Sampling Out: Regulatory Design and the Total Coliform Rule

Lori S. Bennear[†],
Katrina K. Jessoe[‡], and
Sheila M. Olmstead[‡*]

Nicholas School of the Environment, Duke University, Durham, North Carolina 27708,
School of Forestry and Environmental Studies, Yale University, New Haven, Connecticut 06511

Abstract

In this paper we investigate strategic non-compliance with the Total Coliform Rule (TCR) under the Safe Drinking Water Act. The structure of the TCR provides incentives for some water suppliers to avoid non-acute violations by taking additional samples. We estimate the prevalence of strategic over-sampling and its potential impact on violations using monthly data for more than 500 water suppliers in the Commonwealth of Massachusetts from 1993-2003. We find evidence that strategic over-sampling is occurring. There is more over-sampling among suppliers most likely to benefit from it, in particular suppliers for whom violations are determined by the percentage of positive samples and that have at least one positive sample. We estimate a significant number of violations that would have occurred if the supplier had sampled according to their sampling plan, but that were avoided by over-sampling. Regression analysis supports the hypothesis that over-sampling among suppliers likely to benefit lowers the probability of a violation. EPA is currently revising the TCR, so our analysis of potential impacts of strategic compliance with the current rule is particularly timely and suggests that alternative methods of monitoring bacteria should be considered.

[†] Nicholas School of The Environment, Duke University.
[‡] School of Forestry and Environmental Studies, Yale University
[*] Corresponding author phone: 203-432-6247; fax 203-432-0026; e-mail: sheila.olmstead@yale.edu

1. Introduction

Harmful bacteria in drinking water may cause immediate and serious health effects, particularly from acute gastroenteritis which causes cramps, diarrhea, fever, nausea and vomiting and can be deadly to vulnerable individuals. While the prevalence of acute gastroenteritis is difficult to quantify, one current estimate attributes 4.3-11.7 million cases of acute gastroenteritis per year to regulated U.S. drinking water systems (Colford *et al*. 2006).

The 1989 Total Coliform Rule (TCR), under the Safe Drinking Water Act (SDWA), establishes the current U.S. federal standards and sampling protocols for bacteria in drinking water systems. The TCR covers 54,000 community water systems, those serving 25 or more people, which provide piped drinking water to 264 million people (U.S. Environmental Protection Agency 2008). Water suppliers must adhere to minimum TCR sampling and treatment protocols, and state regulators periodically negotiate bacterial sampling plans with each supplier, which may be more stringent than federal requirements.

The TCR is the most frequently violated of any SDWA regulation. U.S. water suppliers incurred 8,310 monthly TCR violations per year, on average, between 1997 and 2003 (U.S. Environmental Protection Agency 2007). But monthly TCR violations do not necessarily represent public health risks. The presence of total coliforms suggests the possible presence of disease-causing agents in drinking water and can be considered as an indicator of treatment effectiveness and the integrity of distribution systems (World Health Organization 2004). While some coliform bacteria, such as *Escherichia coli* (*E. coli*) and fecal coliforms, may cause gasteroenteritis, many are harmless. The TCR requires systems with any positive total coliform samples to test further for dangerous bacteria, such as *E. coli* and fecal coliforms. But positive total coliform samples exceeding the TCR thresholds result in a non-acute monthly violation, even if subsequent diagnostic tests for *E. coli* or fecal coliform are negative.

Complying with the TCR is costly. The U.S. Environmental Protection Agency (EPA) estimated when the rule was introduced that compliance would cost water suppliers $210-230 million annually (in 2007 dollars), with monitoring costs comprising about 60 percent of annual costs (Wade Miller Associates 1989). In addition, water suppliers face potential financial penalties as well as public disapproval when they violate the TCR. For monthly TCR violations suppliers must notify the public of the violation within 14 days and must also include this violation in their annual Consumer Confidence Report to customers. These reporting requirements represent real costs for U.S. water suppliers, who have been shown to reduce violations of bacterial and other contaminant rules when they are required to disclose information about violations to the public (Bennear and Olmstead 2008). Given the substantial costs of TCR compliance and violations, and the uncertain public health benefits of avoiding monthly TCR violations, water suppliers may have an incentive to reduce the costs of TCR compliance and violations by strategically sampling so as to avoid violating the rule.

The economic literature on regulatory enforcement can be traced to Becker (1968), the first to model the process through which those evading the law might balance the benefits and costs of their actions. In the environmental realm, a number of contributions have focused on static analyses of strategic response to enforcement (Downing and Watson 1974, Harford 1978, Viscusi and Zeckhauser 1979), and dynamic models, in which firms and regulators react to each others' actions over time (Harrington 1988, Heyes and Rickman 1999, Livernois and McKenna 1999, and Raymond 2004). The propensity of regulated entities to evade the detection of violations has been demonstrated in contexts as diverse as the personal income tax (Andreoni *et* *al*. 1998), safety regulation of U.S. nuclear power plants (Feinstein 1989), national water quality regulation of U.S. pulp-and-paper plants (Helland 1998), and new source review standards for power plants under the Clean Air Act (Keohane *et al*. 2008). In addition to the potential hazards imposed on society by regulatory avoidance, such efforts to avoid regulatory oversight or penalties are socially wasteful (Malik 1990, Innes 2001).

Our research investigates whether water suppliers respond to the current incentives of the TCR by sampling strategically to avoid monthly TCR violations, an activity referred to by regulators and water suppliers as "sampling out." We estimate the prevalence of sampling out and its potential impact on violations using monthly data for more than 500 water suppliers in the Commonwealth of Massachusetts from 1993-2003. We find evidence that sampling out is occurring. There is more over-sampling among suppliers most likely to benefit from it (in terms of reducing violations). There are a significant number of violations that would have occurred if the supplier had sampled according to their sampling plan, but these violations were avoided by over-sampling. Regression analysis supports the hypothesis that over-sampling among suppliers lowers the probability of a violation. EPA is currently revising the TCR, so our analysis of potential impacts of strategic compliance with the current rule is particularly timely and suggests that alternative methods of monitoring bacteria violation could lower costs by avoiding wasteful strategic sampling without impacting public health.

2. Experimental section

2.1 Regulatory Background

Because coliform contamination can happen anywhere within the distribution system, water suppliers are required under the TCR to collect a minimum number of samples from throughout the distribution system each month. The minimum number of samples is set by federal statute and increases with system size (see Figure 1). In addition, in Massachusetts, water suppliers negotiate a sampling plan with state regulators at least once every three years. The number of samples in the sampling plan may exceed the federal minimum. In Massachusetts between 1993 and 2003, routine sampling required by state regulators exceeds the federal minimum in 27% of supplier-months (where a supplier-month is an observation on a single community water supplier in a given month). Moreover, water suppliers often take more samples than required by their state-approved routine sampling plan, plus necessary repeat samples. This "over-sampling" occurs in roughly 50% of supplier-months.

