"Private" Provision of Publicly Useful Information: An Empirical Analysis of Public Notification Rules for Safe Drinking Water Act

Kenji Adachi and Yoshifumi Konishi

Department of Applied Economics
University of Minnesota

Selected Paper prepared for presentation at the American Agricultural Economics Association (AAEA) Annual Meeting, Portland, OR, July 29-August 1, 2007

Key words: public information, water contamination, Safe Drinking Water Act

I. Introduction

Early public notification of hazardous water contamination incidences can improve consumer welfare considerably. Suppose, for example, that a community water system serves one million consumers and finds that its after-treatment water contains radium 228 at the level of 30pCi/L. Approximately 20 to 45 water users who drink the contaminated water at the rate of 2 liters per day would be expected to contract a cancer after one to two years. Consumers who would like to avoid this level of cancer risk can do so by adopting appropriate treatment devices at homes at relatively low costs by iron exchange water softeners or reverse osmosis treatment.[1] Since consumers can decide their self-protection levels optimally given information, early provision of accurate and relevant information can only improve consumer welfare.

Nonetheless, evidence suggests that private agents hide or delay notifying the public of health-threatening contamination incidences. The economic intuition is trivial. Private agents will keep information "private" for as long as their private benefit of doing so exceeds the private cost. The negative externality of keeping it private, unless internalized by regulatory or economic sanctions, will not be a factor in their decision making. An important empirical question, then, is: What determines "private valuation" of the benefits and costs of keeping private the information about contamination incidences, which otherwise can or should be made public? This question is important in lieu of guiding a regulatory design for information disclosure.

We examine a unique dataset obtained from U.S. EPA, which includes 2,062 observations of maximum contamination level (MCL) violation incidences in Illinois from 1980 to 2006.[2] The dataset was obtained by filing a Freedom of Information Act request. It contains detailed information on public water systems (PWSs) such as type of PWS, population size, water source, ownership, and location. The location information is used to match up with U.S. census data in order to obtain community characteristics such as median household income, education level, and the percentage of young and old subpopulations. Most importantly, it enables us to associate, with each MCL violation incidence, information such as the type of chemical, concentration level, the date on which public notification (PN) is requested, and the date on which public notification is received. Using this information, we construct a variable termed "PN timing."

In the U.S., PN rules concerning drinking water are well established under the consumer right-to-know provisions of the Safe Drinking Water Act. New PN rules, recently enacted in 2000, classify all MCL violations into one of three categories, or "tiers", based on the potential risk of adverse

[1] More detailed information is available, for example, via web-based fact sheets provided by U.S. Environmental Protection Agency and state agencies such as California Dept. of Health Services and New Hampshire Dept. of Environmental Services.
[2] Formally, MCL is defined as the "maximum permissible level of a contaminant in water delivered to any user of a public water system" (U.S. EPA, 2005). The MCLs are set largely by two complementary regulations, called Total Coliform Rule (TCR) and Chemical Rule, as part of the National Primary Drinking Water Regulations.

health effects. In general, tier 1 concerns violations that pose immediate health risks and includes the MCL violations for (i) total coliform, (ii) nitrate and nitrite, and (iii) turbidity. All tier 1 violations require a public notice within 24 hours via broadcast media (such as television and radio) or hand delivery/posting. All other kinds of MCL violations are essentially classified as tier 2, and require a public notice within 30 days via hand delivery/posting or regular mail.

Two surprising facts arise from our database. First, despite the 30-day notification requirement, PWSs issue a public notice for *initial* tier-2 MCL violations on average on 87-th day after receiving a PN request from their primacy agency. Because many of these violation incidences are not trivial (the concentration levels can be two to five times higher than the corresponding MCL standard), this calls upon a question of why public water systems delay a public notice, which can otherwise improve consumer welfare. Second, the distributional characteristics of the PN days before the new rule appear to be significantly different those after the new rule.[3] Before the new rule, the PWSs issue a public notice immediately in almost all instances. After the new rule, on the contrary, the PWSs appear to use discretion in deciding when to issue a notice.

To investigate these issues in depth, we estimate a random-effects model on the panel-like cluster sample, with interaction terms to account for the impact of the regulatory change. Our regression results indicate that (i) PN delay is significantly positively correlated with the severity (concentration above MCL standards) of the contamination incidence after the new rule, but (ii) the coefficient on the severity variable is insignificant before the new rule, and (iii) PN timing decreases with the population size served, the income level of the community, the relative size of the white population, and the relative size of the educated population.

