Sampling Out: Regulatory Avoidance 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, and School of Forestry and Environmental Studies, Yale University, New Haven, Connecticut 06511

Received November 4, 2008. Revised manuscript received May 21, 2009. Accepted May 29, 2009.

This paper investigates strategic noncompliance with the Total Coliform Rule (TCR) under the U.S. Safe Drinking Water Act. The structure of the TCR provides incentives for some piped drinking water systems to avoid violations by taking additional water quality samples. We estimate the prevalence of this behavior and its potential impact on violations using monthly data for more than 500 Massachusetts water systems, 1993−2003. We find evidence that strategic oversampling is occurring. Water systems most likely to avoid violations by oversampling are most likely to oversample. A significant number of additional violations would have occurred if systems had adhered to legal sampling requirements, rather than oversampling. Our analysis of potential impacts of regulatory avoidance under the current rule suggests that alternative policies for monitoring bacteria in drinking water should be considered.

1. Introduction

In environmental economics and policy much attention has been paid to the incentives created by regulatory policy instruments, including incentives for compliance, the choice of lowest cost compliance methods, and the creation and adoption of new technologies. But regulations also can create incentives for strategic avoidance of regulatory action. The propensity of regulated entities to avoid the detection of violations has been demonstrated in contexts such as safety regulation of U.S. nuclear power plants (*1*), water quality regulation of U.S. pulp-and-paper plants (*2*), and new source review standards for power plants under the Clean Air Act (*3*). Undetected violations pose potential hazards to society, and efforts to avoid regulatory penalties are also inefficient (*4, 5*). This paper examines strategic regulatory avoidance of a major U.S. regulation: the Total Coliform Rule.

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. The TCR covers 54 000 community water systems, which collectively provide piped drinking water to 264 million people (*6*). The TCR is the most frequently violated SDWA regulation. U.S. water systems incurred 8310 monthly TCR violations per year, on average, between 1997 and 2003 (*7*). Key regulatory terms under the TCR are explained in Section 2.1 and summarized in Table 1.

The presence of total coliforms in drinking water is an indicator of treatment effectiveness and the integrity of distribution systems (*8*). Some coliform bacteria, such as *Escherichia coli* (*E. coli*) and fecal coliforms can cause acute gastroenteritis (cramps, diarrhea, fever, nausea, and vomiting), which can be deadly to vulnerable individuals. One estimate attributes 4.3−11.7 million cases of acute gastroenteritis per year to regulated U.S. drinking water systems (*9*). But most coliform bacteria are not dangerous. The TCR requires systems with any positive total coliform samples to test those samples and mandatory repeat samples for *E. coli* and fecal coliforms, since these indicate sewage or animal waste contamination, elevating health risks. Samples that test positive for either *E. coli* or fecal coliforms may trigger an acute TCR violation. But the TCR also limits the number of allowable positive total coliform samples in a month. Exceeding the threshold for positive total coliform samples results in a monthly (nonacute) violation, even with subsequent negative tests for *E. coli* and fecal coliform. The literature has recently addressed the statistical challenge of interpreting limited water quality data to determine whether a threshold standard has been violated (*10*−*12*). We examine a different challenge of threshold standards: the incentives they create for strategic behavior.

The U.S. Environmental Protection Agency (EPA) estimates that compliance with the TCR costs water systems $210−230 million annually, in 2007 dollars (*13*). In addition, water systems face potential financial penalties as well as public disapproval when they violate the TCR. Systems must notify the public about a monthly TCR violation within 14 days and must report it in their annual Consumer Confidence Report. Reporting requirements represent real costs for U.S. water systems, who have been shown to reduce violations of contaminant standards when required to disclose violations to the public (*14*). The potential to reduce these costs, paired with uncertain public health damages, creates an incentive for water systems to avoid the detection of monthly TCR violations.

We investigate whether water systems respond to the current incentives of the TCR by taking more samples than legally required to avoid monthly violations. This avoidance behavior is possible in the majority of states, though a handful of states currently prohibit counting additional samples toward a water system's violation threshold. We examine monthly data for water systems in the Commonwealth of Massachusetts (where additional samples are included in the determination of violations) from 1993−2003. We find that water systems do take additional samples to avoid monthly violations and that many more violations would have occurred if systems had instead sampled according to legal requirements.