Violations of the TCR fall into two categories—acute and monthly (non-acute). *Monthly* TCR violations are based on the number of positive samples for total coliform, only, and the relationship between positive coliform samples and violations varies with the number of samples taken in a month. Water suppliers that take more than 40 routine and repeat samples in a month violate the TCR if more than 5 percent of those samples test positive; we refer to this as the 5% rule. Systems taking fewer than 40 monthly routine and repeat samples violate the TCR if two or more of those samples test positive; we refer to this as the 2P rule. Since routine sampling is based on system size, larger systems tend to fall under the 5% rule, and smaller systems tend to fall under the 2P rule, but the relevant rule for any particular system can vary by month, since compliance is determined by samples taken, not required samples. Once a positive sample is drawn, the TCR requires suppliers to conduct a specified number of repeat samples upstream and downstream of the original positive site. If any repeat sample tests positive for *E. coli* or fecal coliforms, or if a routine sample tests positive for *E. coli* or fecal coliform, and any repeat sample tests positive for total coliforms, the system incurs an *acute* TCR violation.

Two components of the monthly maximum contaminant level (MCL) for total coliform, the focus of our study, make strategic behavior possible. First, violations are determined by the number of samples actually taken, rather than the number of samples required either by federal law or by state sampling plans. Second, once more than 40 samples are taken the violation is determined by the percentage of total samples that are positive rather than absolute numbers. Water system operators know more about the likelihood of drawing a positive sample in different parts of the system and at different times of the month, or even times of day, given the timing of disinfection measures. Thus, it is possible that suppliers can strategically locate or time additional samples to remain below a violation threshold.

Of course, even though this type of strategic behavior is possible, suppliers may not routinely engage in this behavior. Furthermore, simply observing over-sampling does not indicate that suppliers are engaged in strategic over-sampling. Over-sampling may be beneficial. Taken randomly, more samples would provide a more complete picture of the presence of bacterial hazards in a drinking water distribution system. Thus, over-sampling water suppliers might reasonably argue that they are over-complying with the TCR, incurring greater testing costs to protect public health.

However, anecdotal evidence suggests that strategic over-sampling, or sampling out, does occur. Figure 2 offers an example of potentially problematic "sampling out", as we define it. We obtained this record of one Massachusetts water supplier's microbiological analysis for the months October 2001-September 2002 from the files in the supplier's regional DEP office. The record shows that from October 2001-June 2002, the supplier took more samples than required by its sampling plan (30 samples per month). All samples tested "absent" (A) for total coliform during this period, so the extra samples cannot be repeat samples, and with no positive samples, this is also not strategic over-sampling. We might say that the supplier over-complied with the TCR in October 2001-June 2002.

In July 2002, the system drew 43 samples, resulting in two positives, but because its sample size was so large, it remained below the 5% threshold (to which it was subject, since it took more than 40 samples) and did not violate the TCR. In August, the system took 48 samples, but with four positives, it did exceed the 5% threshold. The system's sampling decisions, in consultation with the regional regulator, are documented in "sticky notes" attached to the file record which clearly indicate a desire to avoid a July 2002 monthly TCR violation. July and August are months in which the probability of positive total coliform samples is elevated for systems using surface water (as this one does). It is evident that if this system had not sampled out in July 2002, it would have violated the TCR, as it did in August. Our analysis attempts to distinguish over-compliance from strategic sampling that lowers the probability of a violation to better understand if sampling out is a significant problem with the TCR or merely anecdotal.

2.2 Methods

We use three different statistical approaches to differentiate strategic over-sampling from over-compliance. First, we examine whether over-sampling is more frequent among suppliers for whom it is most likely to reduce violations. Because the threshold for violating the TCR depends on the number of samples taken in a month, not all suppliers are equally likely to benefit from taking more samples. Suppliers who take fewer than 40 samples are subject to the 2P rule. Once two positive samples have been drawn, there is no benefit (no reduction in the chance of a violation) from taking more samples, as long as fewer than 40 samples are taken. However, a supplier who routinely takes fewer than 40 samples, if it draws two positive samples in a month, may strategically decide to continue sampling so that total samples exceed 40 and a violation will be determined under the 5% rule. We refer to this group as "jumpers." For jumpers, the possibility of strategically locating and timing samples to get negative results may move the supplier from violation status under the 2P rule to non-violation status under the 5% rule. (The system in Figure 2 is a jumper in July and August 2002, but avoids a violation only in July.)

Additionally, suppliers who routinely take more than 40 samples per month, but in the course of routine sampling exceed the 5% rule, can benefit from strategically sampling to ensure additional negatives, reducing the percent positive below 5%. We refer to this group of suppliers as "dodgers." Thus, the first component of our analysis compares rates of over-sampling in three groups of suppliers—those that always take fewer than 40 samples, those that sometimes take more and sometimes take fewer than 40 samples (potential jumpers), and those that always take more than 40 samples (potential dodgers). If over-sampling is strategic we would expect to see relatively more over-sampling in the latter two groups. We would also expect to see relatively more over-sampling among potential jumpers and potential dodgers during months in which a system has drawn at least one positive coliform sample, since it is only suppliers with at least one positive draw for whom over-sampling may have some strategic benefit.

Evidence of relatively more over-sampling by potential jumpers and potential dodgers, and during months in which at least one positive sample is drawn, is not sufficient to show that over-sampling is done strategically to avoid violations, however. The economics literature suggests that larger firms are more likely to over-comply with environmental regulations (Arora and Cason 1995, Videras and Alberini 2000), so more over-sampling among larger suppliers might still be a sign of over-compliance with environmental regulation. The second step in our analysis examines whether significantly more violations would have occurred if suppliers had not over-sampled. If suppliers are over-sampling to obtain better information, we would not expect over-sampling to reduce violations. If we find that over-sampling suppliers would have incurred a monthly TCR violation had they maintained their routine and required repeat sampling schedule, this can be considered evidence that over-sampling is strategic.