Two of these results are of particular importance. Result (i) is alerting, as it implies that consumers are informed later when they face greater health risks. Because the chemical contaminants covered in this analysis do not pose immediate health risks, water managers have incentives to delay a public notice. Moreover, result (ii), combined with result (i), has an intuitive appeal. Because PWSs have more flexibility under the new rule, they may use discretion in determining the "efficient" timing of a public notice. Their PN timing decisions, then, can be solely based on their own private valuation. Our results imply that it is privately more efficient for the PWSs to delay a public notice when the concentration levels are higher. Unfortunately, this goes against the intended purpose of the new rules, which were meant to inform the public of health-based violations better and earlier when the violations are serious while giving flexibility to reduce costs of public notification.

To check the robustness of our results, we have also run several regression models. In particular, the PN timing variable constructed as above resulted in many "zero" observations, because there is often a long lag between the MCL violation date and the corresponding date on which public notice is requested.

[3] Under the new rules, tier 2 PN deadline is extended from 14 days to 30 days, but PWSs need to conform to more stringent public notice formats.

To take care of this censoring issue, we also ran the Tobit and random-effects Tobit models. We found that Tobit specifications, in fact, reinforce our results. The coefficient on severity after the new rule becomes more significant with the same positive sign; the coefficient on severity before the new rule becomes significant but with a negative sign; and some other variables that were not significant in the ordinary random-effects model become significant.

This paper complements the existing literature in an important way. A large body of literature now exists, which addressed both empirically and theoretically the question of how informational programs affect consumer behavior and welfare (e.g. Colantoni, Davis, and Swaminuthan, 1976; Viscusi and O'Connor, 1984; Viscusi *et al.*, 1986; Smith *et al.*, 1988; Smith and Johnson, 1988; Ippolito and Mathios, 1990; Smith and Desvousges, 1990). These studies find that consumers do respond to the risk information in a rational and sensible way. However, there is only a thin literature on the motivation of information suppliers when there are no direct transfers as the result of information transmission. Though several studies investigated how communication, or transfer of information, arises in various principal-agent setups (e.g. Melumad and Reichelstein, 1989; Melumad and Shibano, 1991; Pitchik and Schotter, 1987), little empirical research appears to exist on the motivation of private or quasi-private agents who can supply publicly valuable information. This is unfortunate, because important environmental risk information often comes from private sources, and regulatory agencies have limited authority over their information disclosure activities.

Our findings also have practical implications for information disclosure policies. As our empirical results suggest, private or quasi-private information providers often have private interests that go against those of the public. Allowing for flexibility in information disclosure without appropriate rules for sanctions is not optimal when information providers have such private interests. It is, of course, difficult to determine to what extent and when some publicly useful information can be or should be considered public. Even in the case of mandatory public notification where all MCL violation information must be made public within 30 days, our empirical evidence suggests that the “flexibility” clause blurred the boundary between private and public information. To help guide efficient information disclosure policies, further empirical analyses of this type are necessary.

II. Data Background

The PN regulations fulfill the consumer right-to-know provisions of the Safe Drinking Water Act. All public water systems must issue a public notice whenever they fail to comply with the National Primary Drinking Water Regulations including health-based water quality standards called Maximum Contaminant Levels (MCLs). MCL is formally defined as the "maximum permissible level of a contaminant in water delivered to any user of a public water system" (U.S. EPA, 2005), and thus essentially defines the allowable level of public health risks associated with the water pollutant.

Under the current rule, all MCL violations are classified into one of three categories, or "tiers," based on the associated health risks. Tier 1 concerns violations that pose immediate health risks. All chemical MCL violations that pose latent, chronic health risks are essentially classified as tier 2.