2. Experimental Section

2.1. Regulatory Background. The TCR requires water systems to collect a minimum number of samples representative of the piped distribution system each month. The minimum number of samples increases with system size (*15*) and applies to all U.S. community water systems. We refer to this set of samples as the "federal minimum" number of samples. But the number of samples taken in a month can exceed the federal minimum for several reasons. First, state enforcement agencies may negotiate with systems to establish a routine monthly sampling plan, which can exceed the federal

* Corresponding author phone: 203-432-6247; fax: 203-436-9150; e-mail: sheila.olmstead@yale.edu.
[†] Nicholas School of the Environment, Duke University.
[‡] School of Forestry and Environmental Studies, Yale University.

10.1021/es803115k CCC: $40.75

Published on Web 06/12/2009

TABLE 1. Definitions of Total Coliform Rule Terms

term	definition
acute violation	occurs when any repeat sample tests positive for *E. coli* or fecal coliforms, or if a routine sample tests positive for *E. coli* or fecal coliforms, and any repeat sample tests positive for total coliforms
monthly (non-acute) violation	occurs when a specified number of samples test positive for total coliform, even if subsequent testing is negative for *E. coli* and fecal coliform. The threshold for monthly violations varies with the number of samples taken. See next two definitions
two positive (2P) rule	systems collecting fewer than 40 samples in a month incur a monthly violation if at least 2 samples test positive for total coliform
five percent (5%) rule	systems collecting at least 40 samples in a month incur a monthly violation if more than five percent of samples test positive for total coliform
federal minimum samples	the TCR sets a minimum number of samples that must be taken in each month based on the population served by the water system. Larger systems must take more samples
routine samples	total number of samples a system is required to take each month. This number may exceed the federal minimum samples based on negotiations between the system and state regulators who implement the TCR
repeat samples	if a required sample tests positive for total coliform the system is required to take three repeat samples (four if the system routinely takes 1 sample per month or fewer). The TCR prescribes the location of these additional samples
oversampling	a system is oversampling if the total number of samples collected in a month exceeds the number of routine samples plus the number of required repeat samples
sampling out	sampling out is oversampling that reduces the probability of a monthly violation
overcompliance	overcompliance is oversampling that does not reduce the probability of a monthly violation

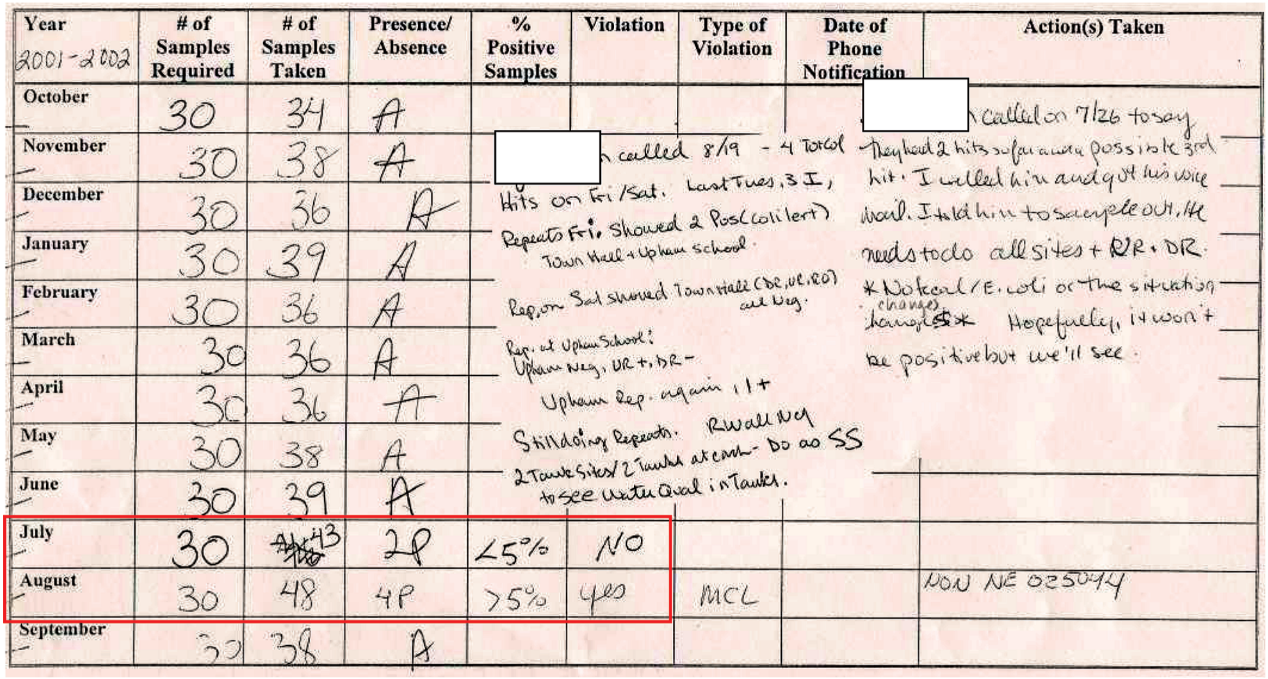
minimum. We refer to the state-negotiated number of samples as "routine samples." Second, if any routine samples test positive for total coliform, systems must take "repeat samples." The TCR prescribes both the number and the location of repeat samples. Finally, some systems may take yet more samples, though legally they cannot do this without regulators' permission. We call this third behavior, drawing more than required routine and repeat samples, oversampling.