The final stage of our analysis extends this line of reasoning by using regression analysis to examine whether over-sampling lowers the probability of a violation while controlling for other factors that influence violations and may vary by supplier and over time. We begin by estimating a pooled linear probability model (1), in which V_{it} is a binary variable equal to one if supplier *i* incurred a monthly TCR violation in month *t*, and zero if not; $oversample_{it}$ is equal to one if supplier *i* over-sampled in month *t*, and zero if not; $imputed_{it}$ is equal to zero if the data for this supplier-month reflect the actual number of positive samples, or one if we have imputed this information; and we assume errors (ε_{it}) are uncorrelated within a supplier over time.

$$V_{it} = \beta_0 + \beta_1 oversample_{it} + \beta_2 imputed_{it} + \varepsilon_{it} \tag{1}$$

We then control for whether a system has at least one positive sample in a month ($anyp_{it}$), since the presence of at least one positive sample creates an incentive for suppliers to strategically over-sample. However, the same underlying process that generates positive samples in a given month likely also generates violations, making *anyp* endogenous. Thus, we estimate a two-stage-least-squares model, instrumenting for *anyp* using a one-month lag and estimating the parameters of the violations model using predicted *anyp* from the first stage (2).

$$\begin{aligned} anyp_{it} &= \gamma_0 + anyp_{it-1} + \nu_{it} \\ V_{it} &= \beta_0 + \beta_1 oversample_{it} + \beta_2 imputed_{it} + \beta_3 \widehat{anyp}_{it} + \beta_4 oversample_{it} * \widehat{anyp}_{it} + \varepsilon_{it} \end{aligned} \tag{2}$$

We then estimate a series of panel linear probability models to control for supplier heterogeneity in the propensity to violate, also adding a series of year dummies, d_t, to control flexibly for any trend in violations over time, and again instrumenting for *anyp* (3).

$$V_{it} = \beta_0 + \beta_1 oversample_{it} + \beta_2 imputed_{it} + \beta_3 \widehat{anyp}_{it} + \beta_4 oversample_{it} * \widehat{anyp}_{it} + \sum_{t=1}^{T} \lambda_t d_t + u_i + \varepsilon_{it} \qquad (3)$$

The error structure in (3) includes u_i, a supplier heterogeneity parameter, and ε_{it}, the idiosyncratic error term. We estimate models in which u_i is a supplier random effect, and also one in which u_i is a supplier fixed effect.

We also estimate (3) adding additional control variables that may influence the probability of a violation: *summer*, equal to 1 in July, August and September and zero otherwise (since coliform levels tend to be elevated during these moths); and *jump_dodge*, which identifies suppliers who are either sometimes or always under the 5% rule (equal to 1), or always under the 2P rule (equal to 0), interacting this with *oversample* and *oversample*anyp*. We also control for any difference between a supplier's federal minimum sampling requirement and their state sampling plan. Suppliers may reduce violations by successfully negotiating with regulators to establish a sampling schedule less likely to produce violations. That is, strategic over-sampling may occur in part during the month in which sampling actually takes place, and in part well in advance, when the supplier negotiates with the state regulator about how many samples should routinely be taken, and from which sites. The variable *feddeviation* will capture any effect on violations of this *ex ante* negotiation.

2.3 Data

With the help of the Massachusetts Department of Environmental Protection (DEP), we collected data on Massachusetts water suppliers' total coliform sampling and violations between

1993 and 2003. We were able to obtain these data for a total of 559 community water suppliers. Because total coliform maximum contaminant level (MCL) enforcement is based on monthly statistics, we created a panel of 55,993 supplier-months. We sought monthly data on the number of total coliform samples drawn by each system, the number of routine samples required for each system (the state sampling plan), the number of positive coliform samples, and whether the supplier violated or not.

All of this information was obtainable from the state DEP offices in Boston, except for the last item. The DEP headquarters does not retain records of the number of positive coliform samples in each month – only whether *at least one* sample in a month tested "present" (P) or all were absent (A). Thus, for all Massachusetts systems each month, we know the number of samples taken, whether the aggregate results included at least one positive, or none, and whether they violated the TCR. This information was insufficient to test for strategic over-sampling. By contacting the four regional DEP offices, we were able to obtain paper records of the number of positive samples for a portion of our full observation period for 245 suppliers: 133 in the Western region, partial records between 1993 and 1997; and 112 in the Northeast region, partial records between 1997 and 2003. For a set of 13,970 supplier-months, we have these full data on coliform samples and complete test results.

Some of our statistical analysis proceeds using only the system-months for which full data are available. In other cases, we impute the number of positive samples for supplier-months in which these data were not available from the regional DEP offices. Figure 3 describes our imputation method. For suppliers with no positive coliform observations, our imputed number of positive draws is set to zero. For suppliers with at least one positive draw, the imputation method varies depending on the violation rule relevant for the supplier in that month, and on whether we observe a violation by the supplier in that month. Suppliers with a monthly TCR violation are assumed to have stopped sampling once they violated, so those subject to the 2P rule with a violation are assumed to have two positives, and those subject to the 5% rule, exactly the number of samples that would have put them over this threshold. The group of greatest interest are the suppliers at the bottom of Figure 3 – those who have at least one positive sample, but do not violate the TCR. This is the group in which we would expect to predict unobserved violations if suppliers are successfully sampling out. Suppliers without a violation and subject to the 2P rule are assumed to have a single positive sample. Suppliers subject to the 5% rule with at least one positive sample, but no violation, are assumed to have the largest number of positive samples that will keep them under the violation threshold.

To the extent that our imputation method estimates positive samples that are different from the actual number of positive samples, we would expect to underestimate the number of positives for those who violate (the middle two groups in Figure 3) – they may keep sampling either to obtain better information about the source of contamination in the distribution system, or in an unsuccessful effort to sample out. We expect to overestimate the number of positive draws for the final group, those subject to the 5% rule who do not violate, because we assume they are getting just to the 5% threshold without crossing it.

We examine the accuracy of our imputation method in a variety of ways. First, how accurately do we predict the number of positive coliform samples in a supplier-month? In Table 1, we compare imputed and actual positives within those supplier-months for which we can make this comparison – those for which the full sampling data were available from the regional offices. Consider the first row in Table 1, which compares the actual number of positives (*actual_P*) to our imputed number (*imputed_P*). For the full sample, the mean of imputed positives (0.096) is very close to the mean of actual positives (0.094). The distribution of *imputed_P* is tighter with a smaller standard deviation and a lower maximum (and equal minimum of zero). This is to be expected because the imputation method necessarily clusters *imputed_P* between 0 and 2 for small suppliers governed by the 2P rule and around 5% of total samples for suppliers governed by the 5% rule. This imposed clustering decreases the spread in the number of positive samples. Figure 4 compares the mean number of actual positive samples to the mean number of imputed positive samples over time and this comparison suggests the imputation method works fairly well. This is reassuring, but it is largely due to the fact that positive samples are uncommon. Our imputation method perfectly predicts zero positive samples, so when a system has no positive samples in a month, *imputed_P* always equals *actual_P* (see row 2 of Table 1).