Important changes were made to the PN regulations in May, 2000. Among others is the revision to the tier-2 violation notification timing. Prior to the revision, all PWSs were required to issue a public notice within 14 days via hand delivery/posting or regular mail. No extension of the 14-day deadline was allowed. Under the new rule, PWSs are now required to issue a public notice "as soon as practical but no later than 30 days after the system learns of the violation" (U.S. EPA, 2000a and 2000b). The new rule also gives the primacy agency discretion to extend the time period allowed for the tier 2 notice from 30 days to up to three months for the initial notice in some appropriate circumstances. These changes were made so as to give the PWSs flexibility in meeting the PN rules in the cost-effective manner. For example, the extensions now make it possible for the PWSs to include the initial notice in the same mailing as the water bill. An important question, then, is how this regulatory change has affected the PWSs' public notification decisions.

To investigate this question, we collected detailed MCL violation and compliance data on Illinois water systems by filing a Freedom of Information Act request. The obtained dataset provides the basic PWS information, violation type, chemical name, sample result, and reported enforcement actions/dates for each MCL violation. To supplement the PWS information, the SDWIS/Fed Inventory data were also used. The U.S. EPA staff also provided a comprehensive list of contaminant codes, which helped match each contaminant with its corresponding MCL standard.

The data set allows us to construct a variable, termed "PN days" or "PN timing," by calculating the difference between the date on which the primacy agency recorded the enforcement action "public notice requested" and the date on which it "received a public notice." As the new PN regulations became fully effective only after May 6, 2002, we classify the violation and enforcement records according to whether the violations occurred before and after this date (instead of May 2000). We use only tier-2 MCL violations, as tier-1 MCL violations for total coliform do not contain information on sample results.

Several surprising observations arise from the data set. First, the records show that despite the mandatory public notification deadline, the PWSs do not necessarily issue a public notice within 30 days after the new rule and the PN days vary significantly across observations with a standard deviation of 99.7 days. As Table 1 shows, some of the MCL violations are not trivial --- the contamination levels found may be two to five times higher than the respective MCL standards. Moreover, for many of the violations, MCL compliance is not achieved when the PN is issued late. Second, the distributional characteristics of the PN days before the new rule appear to be significantly different those after the new rule (Figure 1). Before the new rule, the PWSs issue a public notice immediately in almost all instances. After the new rule, on the contrary, the PWSs appear to use discretion in deciding when to issue a notice. This is despite the fact that pollution characteristics are very similar between the two subsets. Our primary interest is to explain this difference in terms of the observed explanatory variables.

Lastly, to illustrate the health-risk consequences of the delays in public notification, we provide rough cancer risk estimates based on the EPA's published risk coefficients for radioactive pollutants. Table 2 reports the cancer morbidities in cases/millions for radioactive pollutants, estimated at mean, min, and max concentration levels actually found in the data set for various milestone timings. As there are many radioactive pollutants in the Gross Alpha Activity, we picked two examples for this category. For example, in the case of Radium-228, at the maximum concentration level found in the data set, approximately 5 out of millions would be expected to contract a cancer at the mean PN days and 44 out of millions at the maximum PN days if they consume 2L/day of this contaminated water.

Table 1. Summary of Contamination Levels Found in the Dataset

(a) After the New Rule

Code	Contaminant Name	Unit	MCL	Obs.	Mean	Std. Dev.	Min	Max
1005	Arsenic*	(mg/L)	0.01	0	---	---	---	---
1009	Chlorite	(mg/L)	1.00	1	1.13	---	1.13	1.13
1010	Barium	(mg/L)	2.00	3	2.50	0.10	2.40	2.60
1025	Fluoride**	(mg/L)	2.00	0	---	---	---	---
1040	Nitrate	(mg/L)	10.00	6	11.88	0.97	10.60	13.60
1041	Nitrite	(mg/L)	1.00	3	1.62	0.10	1.50	1.69
2037	Simazine	(μg/L)	4.00	1	6.80	---	6.80	6.80
2050	Atrazine	(μg/L)	3.00	4	0.00	0.00	0.00	0.01
2456	Haloacetic Acids	(μg/L)	60.00	24	85.25	17.37	62.00	117.30
2950	Total Trihalomethanes	(μg/L)	80.00	36	121.31	31.49	81.00	193.00
2964	Dichloromethane	(μg/L)	5.00	3	6.00	0.00	6.00	6.00
4000	Gross Alpha, Excl. Radon & U	(pCi/L)	15.00	385	27.66	12.74	15.50	73.50
4010	Radium 226 and Radium 228 (Combined)	(pCi/L)	5.00	881	10.65	4.64	5.50	28.20