TCR violations may be either acute or monthly (nonacute). A water system incurs an *acute* TCR violation if any repeat sample tests positive for *E. coli* or fecal coliforms (these tests are required after a positive routine sample), or if a routine sample tests positive for *E. coli* or fecal coliform, and any repeat sample tests positive for total coliforms. This paper focuses on *monthly* TCR violations, which are based on the number of positive total coliform samples, only. Water systems that take at least 40 samples in a month incur a monthly TCR violation if more than five percent of those samples test positive; we refer to this as the 5% rule. Systems taking fewer than 40 samples incur a monthly TCR violation if two or more coliform samples test positive; we refer to this as the 2P rule.

Two characteristics of the monthly standard make regulatory avoidance possible. First, in most states, violations are determined by the number of samples taken, rather than the number of samples legally required. Second, once at least 40 samples are taken, violations are based on the percentage of positive samples, rather than absolute numbers. Water system operators know more than regulators about the likelihood of drawing a positive sample in different parts of the system and at different times of the day or month, given the timing of disinfection measures or weather events. Thus, it is possible that systems can strategically draw additional negative samples to remain below a violation threshold.

Oversampling does not necessarily indicate that water systems are strategically avoiding regulatory compliance. Taken randomly, more samples provide a more complete picture of the presence of bacterial hazards in a drinking water distribution system. Thus, oversampling water systems may be overcomplying with the TCR, increasing testing costs to protect public health.

Oversampling to avoid violations is referred to by water system staff and regulators as "sampling out". Water systems that sample out, if any, are a subset of systems that oversample. Figure 1 offers an example of both overcompliance and sampling out. We obtained this record of one Massachusetts water system's microbiological analysis from the files in the system's regional office of the Department of Environmental Protection (DEP). From October 2001 to June 2002, the system took more than its 30 required routine samples. All samples tested "absent" for total coliform during this period, so the extra samples are not repeat samples, and with no positive samples, this is also not sampling out. This



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	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 1. Microbiological analysis record for a single Massachusetts water system, October 2001 to September 2002.

system appears to have overcomplied with the TCR from October 2001 to June 2002.

In July 2002, the system drew 43 samples, obtaining two positives, but because its sample size was so large, it remained below the 5% threshold (to which it was subject, since it took at least 40 samples) and did not violate the TCR. In August, the system took 48 samples, obtained four positives, and exceeded the 5% threshold. The system's July 2002 sampling decisions, in consultation with the regional regulator, are documented in notes attached to the record, which indicate a desire to avoid a monthly TCR violation. If this system had not sampled out in July 2002, it would have violated the TCR, as it did in August. Figure 1 provides evidence of the temporal dimension of sampling out. The notes suggest that in July the water system "had 2 hits so far with a possible 3rd hit", and subsequently sampled out. Without these notes, we could not determine whether the system took additional samples after approaching the violation threshold, or before. Sampling out may be evidence that water systems either routinely take additional samples so as to reduce the probability of a violation, or do so only when the risk of a violation is elevated.