What if we look more closely at the supplier-months with some positive samples? Here again, the actual (2.60) and imputed (2.66) mean number of positive draws are very close, though the distribution is tighter for *imputed_P*. If we look at the four groups defined at the bottom of Figure 3, we can see that we do better for some groups than for others. As expected, we under-predict positive samples for suppliers with a monthly TCR violation. Suppliers under both rules may continue to sample after reporting a violation in order to obtain better information about the location and extent of contamination, or they may unsuccessfully attempt to sample out, and violate. For those with no monthly TCR violation under the 2P rule, we under-predict positive samples; it appears that in some cases, suppliers under the 2P rule obtain two or more positive samples, but for some reason do not report a violation to the DEP (mean positive samples for this group are greater than one). As expected, we also over-predict the number of positive samples for those with no monthly TCR violation under the 5% rule, given our assumption that they would obtain the number of positive samples that keeps them just below the threshold.

Perhaps more important than how well our imputation method predicts positive samples is how well it predicts expected monthly TCR violations relative to the actual data. In Table 2, we predict violations from our imputed positive samples and compare those to predicted violations using the actual number of positive samples. The results of our over- and under-predictions of the number of positive samples, to the extent that they result in mistakes predicting expected violations, show up in the last two columns of Table 2: cases in which *imputed_P* indicates that we would not expect a violation, but *actual_P* indicates that we would (column 3), or the reverse is true (column 4). Our imperfect imputation method looks "less imperfect" in this light. The only group of suppliers for which we would significantly mis-predict whether suppliers are above or below the violation threshold using *imputed_P*, relative to using *actual_P*, is the group under the 2P rule with at least one P, but no violation. For this group, 15% of the time, we impute a number of positive draws that suggests suppliers would not violate, but the actual number of positive draws suggests that they would. This is the anomalous group identified in the discussion of Table 1 – suppliers under the 2P rule who obtain two or more positive samples, but for some reason do not report a violation to the DEP. We assumed one positive sample in these supplier-months, but *actual_P*>1 in 27 supplier-months for this group. For all other groups, *imputed_P* and *actual_P* differ in terms of predicting expected violations for only 1-6% of supplier-months.

No imputation method is perfect. However, given the records available to us for this analysis, we believe this is the best and most defensible imputation possible. While the use of imputed data to estimate the variable *oversample* introduces measurement error, errors in imputation are small and are unlikely to introduce strong bias. To further ensure that our analysis is not driven by imputation problems, we perform our analysis both for the subset of supplier-months in which complete records are available (using *actual_P*), and for all supplier-months, using *actual_P* for supplier-months in which it is available and *imputed_P* where it is not or when separate analysis is not feasible we condition estimates on whether data are imputed or not.

3. Results

3.1 Over-sampling by Supplier Category

We measure over-sampling as the difference between the actual number of samples taken in a month and the number of samples specified in the state-negotiated sampling plan, less any required repeat samples from positive draws. This captures the number of samples that the water supplier is voluntarily taking beyond what is required by federal and state law. It is important to note that if we observe over-sampling, this, in itself, may be illegal behavior. Suppliers may not deviate from their monthly sampling plan without the specific approval of state regulators. The "sticky note" on Figure 2 suggests that, at least some of the time, suppliers take additional samples with regulators' approval. But this may not always be the case.

We use two different measures of over-sampling intensity. The first is simply an indicator variable that takes a value of one if the water supplier over-sampled in a particular month. The second is the number of extra samples taken. These measures of over-sampling are obtained for all supplier-months and also broken down into three supplier categories—suppliers who are always governed by the 2P rule, potential jumpers, and potential dodgers. (We consider a supplier a potential dodger if it always takes more than 40 samples, and an "always 2P" system if it always draws fewer than 40 samples. A supplier is a potential jumper if its total number of monthly samples is sometimes greater than, and sometimes less than 40.)

Note that because this part of the analysis relies on the number of positive coliform samples obtained in a given month in order to calculate the number of repeat samples required by the TCR (and thus the number of total samples required), we present results separately for supplier-months in for which we have actual numbers of positive samples (Table 3, rows 1 and 2) and for all supplier-months, using actual and imputed data (Table 3, rows 3 and 4).

The pattern of over-sampling is similar for the two sets of data. A much larger percentage of potential jumpers and dodgers over-sample. Using the regional data (Table 3, row 1), only 16% of suppliers always governed by the 2P rule over-sample, compared to 76% of jumpers and 84% of potential dodgers. Using actual and imputed data for the whole state (Table 3, row 3), 38% of suppliers always governed by the 2P rule over-sample, while 83% of jumpers and 89% of potential dodgers over-sample.

The number of additional samples taken is also greater for potential jumpers and dodgers. Using the smaller regional dataset, suppliers always under the 2P rule take, on average, fewer than 4 extra samples, conditional on over-sampling, while jumpers take an average of 8 additional samples and those always under the 5 percent rule take an additional 14 samples (Table 3, row 2) (average numbers of extra samples are 3, 10 and 17 for the same respective groups using data for the whole state as shown in Table 3, row 4).

The data from Table 3 suggest that over-sampling is occurring more frequently among the groups of suppliers that are more likely to reduce violations by over-sampling – those for whom expected marginal benefits of over-sampling are higher. However, there are other plausible explanations for this finding. For example, the marginal costs of additional sampling may well be lower for larger suppliers. In months where there is some reason to be concerned about water quality a purely socially-interested water supplier might be expected to take more samples and the frequency and quantity of these extra samples should be inversely related to their marginal costs. So finding more over-sampling among larger suppliers does not necessarily mean that this over-sampling is strategic. The next two sections present analysis designed to further distinguish strategic over-sampling from over-compliance.

3.2 Over-sampling and Violation Frequency

If over-sampling is the result of over-compliance by water suppliers then we would not expect this behavior to systematically lower the probability of a violation in the current month. Strategic over-sampling is defined as over-sampling that results in a lower probability of violation without reducing health risks. To distinguish strategic over-sampling from over-compliance we estimate violation frequency assuming that each water supplier only took the number of samples required in its negotiated sampling plan. This "predicted violation frequency" is equal to the number of positive samples divided by the number of required samples. The predicted violation frequency is then compared to actual violations, based on the number of positive samples divided by the actual number of samples taken.

There are four possible results. First, a supplier may have no violation and we predict they would not have had a violation even if they adhered to the sampling plan. Second, there may be an actual violation where we do not predict one. Third, there may be an actual violation where we do predict one. And finally, there may be no actual violation where we predict that the supplier would have violated, had it adhered to its sampling plan. The final category is of the most interest as it is most suggestive of strategic over-sampling. The calculations are done separately for suppliers that over-sample and those that do not. Because the calculation of violations depends on the number of positive samples, we do the analysis separately for the regions and time periods for which we have actual positive sample data.