(b) Before the New Rule

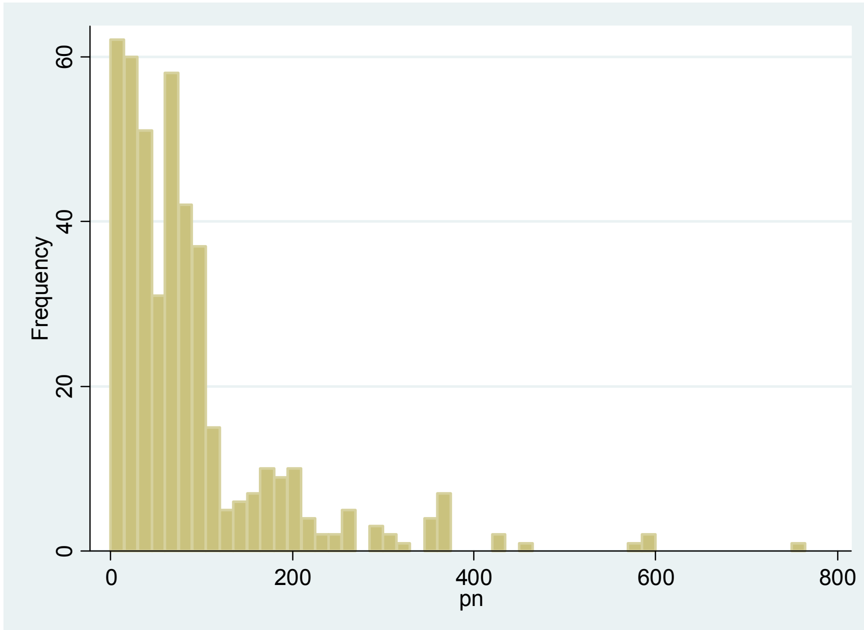
Code	Contaminant Name	Unit	MCL	Obs.	Mean	Std. Dev.	Min	Max
1005	Arsenic*	(mg/L)	0.01	4	0.07	0.01	0.06	0.08
1009	Chlorite	(mg/L)	1.00	0	---	---	---	---
1010	Barium	(mg/L)	2.00	40	4.53	5.35	1.50	25.50
1025	Fluoride**	(mg/L)	2.00	92	3.56	3.67	2.00	27.00
1040	Nitrate	(mg/L)	10.00	114	12.74	1.97	10.50	21.70
1041	Nitrite	(mg/L)	1.00	3	1.70	0.17	1.50	1.80
2037	Simazine	(μg/L)	4.00	0	---	---	---	---
2050	Atrazine	(μg/L)	3.00	14	0.01	0.00	0.01	0.01
2456	Haloacetic Acids	(μg/L)	60.00	0	---	---	---	---
2950	Total Trihalomethanes	(μg/L)	80.00	9	91.22	8.21	87.00	110.00
2964	Dichloromethane	(μg/L)	5.00	0	---	---	---	---
4000	Gross Alpha, Excl. Radon & U	(pCi/L)	15.00	169	24.57	12.44	15.70	99.20
4010	Radium 226 and Radium 228 (Combined)	(pCi/L)	5.00	270	10.26	3.90	5.50	25.80

* Arsenic MCL standard was strengthened from 0.05 mg/L to 0.01 mg/L in 2001.

** Secondary fluoride standard is used; National MCL is 4 mg/L.

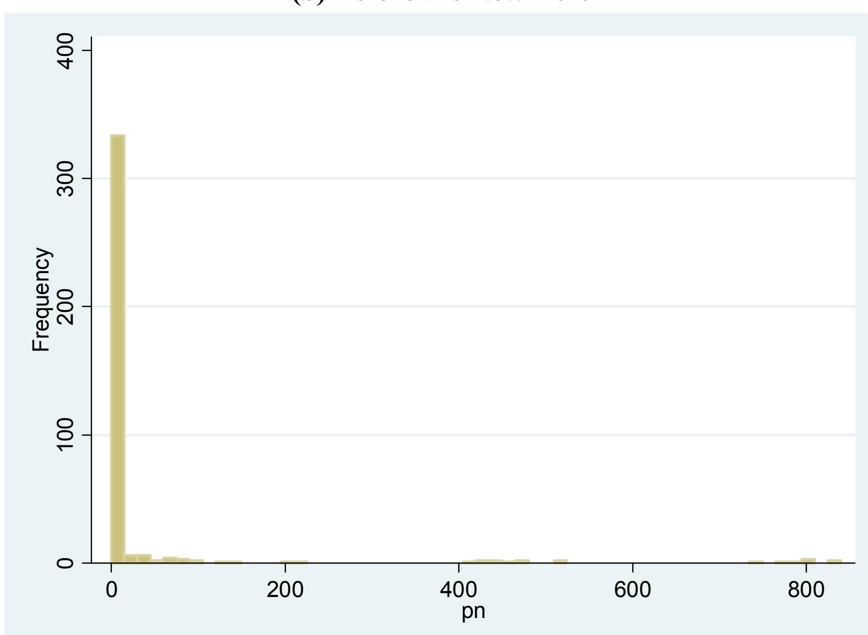
Figure 1. Public Notification Days, Initial MCL Violations, IL