In distinguishing overcompliance from sampling out, it is helpful to think about three categories of water systems. For systems always 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. However, a system that routinely takes fewer than 40 samples may in a given month take at least 40 (for any of the reasons described above); in this month, a violation for this system will be determined under the 5% rule. We refer to the group of water systems that sometimes take fewer than 40 samples, and sometimes take 40 or more, as "jumpers." These systems are jumping back and forth between the 2P and 5% rules. (The system in Figure 1 is a jumper.) Jumpers may reduce violations by oversampling, if this behavior moves them from violation status under the 2P rule to nonviolation status under the 5% rule (the system in Figure 1 avoids a violation by oversampling in July). The last group of systems, those always under the 5% rule, may benefit from strategically sampling to ensure additional negatives, reducing the percent positive to 5% or below. Thus, only the last two groups (jumpers and those always subject to the 5% rule) may reduce the probability of a violation by oversampling; those always subject to the 2P rule cannot.

2.2. Methods. We use two statistical approaches to distinguish sampling out from overcompliance. First, we examine whether oversampling is more frequent among systems for which it is most likely to reduce violations. We compare rates of oversampling in the three groups of systems defined above: those that always take fewer than 40 samples, jumpers, and those that always take at least 40 samples. If oversampling is strategic, it should happen more often among the latter two groups, the only water systems that can reduce the probability of a violation by oversampling. Strategic oversampling should also happen more often among these latter two groups during months with at least one positive sample, since oversampling may reduce the probability of a violation only for systems with a positive draw.

Jumpers and systems always governed by the 5% rule systematically serve larger populations than systems that are always governed by the 2P rule. Thus, this first stage of the analysis cannot rule out the possibility that larger systems oversample more than smaller systems for reasons unrelated to regulatory avoidance. For example, larger systems may have lower marginal costs of oversampling (more sampling staff out on more days lowers the cost of additional samples taken on any given day) or greater marginal benefits of oversampling (learning more about problems in the distribution system may lower future nonpecuniary costs of violations such as angry phone calls from customers). Indeed, in other contexts researchers have found that larger firms are more likely than smaller firms to overcomply with environmental regulations (*16, 17*). The second step in our analysis is designed to further distinguish overcompliance from sampling out by examining whether significantly more violations would have occurred if systems had not oversampled. If oversampling is predominantly overcompliance, it should not systematically reduce violations. If oversampling is really sampling out, however, we would expect it to do so.

2.3. Data. With assistance from the Massachusetts DEP, we collected data on 559 Massachusetts water systems' monthly total coliform sampling and violations from 1993 to 2003, creating a panel of 55 993 system-months. Community water systems in Massachusetts range from small private

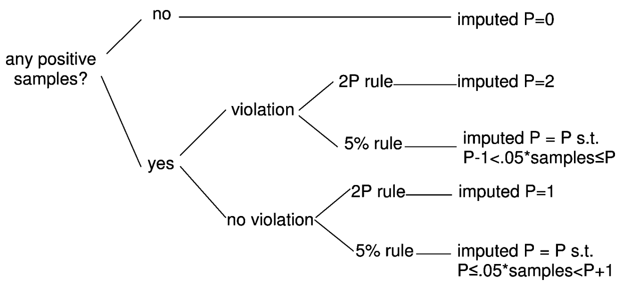


FIGURE 2. Imputing the number of positive samples.

systems to the largest public regional water system, which serves more than 2 million people. We sought monthly data on whether a system violated the TCR or not, the number of total coliform samples drawn by each system, the number of routine samples required for each system, and the number of positive total coliform samples.

All of this information was obtainable from DEP headquarters, except for the last item. The DEP headquarters does not collect data on 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). This information was insufficient to detect oversampling, since the number of required repeat samples depends on the number of positive samples. From regional DEP offices, we obtained paper records of the number of positive samples for a portion of our full observation period for 245 systems: 133 in the Western region, partial records, 1993−1997; and 112 in the Northeast region, partial records, 1997−2003. For a subsample of 13 970 system-months, we have these full data on the number of positive samples.

We conduct the analysis using this subsample. However, the subsample is not random. It is drawn from only two of the four regional DEP offices, and regions may differ in TCR enforcement and implementation. To address potential concerns about selection bias introduced by using this restricted sample, we also impute the number of positive samples for system-months in which these data were not available and run the analysis for all Massachusetts system-months, 1993−2003.

Figure 2 describes our imputation method. For systems with no positive coliform samples, we impute zero positive draws. For systems with at least one positive draw, the imputation method varies depending on the violation rule relevant for the system in that month, and on whether we observe a violation. Systems with a monthly TCR violation are assumed to stop sampling once they violate (as the TCR allows), so those subject to the 2P rule with a violation are assumed to have two positives, and those subject to the 5% rule are assumed to have exactly the number of positive samples that exceeds this threshold. Systems without a violation and subject to the 2P rule are assumed to have a single positive sample. Systems 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.