We present these findings in Table 4, reporting results for the actual data, only, at the top, and for the actual and imputed data for the whole state at the bottom. The first thing to notice is that 99% of the observations have predicted violations equal to actual violations, for both samples. That is, the vast majority of supplier-months would have experienced no change in violation status if the supplier had taken the required number of samples. However, among the 94 supplier-months (in the regional data) where violation status with required sampling differs from violation status with actual sampling, 91 of these were cases where a violation would have occurred if the supplier had taken the required number of samples, but with the actual number of samples there was no violation. Furthermore, the majority of these cases (70 of 91 or 77%) are for supplier-months in which over-sampling did occur. These are the strategic over-sampling cases. The picture is similar looking at data for the whole state; where predicted violation status diverges from actual violations, most (64%) are cases in which we would predict a violation, but none occurs. And 94% of these instances occur in supplier-months characterized by over-sampling.

To get a better sense of the frequency we report these results in Table 5 only for supplier-months with at least one positive sample, since it is this group in which strategic over-sampling is likely. Once again for the majority of supplier-months (81% for the regional sample, and 73% for the full sample) there is no difference between violations and predicted violations. For the 31,581 supplier-months where over-sampling occurred, there were 1,531 supplier-months in which firms drew at least one positive sample. The over-sampling resulted in a potential violation status change in less than 2% of cases (518 supplier months). Of those 518 status changes, 325 observations were cases where there was no actual violation, but there would have been a violation if the supplier had adhered to their sampling plan. These 325 supplier-months represent 21% of the 1,531 supplier-months with at least one positive coliform sample. (If we do the same exercise with the regional data only, status changes in which there was no actual violation, but would have been under adherence to the supplier's sampling plan would occur in 20% of months with at least one positive sample.)

To get a sense of the magnitude of the violations avoided by strategic over-sampling, we can compare the total number of actual violations to our estimates of the number of violations potentially avoided by over-sampling. There were 900 violations of the TCR in our Massachusetts sample from 1993-2003, and we estimate an additional 325 violations that were avoided by suppliers that over-sampled and had at least one positive sample. That is, there could be as many as a third again as many violations of the TCR relative to what is actually reported. Extrapolating that ratio to the national level where there are 8,310 monthly (non-acute) TCR violations per year, on average, these results suggest that as many as 3,000 violations might go undetected each year due to strategic over-sampling.

The analysis presented in this section is suggestive of strategic over-sampling, but other explanations for these findings are possible. In particular, we are not controlling for other factors that might be correlated both with the decision to oversample and also with the likelihood of violation. In the next section, we use regression analysis to examine the likelihood of violation as a function of over-sampling holding other supplier and temporal characteristics constant.

3.3 Regression Analysis

In analyzing the effects of over-sampling on monthly TCR violations econometrically, we face two significant challenges. First, there is very little variation in the dependent variable – violations occur on only 1.3% of supplier-months in Massachusetts between 1993 and 2003. Second, violations may arise due to a complex set of factors that vary by supplier and over time, so any model that tells a convincing story about the effects of over-sampling on violations will include a large number of control variables, which will soak up a good deal of the variation in violations that we may hope to explain with over-sampling. These problems would be less severe with additional data, but with our Massachusetts data, alone, it may be difficult to identify a significant effect of over-sampling on violations once we control for confounding factors.

Table 6 reports the results of our econometric models. Given limited variation in violations, we use the full state sample in all models, controlling for imputed data by including a dummy variable equal to 1 for supplier-months where the number of positive samples is imputed. In column 1, we estimate equation (1). This essentially reproduces the results at the bottom of Table 4. Not surprisingly, we estimate a positive and significant coefficient on *oversample*; there are many more violations during supplier-months with over-sampling (757) than those without (143), according to Table 4. We then estimate equation (2), controlling for the presence of at least 1 positive sample, using reasoning similar to that in Table 5, since it is only suppliers with at least one P for whom over-sampling may have some strategic benefit. Conditional on having at least one positive sample, we estimate a negative (but statistically insignificant) coefficient on *oversample*, but not the expected negative coefficient on the interaction between *oversample* and *anyp*.

In column 3, we estimate equation (3), controlling more carefully for other factors that may affect the probability of a violation. Though the over-sampling effect is still not statistically significant, we obtain the expected negative signs for over-sampling, and for over-sampling with at least one positive sample. As expected, having at least one positive sample strongly increases the probability of a violation. We control for additional factors in column 4. Results indicate that suppliers who are at least sometimes under the 5% rule (jumpers and dodgers), and have at least one positive coliform sample in a month, may reduce the probability of a violation that month by over-sampling – precisely the group we would expect to benefit from strategic over-sampling. In this model, *jump_dodge* also captures some strategic variation in sampling, since the extra samples that make a supplier a jumper may be repeat samples, random extra samples, strategic extra samples, or some combination of these. The linear coefficient on this variable is negative, as well. As expected, *summer* and *anyp* both increase the probability of a violation. Having a sampling plan that exceeds the federal minimum number of samples does not appear to decrease the likelihood of a violation; unlike over-sampling, these extra samples are always negotiated directly with regulators, reducing the likelihood that they are successfully strategic.

Finally, we estimate a fixed-effects model in column 5. The coefficient estimates in this model are all identified only off of variation in violations within suppliers over time, a significant challenge to the data with so little variation in violations either within or between suppliers. We can no longer identify the linear effect of *jump_dodge* in this model, since it varies only by supplier and not over time. Otherwise, the model results are quite similar to those in column 4, except, importantly, the negative but now insignificant coefficient on the interaction between *oversample*, *jump_dodge*, and *anyp*.

4. Discussion

Our empirical results for the econometric models reported in Table 6, as well as the analysis of summary statistics in Tables 3-5, suggest that: (1) over-sampling occurs on a large scale in Massachusetts, supporting the anecdotal evidence in Figure 2; (2) water suppliers most likely to benefit from strategic over-sampling (those for whom it has the greatest likelihood of reducing the probability of a violation) are most likely to over-sample; (3) additional monthly TCR violations may have occurred among Massachusetts water suppliers between 1993 and 2003, had suppliers adhered to state sampling plans rather than over-sampling; and (4) over-sampling may systematically reduce the probability of a violation for suppliers most likely to engage in it for that purpose, controlling for other factors.

The magnitude of the potential violations avoided by over-sampling is significant. There were a total of 900 violations in Massachusetts in our sample, and we estimate that there may have been an additional 325 violations avoided by over-sampling. Extrapolating that to the national level suggests that as many as 3,000 violations are avoided each year by strategic over-sampling.