(a) After the New Rule



	Obs.	Mean	Std. Dev.	Min	Max
PN Days	440	87.025	97.773	0	751

(b) Before the New Rule



	Obs.	Mean	Std. Dev.	Min	Max
PN Days	383	48.859	199.934	0	1,843

Table 2. Cancer Morbidity Risk Estimates

	Radium 226 & 228 (Combined)							
	Radium-226				*Radium-228*			
Risk Coefficients	3.85E-10				1.04E-09			
Concentration Levels	MCL	Mean	Min	Max	MCL	Mean	Min	Max
	5.0	10.6	5.5	28.2	5.0	10.6	5.5	28.2
Cancer Morbidity Risk (in one million)								
at 30-th day	0	0	0	1	0	1	0	2
at 90-th day	0	1	0	2	1	2	1	5
at mean PN compliance day *	0	1	0	2	1	2	1	5
at max PN compliance day *	3	6	3	16	8	17	9	44
at mean MCL compliance day	2	4	2	10	5	10	5	27
at max MCL compliance day	4	8	4	21	10	21	11	56
over life time (70 years)	98	209	108	555	266	566	292	1,499

	Gross Alpha Activity, Exl. Radon & Uranium							
	Po-210 (Inorganic polonium)				*Th-232 (Thorium)*			
Risk Coefficients	3.77E-10				1.01E-10			
Concentration Levels	MCL	Mean	Min	Max	MCL	Mean	Min	Max
	15.0	27.7	15.5	73.5	15.0	27.7	15.5	73.5
Cancer Morbidity Risk (in one million)								
at 30-th day	0	1	0	2	0	0	0	0
at 90-th day	1	2	1	5	0	1	0	1
at mean PN compliance day *	1	2	1	5	0	0	0	1
at max PN compliance day *	7	12	7	33	2	3	2	9
at mean MCL compliance day	5	10	6	27	1	3	2	7
at max MCL compliance day	11	21	12	55	3	6	3	15
over life time (70 years)	289	533	299	1,417	77	143	80	379

* Mean and maximum PN days only for initial violations after the new rule.
Source: California EPA, 2006; U.S. EPA, 1999; SDWIS/Fed, EPA.

III. Econometric Model

We estimate regression models to examine the impact of a regulatory change on the parameters of the empirical PN timing function. Our dataset can be seen as a cluster-sample: individual violation may be clustered by the water system for which MCL violations occur. In order to account for cluster-sample nature of the dataset, observations within a cluster are assumed to be correlated due to an unobserved cluster effect. Thus, the equation of interest takes the following form:

$$
\begin{aligned}
PN_{gm} = {} & \beta_0 + \beta_1 severity_{gm} + \beta_2 newrule_{gm} + \beta_3 (severity \times newrule)_{gm} + \beta_4 rep_{gm} + \boldsymbol{\beta}_5' \boldsymbol{chem}_{gm} \\
& + \boldsymbol{z}_g' \boldsymbol{\gamma} + \boldsymbol{w}_g' \boldsymbol{\delta} + v_{gm}, \qquad m = 1,\ldots,M_g;\ g = 1,\ldots,G
\end{aligned} \tag{1}
$$