We expect to underestimate the number of positives for those who violate (the middle two groups in Figure 2), 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 systems subject to the 5% rule who do not violate but have at least one positive draw, because we assume they remain just under the 5% threshold.

Table 2 compares imputed and actual positives within the subsample of 13 970 system-months for which full data were available. The first row in Table 2 indicates that the mean of imputed positives (0.096) is very close to the mean of actual positives (0.094). This is largely because positive samples are uncommon, and our imputation method perfectly predicts zero positive samples (row 2 of Table 2). The actual (2.60) and imputed (2.66) mean numbers of positive draws are also very close. But we do better for some groups than for others. As expected, we under-predict positive samples for systems with a monthly TCR violation. For those with no violation under the 2P rule, we under-predict positive samples; it appears that in some cases, systems under the 2P rule obtain two or more positive samples, but do not report a violation to the DEP (mean positive samples for this group are greater than one). As expected, we also overpredict the number of positive samples for those with no violation under the 5% rule.

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. Table 3 compares predicted violations from our imputed positive samples to those using the actual number of positive samples. Imputation errors show up in the last two columns. The only group of systems for which we significantly mis-predict whether systems are above or below the violation threshold using the imputed data 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 systems would not violate, but the actual positives suggest that they would. This is the anomalous group identified in Table 2, systems under the 2P rule that obtain two or more positive samples, but do not report a violation to the DEP. For all other groups, imputed and actual positives generate different violation predictions in only 1−6% of system-months.

The use of imputed data to predict whether systems are oversampling introduces measurement error. Errors in imputation are small and are unlikely to introduce strong bias. In any case, the use of the imputed data does not significantly affect the findings. The imputed data are included to serve as a robustness check against the possibility that our findings are only present in the nonrandom subsample for which the number of positive samples was available.

3. Results

3.1. Oversampling by System Category. We use two different measures of oversampling intensity. The first is an indicator variable that takes a value of one if the water system oversampled in a particular month and zero otherwise. The second is the number of extra samples taken, conditional on oversampling. These measures of oversampling are obtained for all system-months and also broken into three categories: systems always governed by the 2P rule, jumpers, and those always governed by the 5% rule. We present results separately for the subsample with actual numbers of positive samples (Table 4, rows 1 and 2) and for all system-months, using actual and imputed data (Table 4, rows 3 and 4).

The pattern of oversampling is similar for the two samples. A much larger percentage of jumpers and systems always facing the 5% rule oversample. Using the subsample (Table 4, row 1), only 16% of systems always governed by the 2P rule oversample, compared to 76% of jumpers and 84% of those always under the 5% rule. The differences in means reported for the three groups in row 1 of Table 4 are all significant at $\alpha<0.01$, by two-tailed *t* tests for differences in means. Using actual and imputed data for the whole state (Table 4, row 3), 38% of systems always governed by the 2P rule oversample, while 83% of jumpers and 89% of systems always under the 5% rule oversample. The differences in means for the three groups reported in row 3 of Table 4 are all significant at $\alpha < 0.01$, by two-tailed *t* tests for differences in means.

The number of additional samples taken by oversamplers

TABLE 2. Comparison of Imputed and Actual Positive Coliform Samples

	variable by system group	obs	mean	med	std dev	min	max
	all system-months with full data						
(1)	*actual_P*	13 970	0.094	0	0.808	0	33
	imputed_P	13 970	0.096	0	0.687	0	23
	system-months with no positive samples						
(2)	*actual_P*	13 464	0	0	0	0	0
	imputed_P	13 464	0	0	0	0	0
	system-months 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						
(3)	*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 3. Comparison of Imputed and Actual Violations

			fraction of predictions			
			(1)	(2)	(3)	(4)
		imp	viol	no	no	viol.
system group	obs	act	viol	no	viol	no
all system-months with full data	13 970		0.006	0.997	0.003	0.000
system-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
system-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	3089		0.000	1.000	0.000	0.000

is also greater for jumpers and systems always under the 5% rule. Using the smaller regional data set, systems always under the 2P rule take, on average, fewer than four extra samples, conditional on oversampling. Jumpers take an average of 8 additional samples and those always under the 5% rule take an additional 14 samples (Table 4, row 2) (average numbers of extra samples are 3, 10, and 17 for the groups, respectively, using data for the whole state as shown in Table 4, row 4). The differences in means reported in rows 2 and 4 of Table 4 are all significant at $\alpha<0.01$, by two-tailed t tests for differences in means.

