial regression methods were considered to be viable—where a2 indicators were assumed to follow a Poisson distribution with the mean μ and variance μ, then $\text{Log}(\mu) = \beta 0 + \beta 1 \times 1 + \beta 2 \times 2 + \cdots + \beta n \times n$, and a Negative Binomial distribution with the mean μ and variance $\mu + \alpha\mu^{**}2$, then, $\text{Log}(\mu) = \beta 0 + \beta 1 \times 1 + \beta 2 \times 2 + \cdots + \beta n \times n$, respectively. The models were applied using SAS software, which required the data sets to be recreated in the appropriate format. For each of the facilities, compliance, noncompliance, and the total number of indicators were counted. For these analyses, the number of compliances was the dependent variable and the independent variable was a dummy variable identifying the baseline and post intervention settings. The number of indicators in each category was used as offset as they varied from facility to facility. To compare the fit of the models, the deviance, likelihood, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) measures were used. Between these two methods, negative binomial regression analysis was judged to be appropriate based on AIC. The p value for the regression coefficient compared compliance at baseline and postintervention.

REFERENCES

1. U.S. Environmental Protection Agency. Office of Underground Storage Tanks. Cleaning Up Underground Storage Tank System Releases. Last updated on Wednesday, December 07, 2011. Available at: http://www.epa.gov/swerust1/cat/index.htm, Accessed on January 28, 2009.
2. Rhode Island Department of Environmental Management. (2005). Pascoag Water District Environmental Response Plan. Available at: http://www.dem.ri.gov/programs/benviron/waste/ pascoag/factsht.pdf, Accessed on February 2, 2009.
3. US Environmental Protection Agency. Office of Underground Storage Tanks. FY 2008 Annual Report on the Underground Storage Tank Program. (2009). Available at: http://www.epa.gov/swerust1/pubs/OUST_FY08_Annual Report-_Final_3–19-09.pdf, Accessed on March 14, 2009.
4. Massachusetts Department of Environmental Protection. Environmental Results Program. (No Date Given). Available at: http://mass.gov/dep/service/about11.htm, Accessed on August 20, 2009.
5. Rhode Island Department of Environmental Management. Environmental Results Program Certification Checklist. (rev. 2008). Available at: http://www.dem.ri.gov/programs/benviron/assist/usterp/ pdf/ustcl04.pdf, Accessed on December 24, 2009.
6. Rhode Island Department of Environmental Management. Environmental Results Program Certification Workbook. (2004). Available at: http://www.dem.ri.gov/programs/ benviron/assist/usterp/pdf/ustwb04.pdf, Accessed on December 24, 2009.
7. Rhode Island Department of Environmental Management. UST Steering Committee Meeting Minutes. (2003). Available at: http://www.dem.ri.gov/programs/benviron/assist/usterp/pdf/052003mn.pdf, Accessed on November 5, 2009.
8. Rhode Island Department of Environmental Management. Rules and Regulations for Underground Storage Facilities used for Petroleum Products and Hazardous Materials. (2008). Available at: http://www.dem.ri.gov/pubs/regs/regs/waste/ustreg08.pdf, Accessed on January 25, 2010.
9. Enander RT, Gagnon RN, Hanumara RC, Park E, Armstrong T, Gute DM. Environmental health practice: Statistically based performance measurement. American Journal of Public Health, 2007; 97(5):1–6.
10. US Environmental Protection Agency. Office of Underground Storage Tanks. UST Performance Measures. (2008). Available at: http://www.epa.gov/OUST/cat/camarchv.htm, Accessed on March 12, 2009.
11. Greenberg M, Lowrie K. Editor's choice. Time to worry: Whither risk analysis and the compression of government funding in the United States? Risk Analysis: An International Journal, 2011; 31(3):344.
12. US Environmental Protection Agency. National Center for Environmental Innovation. Environmental Results Program Results Analyzer and Sample Planner. (2007). Available at: http://www.epa.gov/erp/toolsandresources.htm, Accessed on February 3, 2010.
13. Holm S. A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics, 1979; 6: 65–70.