EPA is currently revising the TCR, so our analysis of potential impacts of strategic compliance with the current rule is timely. Our results suggest that suppliers may spend significant resources avoiding monthly TCR violations by over-sampling. The health impacts of this violation-avoidance are unknown because there is limited scientific evidence linking total coliform (as opposed to fecal coliform or *E. coli*) to health outcomes, but the rule warrants revision regardless of EPA's view of these health impacts.

If EPA believes there is a direct link between total coliform and health impacts, then significant violation avoidance is undesirable from a human health standpoint. In this case, the regulation of total coliform should be revised to reduce the possibility of strategic violation avoidance by making violations contingent upon sampling plans rather than actual samples taken, or by making violations a function of the absolute number of positive samples rather than percentages. Either method would eliminate the incentive to strategically over-sample.

If EPA believes there is limited connection between total coliform and health impacts, then significant resources are being wasted avoiding violations that have minimal impact on public health. In addition, those suppliers who do not avoid violations bear costs from violations that do not represent real threats to public health. These violations must be disclosed to customers twice—once within 14 days and again as part of the supplier's annual Consumer Confidence Report. If there is no scientific link between total coliform and health impacts then the rule needs to be revised to ensure that failures of the total coliform screening test do not trigger violations unless further diagnostic tests reveal more significant problems.

Acknowledgments

This research was funded by the National Science Foundation (#SES-0648256) and the Hixon Center for Urban Ecology, Yale University. Damon Guterman and Andrew Durham at the Massachusetts Department of Environmental Protection (DEP) provided invaluable data and advice. We are also grateful for the assistance of staff in the Western and Northeastern regional offices of the Massachusetts DEP, especially Rick Larson. For comments, we thank participants at the National Bureau of Economic Research 2008 Summer Institute, Camp Resources XV, the Yale Environmental Economics Seminar, and the Duke Law and Economics Workshop. The authors, alone, are responsible for any errors.

Table 1. Comparison of imputed and actual positive coliform samples

	Variable by Supplier Group	Obs.	Mean	Med.	Std. Dev.	Min.	Max.
(1)	All supplier-months with full data						
	actual_P	13970	0.094	0	0.808	0	33
	imputed_P	13970	0.096	0	0.687	0	23
(2)	Supplier-mos. with no positive samples						
	actual_P	13464	0	0	0	0	0
	imputed_P	13464	0	0	0	0	0
(3)	Supplier-mos. with ≥1 positive sample						
	actual_P	506	2.603	1	3.392	1	33
	imputed_P	506	2.662	1	2.494	1	23
	With a monthly TCR violation						
	under the 2P rule						
	actual_P	47	4.021	3	2.674	1	12
	imputed_P	47	2	2	0	2	2
	under the 5% rule						
	actual_P	31	8.903	7	5.647	2	28
	imputed_P	31	3.806	3	1.138	3	7
	With no monthly TCR violation						
	under the 2P rule						
	actual_P	185	1.378	1	1.289	1	11
	imputed_P	185	1	1	0	1	1
	under the 5% rule						
	actual_P	243	2.457	2	3.284	1	33
	imputed_P	243	3.909	3	2.982	2	23

Table 2. Comparison of imputed and actual violations

Supplier group	Obs.	Imp.: Act.:	Fraction of predictions (1) Viol. Viol.	 (2) No No	 (3) No Viol.	 (4) Viol. No
All supplier-months with full data	13,970		0.006	0.997	0.003	0.000
Supplier-months under 2P rule						
with a violation	47		0.957	0.000	0.000	0.043
with one P, no violation	185		0.000	0.854	0.146	0.000
with no Ps	10,375		0.000	1.000	0.000	0.000
Supplier-months under 5% rule						
with a violation	31		0.968	0.000	0.000	0.032
with at least one P, no violation	243		0.045	0.881	0.062	0.012
with no Ps	3,089		0.000	1.000	0.000	0.000

Table 3. Over-sampling among suppliers

	Variable by Supplier Group	Obs.	Mean	Med.	St. Dev.	Min.	Max.
Suppliers with full data available							
(1)	Over-sampling (0/1)						
	All supplier-months	13970	0.45	0	0.50	0	1
	Always under 2P rule	7561	0.16	0	0.36	0	1
	Jump between 2P and 5 percent	4251	0.76	1	0.43	0	1
	Always 5 percent rule	2158	0.84	1	0.36	0	1
(2)	Extra samples (#) if over-sampling						
	All over-sampled supplier-months	6108	8.83	6	10.62	1	297
	Always under 2P rule	1049	3.69	2	3.20	1	27
	Jump between 2P and 5 percent	3239	7.60	5	7.63	1	106
	Always 5 percent rule	1820	14.00	10	14.98	1	297
All suppliers							
(3)	Over-sampling (0/1)						
	All supplier-months	55993	0.57	1	0.50	0	1
	Always under 2P rule	33511	0.38	0	0.49	0	1
	Jump between 2P and 5 percent	18113	0.83	1	0.38	0	1
	Always 5 percent rule	4369	0.89	1	0.31	0	1
(4)	Extra samples (#) if over-sampling						
	All over-sampled supplier-months	20664	8.55	4	14.35	1	677
	Always under 2P rule	7084	2.62	2	2.45	1	33
	Jump between 2P and 5 percent	10614	10.17	6	13.45	1	328
	Always 5 percent rule	2966	16.88	11	24.70	1	677

Table 4. Predicted potential "missing" violations, over-samplers vs. others

Supplier group	Obs.	Imp.: Act.:	(1) Viol. Viol.	(2) No No	(3) No Viol.	(4) Viol. No
			Number of predictions			
Suppliers with full data available						
All supplier-months	13,970		75	13,801	3	91
Supplier-months with over-sampling	6,241		43	6,127	1	70
Supplier-months with no over-sampling	7,729		32	7,674	2	21
All suppliers						
All supplier-months	55,754		704	54,701	196	346
Supplier-months with over-sampling	31,581		564	30,691	193	325
Supplier-months with no over-sampling	24,173		140	24,010	3	21

Note: Full sample N drops to 55,754 supplier-months (from 55,993) due to 239 missing observations for whether the supplier incurred a monthly TCR violation or not.