3.2. Oversampling and Violation Frequency. If oversampling is predominantly overcompliance, it should not systematically lower the probability of a violation, as sampling out would. To distinguish sampling out from overcompliance we predict violation frequency assuming universal adherence to systems' required routine and repeat samples and compare this to actual violations, which are based on the number of samples taken. The calculations are done separately for systems that oversample and those that do not.

We present these findings in Table 5, reporting results for the subsample 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 the subsample ((75 + 13 801)/13 970) and all systems ((511 + 54 701)/55 754). The vast majority of system-months would have experienced no change in violation status if the system had taken the required number of samples. However, among the 94 (91 + 3) system-months (using the subsample) where predicted 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 system had taken the required number of routine and repeat samples, but with the actual number of samples no violation occurred. Furthermore, the majority of these cases (70 of 91, or 78%) are for system-months in which oversampling did occur. These are the cases in which systems may have sampled out. The picture is similar for the whole state; where predicted violation status diverges from actual violations, most (346/(346 + 196), or 64%) are cases in which we would predict a violation, but none occurs. And 94% of these instances (325 of 346) occur in system-months characterized by oversampling.

If we restrict this analysis to system-months with at least one positive sample, we gain a better sense for the frequency of unobserved violations due to sampling out. For the 31 581 system-months in which oversampling occurred (Table 5, all systems), there were 1531 system-months with at least one positive sample. The oversampling resulted in a potential violation status change in less than 2% of system-months (325 + 193 = 518). Of those 518 status changes, 325 observations were cases where there was no actual violation, but there would have been a violation if the system had taken only their required routine and repeat samples. These 325 system-months rep-

TABLE 4. Over-Sampling among Systems

	variable by system group	obs	mean	med	st dev	min	max
	Systems with Full Data Available						
	oversampling (0/1)						
	all system-months	13 970	0.45	0	0.50	0	1
(1)	always under 2P rule	7561	0.16	0	0.36	0	1
	jump between 2P and 5% rules	4251	0.76	1	0.43	0	1
	always 5% rule	2158	0.84	1	0.36	0	1
	extra samples (no.) if oversampling						
	all oversampled system-months	6108	8.83	6	10.62	1	297
(2)	always under 2P rule	1049	3.69	2	3.20	1	27
	jump between 2P and 5% rules	3239	7.60	5	7.63	1	106
	always 5% rule	1820	14.00	10	14.98	1	297
	All Systems						
	oversampling (0/1)						
	all system-months	55 993	0.57	1	0.50	0	1
(3)	always under 2P rule	33 511	0.38	0	0.49	0	1
	jump between 2P and 5% rules	18 113	0.83	1	0.38	0	1
	always 5% rule	4369	0.89	1	0.31	0	1
	extra samples (no.) if oversampling						
	all oversampled system-months	20 664	8.55	4	14.35	1	677
(4)	always under 2P rule	7084	2.62	2	2.45	1	33
	jump between 2P and 5% rules	10 614	10.17	6	13.45	1	328
	always 5% rule	2966	16.88	11	24.70	1	677

TABLE 5. Predicted Vs. Actual Violations, Over-Samplers Vs. Others[a]

			number of predictions			
			(1)	(2)	(3)	(4)
		pred	viol	no	no	viol
system group	obs	act	viol	no	viol	no
Systems with Full Data Available						
all system-months	13 970		75	13 801	3	91
system-months with oversampling	6241		43	6127	1	70
system-months with no oversampling	7729		32	7674	2	21
All Systems						
all system-months	55 754		511	54 701	196	346
system-months with oversampling	31 581		372	30 691	193	325
system-months with no oversampling	24 173		139	24 010	3	21

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

resent 21% of the 1531 system-months with at least one positive coliform sample. (If we do the same exercise with the regional data only, this happens in 20% of months with at least one positive sample, at about the same rate.)

There were 707 monthly TCR violations in our Massachusetts sample from 1993 to 2003, and we estimate an additional 325 violations avoided by systems that oversampled. This suggests that only about 69% of "true" TCR violations in Massachusetts are actually observed. Putting this a different way, almost one-third of TCR violations in the state may go undetected.

We recommend caution in extrapolating this ratio to the national level, since we have data from a single state, and regulatory avoidance will vary across states (for example, sampling out is impossible in states like New York and Arizona, which do not count extra samples toward a system's violation threshold). Nonetheless, if we do extrapolate this ratio to the national level where there are 8310 monthly TCR violations per year, on average, these results suggest that U.S. water systems may "sample out" of 3,000 to 4,000 monthly TCR violations each year.

4. Discussion

Our empirical results suggest that (1) oversampling occurs on a large scale in Massachusetts; (2) water systems most likely to succeed in regulatory avoidance are most likely to oversample; and (3) additional monthly TCR violations may have occurred among Massachusetts water systems, 1993−2003, had systems not oversampled. The balance of evidence suggests that much of the oversampling we observe is sampling out, rather than overcompliance with the TCR. In Massachusetts, we estimate that almost one-third of monthly TCR violations may go undetected due to sampling out. Extrapolating that to the national level suggests that thousands of monthly TCR violations may be avoided each year by sampling out, though we are cautious about interpreting our estimates for Massachusetts for the nation as a whole.

EPA is currently revising the TCR, so our analysis of regulatory avoidance is timely. If positive total coliform samples can be directly linked to health impacts, then violation avoidance is undesirable from a human health standpoint. Positive total coliform samples, with no detected *E. coli* or fecal coliforms, may provide evidence of health risks for some types of systems, but not for others, depending on characteristics like source water quality, disinfection technologies, and system age and structure. In this case, a monthly standard could be specific to groups of systems, based on such characteristics, and would need to be updated as characteristics change (e.g., systems age, land use changes).

Where positive total coliform samples do not provide evidence of health risks, the monthly standard could be eliminated. Resources are wasted avoiding violations with minimal public health impact. In addition, systems that do not avoid violations bear costs from these violations, even though they do not represent threats to public health. The TCR already requires tests for *E. coli* and fecal coliforms when a total coliform sample tests positive; the TCR could retain this requirement, even if the monthly standard is eliminated for some systems. These tests provide the basis for acute violations under the current rule. The TCR prescribes the number and location of repeat samples for *E. coli* and fecal coliform, so sampling out is not possible for such tests. To retain some regulatory consequence for positive total coliform samples, absent positive tests for *E. coli* and fecal coliforms, systems could still be required to report total coliform results in annual Consumer Confidence Reports.

If the monthly total coliform standard is retained, it should be revised. Violations should be contingent upon required samples rather than actual samples taken (as is currently done in several states), or upon the absolute number of positive samples rather than percentages. Either change would eliminate the incentive to sample out.

Acknowledgments

This research was funded by the National Science Foundation (nos. SES-0648256 and SES-0647855) and the Hixon Center for Urban Ecology, Yale University. Damon Guterman and Andrew Durham at the Massachusetts 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, the Duke Law and Economics Workshop, and the University of Massachusetts Environmental and Resource Economics Seminar. The authors, alone, are responsible for any errors.

Literature Cited

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

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

(3) Keohane, N. O.; Mansur, E. T.; Voynov, A. Averting regulatory enforcement: evidence from new source review. *J. Econ. Manage. Strategy* **2009**, *8*, 75–104.

(4) Malik, A. S. Screening and optimum enforcement. *RAND J. Econ.* **1990**, *21*, 341–353.

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

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

(7) U.S. Environmental Protection Agency. *Analysis of Compliance and Characterization of Violations of the Total Coliform Rule*; Environmental Protection Agency: Washington, DC, 2007.

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

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

(10) Borsuk, M. A.; Stow, C. A.; Reckhow, K. H. Predicting the frequency of water quality standard violations: a probabilistic approach for TMDL development. *Environ. Sci. Technol.* **2002**, *36*, 2109–2115.

(11) Smith, E. P.; Ye, K.; Hughes, C.; Shabman, L. Statistical assessment of violations of water quality standards under Section 303(d) of the Clean Water Act. *Environ. Sci. Technol.* **2000**, *35*, 606–612.

(12) McBride, G. B.; Ellis, J. C. Confidence of compliance: a Bayesian approach for percentile standards. *Water Res.* **2001**, *35*, 1117–1124.

(13) Wade Miller Associates, Inc. *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, 1989.

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

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

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

(17) Videras, J.; Alberini, A. The appeal of voluntary environmental programs: which firms participate and why. *Contemp. Econ. Policy* **2000**, *18*, 449–461.

ES803115K