Table 5. Predicted potential "missing" violations,
over-samplers vs. others with at least one positive sample

Supplier group	Obs.	Imp.: Act.:	(1) Viol. Viol.	(2) No No	(3) No Viol.	(4) Viol. No
			Number of predictions			
Suppliers with full data available						
All supplier-months	506		75	337	3	91
Supplier-months with over-sampling	350		43	236	1	70
Supplier-months with no over-sampling	156		32	101	2	21
All suppliers						
All supplier-months	2,026		511	973	196	346
Supplier-months with over-sampling	1,531		372	641	193	325
Supplier-months with no over-sampling	495		139	332	3	21

Table 6. Econometric model results

Variable	**Pooled (1)**	**Pooled (2)**	**Panel RE (3)**	**Panel RE (4)**	**Panel FE (5)**
Oversample	0.0111*** (0.0009)	-0.0028 (0.0025)	-0.0011 (0.0014)	-0.0013 (0.0022)	0.0000 (0.0030)
Oversamp*anyp		0.1732 (0.1326)	-0.0031 (0.0300)	-0.0215 (0.0422)	-0.0501 (0.0456)
Oversamp*jump_dodge				0.0142*** (0.0034)	0.0154*** (0.0043)
Oversamp*jump_dodge*anyp				-0.0757** (0.0347)	-0.0242 (0.0406)
Jump_dodge				-0.0111*** (0.0025)	
Anyp		0.1788 (0.1317)	0.3674*** (0.0284)	0.3620*** (0.0312)	0.3964*** (0.0347)
Feddeviation				0.0000 (0.0001)	0.0000 (0.0001)
Summer				0.0042*** (0.0011)	0.0023* (0.0012)
Imputed	0.0077*** (0.0008)	0.0083*** (0.0008)	0.0130*** (0.0010)	0.0095*** (0.0012)	0.0148*** (0.0015)
Supplier effects	None	None	Random	Random	Fixed
Year dummies	No	No	Yes	Yes	Yes
Number of observations	55754	53849	53849	43528	43528
Number of groups	N/A	N/A	540	521	521
R^2	0.004	0.32	0.33	0.26	0.26

Notes: Dependent variable is whether a supplier incurs a monthly TCR violation (1) or not (0). Columns 1 and 2 report estimates from pooled linear probability models, and columns 3 and 4 from panel linear probability models, with standard errors (robust in columns 1 and 2) in parentheses. In columns 2-5, we instrument for *anyp*, *jump_dodge*, and all interactions with these variables using one-month lags. All models include a constant. The reduction in sample size from column 1 to column 2 results from predicting anyp using lagged values (the first month for each supplier must be dropped). The reduction in sample size from column 3 to column 4 results from missing data for the additional controls. Asterisks denote statistical significance: ***at 0.01; ** at 0.05; and * at 0.10.

Figure 1. Federal Minimum Sampling Requirements under the TCR

Public Water System ROUTINE Monitoring Frequencies					
Population	Minimum Samples/ Month	Population	Minimum Samples/ Month	Population	Minimum Samples/ Month
25-1,000*	1	21,501-25,000	25	450,001-600,000	210
1,001-2,500	2	25,001-33,000	30	600,001-780,000	240
2,501-3,300	3	33,001-41,000	40	780,001-970,000	270
3,301-4,100	4	41,001-50,000	50	970,001-1,230,000	300
4,101-4,900	5	50,001-59,000	60	1,230,001-1,520,000	330
4,901-5,800	6	59,001-70,000	70	1,520,001-1,850,000	360
5,801-6,700	7	70,001-83,000	80	1,850,001-2,270,000	390
6,701-7,600	8	83,001-96,000	90	2,270,001-3,020,000	420
7,601-8,500	9	96,001-130,000	100	3,020,001-3,960,000	450
8,501-12,900	10	130,001-220,000	120	≥ 3,960,001	480
12,901-17,200	15	220,001-320,000	150		
17,201-21,500	20	320,001-450,000	180		

*Includes PWSs which have at least 15 service connections, but serve <25 people.

U.S. Environmental Protection Agency, Office of Water (2001), Total Coliform Rule: A Quick Reference Guide, EPA 816-F-01-035, Washington, DC, p. 2.

Figure 2. Microbiological Analysis Record for a Single Massachusetts Water Supplier, October 2001-September 2002

Year 2001-2002	# of Samples Required	# of Samples Taken	Presence/ Absence	% Positive Samples	Violation	Type of Violation	Date of Phone Notification	Action(s) Taken
October	30	34	A					
November	30	38	A					
December	30	36	A					
January	30	39	A					
February	30	36	A					
March	30	36	A					
April	30	36	A					
May	30	38	A					
June	30	39	A					
July	30	~~41~~ 43	2P	<5%	NO			
August	30	48	4P	>5%	yes	MCL		NON NE 025044
September	30	38	A					

[redacted] called 8/19 - 4 Total Hits on Fri/Sat. Last Tues, 3 I,
Repeats Fri. Showed 2 Pos (colilert) Town Hall + Upham school.
Rep. on Sat showed Town Hall (DR, UR, RO) all Neg.
Rep. at Upham School: Upham Neg, UR +, DR -
Upham Rep. again, 1+
Still doing Repeats. RW all Neg
2 Tank Sites/2 Tanks at each - Do as SS to see Water Qual in Tanks.

[redacted] called on 7/26 to say they had 2 hits so far and a possible 3rd hit. I called him and got his voice mail. I told him to sample out, He needs to do all sites + UR + DR.
* No fecal/E. coli or the situation changes * Hopefully, it won't be positive but we'll see.

Figure 3. Imputing the number of positive samples

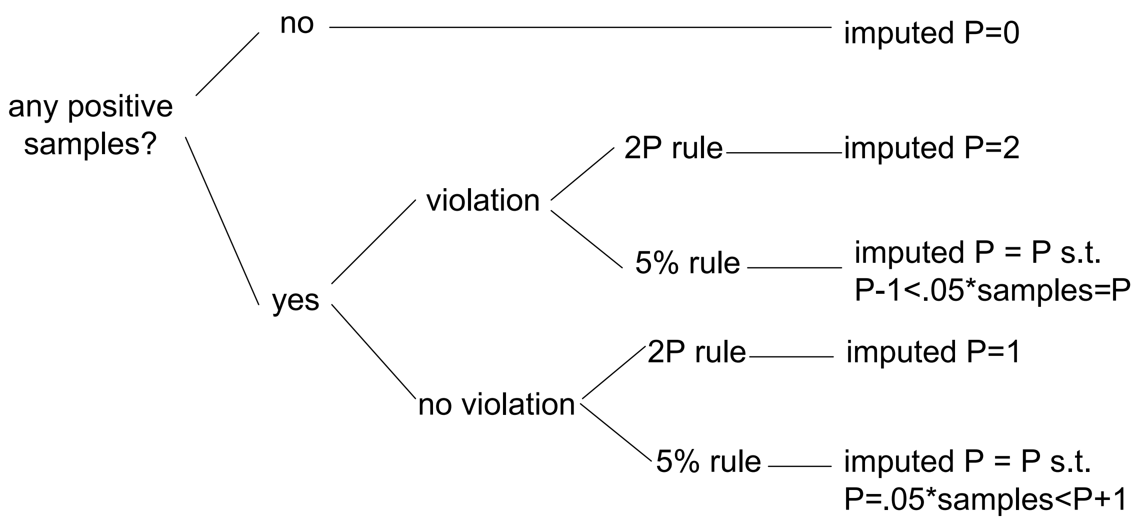
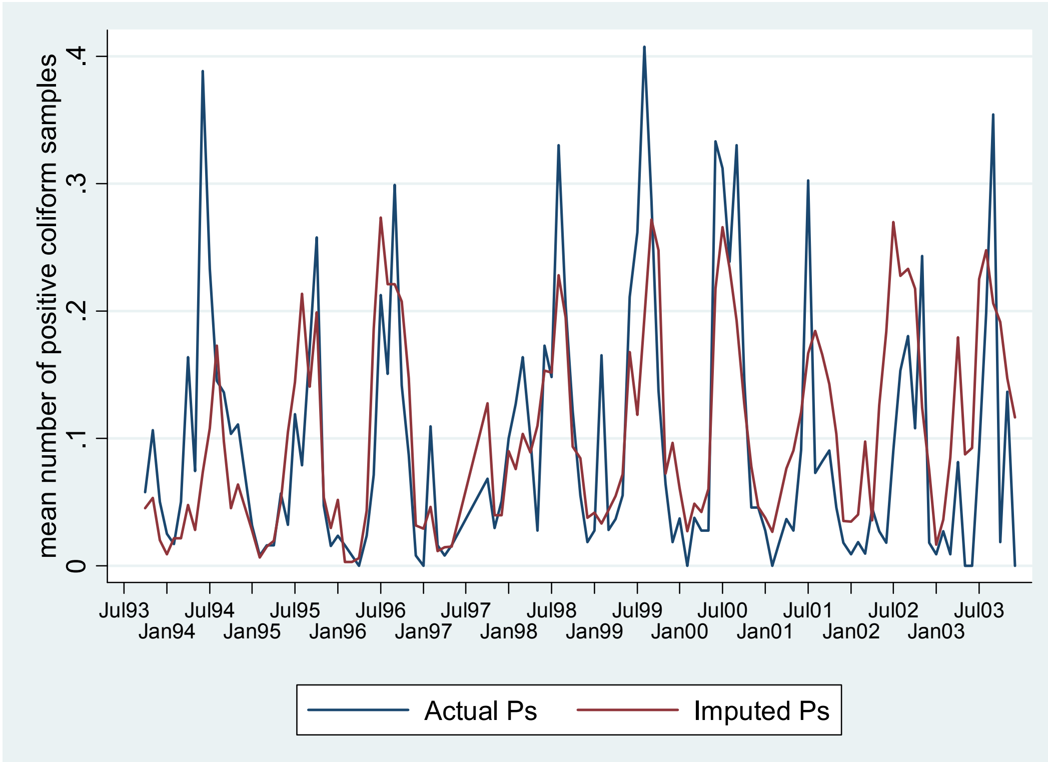


Figure 4. Actual vs. imputed positive coliform samples



References

Andreoni, J., B. Erard and J. Feinstein (1998), Tax Compliance, *J. Econ. Lit.* 36: 818-860.

Arora, S. and T. N. Cason (1995), An experiment in voluntary environmental regulation: participation in EPA's 33/50 program, *J. Environ. Econ. Manage.* 28, 271-286.

Becker, G. S. (1968), Crime and punishment: An economic analysis, *J. Polit. Econ.* 76(2): 169-217.

Bennear, L. S. and S. M. Olmstead (2008), The impacts of the "right to know": information disclosure and the violation of drinking water standards, *J. Environ. Econ. Manage.* 54, 181-198.

Colford, J. M., Jr., S. Roy, M. J. Beach, A. Hightower, S. E. Shaw, and T. J. Wade (2006), A review of household drinking water intervention trials and an approach to the estimation of endemic waterborne gastroenteritis in the United States, *J. Wat. Health* 4(Suppl. 2), 71-88.

Downing, P. and W. D. Watson (1974), The economics of enforcing air pollution controls, *J. Environ. Econ. Manage.* 1(3): 219-236.

Feinstein, J. S. (1989), The safety regulation of U.S. nuclear power plants: Violations, inspections, and abnormal occurrences," *J. Polit. Econ.* 97(1): 115-154.

Harford, J. (1978), Firm behavior under imperfectly enforceable standards and taxes, *J. Environ. Econ. Manage.* 5(1): 26-43.

Harrington, W, (1988), Enforcement leverage when penalties are restricted, *J. Public Econ.* 37(1): 29-53.

Helland, E. (1998), The enforcement of pollution control laws: Inspections, violations, and self-reporting," *Rev. Econ. Stat.* 80(1): 141-153.

Heyes, A. and N. Rickman (1999), Regulatory dealing – revisiting the Harrington paradox, *J. Public Econ.* 72(3): 361-378.

Innes, R. (2001), Violator avoidance activities and self-reporting in optimal law enforcement, *J. Law Econ. Org*. 17(1): 239-256.

Keohane, N. O., E. T. Mansur, and A. Voynov (2007), Averting regulatory enforcement: Evidence from new source review, *J. Econ. Manage. Strat.*, forthcoming.

Livernois, J. and C. J. McKenna (1999), Truth or consequences: Enforcing pollution standards with self-reporting, *J. Public Econ.* 71(3): 415-440.

Malik A. S. (1990), Screening and optimum enforcement, *RAND J. Econ*. 21(3): 341-353.

Raymond, M. (2004), Regulatory compliance with costly and uncertain litigation, *J. Regul. Econ.* 26(2): 165-176.

Wade Miller Associates, Inc. (1989), Economic impact analysis: Benefits and costs of final total coliform rule, Prepared for the U.S. Environmental Protection Agency, Office of Drinking Water, Washington, DC (8 March).

World Health Organization (2004), *Guidelines for Drinking-water Quality: Volume 1, Recommendations*, WHO, Geneva.

U.S. Environmental Protection Agency (2007), *Analysis of Compliance and Characterization of Violations of the Total Coliform Rule*, April, Washington, D.C.

Videras, J. and A. Alberini (2000), The appeal of voluntary environmental programs: Which firms participate and why? *Contemp. Econ. Policy* 18(4): 449-461.

Viscusi, W. K. and R. J. Zeckhauser (1979), Optimal standards with incomplete enforcement, *Pub. Policy* 27(4).

U.S. Environmental Protection Agency (2008), "Public Drinking Water Systems: Facts and Figures" http://www.epa.gov/ogwdw/pws/factoids.html.