where PN_{gm} denotes the PN timing of m-th MCL violation of water system g; *severity* is the contamination level normalized with respect to MCL standards, i.e. (contamination level found - MCL standard)/MCL; *newrule* is a binary dummy variable indicating if a violation occurs after the new PN regulations became fully effective; $severity \times newrule$ is an interaction variable; *rep* is a binary dummy describing whether a violation is initial or recurring; ***chem*** is a 9×1 vector of dummy variables describing the contaminants, namely, Atrazine, Barium, Fluorde, Gross Alpha Activity, Haloacetic Acids, Nitrate, Nitrite, Radium 226 and Radium 228 (combined), and Total Trihalomethanes; $\boldsymbol{z_g}$ is a vector of variables describing system g; $\boldsymbol{w_g}$ is a vector of variables that approximate the characteristics of the population served by system g; v_{gm} is a composite error term; m indexes observations within cluster; M_g is the number of observations in each system g; and G is the number of water systems.

$\boldsymbol{z_g}$ includes three explanatory variables: source of water (*gw*), ownership (*gov*), and the size of population served (*pop*). *gw* is a dummy variable that describes whether a water source is groundwater (*gw* = 1) or surface water (*gw* = 0). There are five types of ownership: privately owned, state owned, federally owned, local government owned, and half-private half-public. In the usable dataset, we do not observe any half-private half-public water systems. Thus, we classified PWSs into publicly-owned (*gov* = 1) or privately-owned (*gov* = 0). To classify the size of population served by a water system, we apply the size category defined by EPA. *pop* is an ordinal categorical variable ranging 1-8 with 8 being the largest.

$\boldsymbol{w_g}$ includes variables that describe the population characteristics of the community served: *white*, the percentage of white population in the community; *age5*, the percentage of population under age five; *educ*, the percentage of community population with high school diploma; and *income*, the median household income of the community. To construct these variables, we used the U.S. Census 2000 data as approximation of the community population characteristics. The PWS address information was used to match the smallest geographic areas (town, city, village, CDP) available in the Census.[4]

Table 3 reports descriptive statistics of the variables used in the study.

Table 3. Descriptive statistics of variables

Variable	Units	Mean	Std. Dev.	Min	Max
pn	days	59.40	113.60	0.00	1843.00
severity	--	1.00	1.26	0.00	24.50
newrule	--	0.65	0.48		
rep	--	0.60	0.49		
chem1	--	0.02	0.14		

[4] In a few occasions, the PWS address information in the compliance and enforcement data only provide street names. We used GIS information to match the closest township and then applied the Census information from that township.

chem2	--	0.04	0.21		
chem3	--	0.06	0.23		
chem4	--	0.00	0.05		
chem5	--	0.01	0.09		
chem6	--	0.01	0.11		
chem7	--	0.02	0.15		
chem8	--	0.56	0.50		
chem9	--	0.27	0.44		
gw	--	0.88	0.32		
gov	--	0.74	0.44		
pop	--	3.76	2.02	1.00	8.00
white	%	0.93	0.10	0.42	1.00
age5	%	0.07	0.02	0.02	0.12
educ	%	0.35	0.12	0.07	0.65
income	$1,000	51.45	24.91	21.25	148.15

The composite error term v_{gm} in (1) may contain an unobserved common group effect, and thus, can be written:

$$v_{gm} = c_g + u_{gm}, \qquad m = 1,\ldots, M_g, \tag{2}$$

where c_g is the unobserved group effect and u_{gm} is an idiosyncratic error. As with the case of panel-data methods, the assumptions on correlation between the explanatory variables in (1) and the composite error term is crucial to cluster-sample models. Because our interest is the impact on the PN timing not only of covariates that vary within as well as across water systems, but also of the ones that vary only at the water system level, we estimate the parameters of equation (1) by pooled ordinary least squares (pooled OLS) and by random effects (RE) with a fully robust variance matrix. With the variance estimator of RE, standard errors are "robust to arbitrary heteroskedasticity and with-in group correlation" (Wooldridge, 2006).

Although the linear model in (1) produces results that are easy to interpret, the estimated parameters may not be consistent because of the censored observations for which the observed PN timing equals zero. In Wooldridge's (2002) terms, these observations are the "corner-solution outcomes" of the response variable. As discussed in the previous section, the dataset contains large positive probability mass at $PN = 0$. Inconsistency cannot be avoided even if we estimate the above model restricted to the subsample with $PN > 0$ (Wooldridge, 2002; Cameron and Trivedi, 2005). In order to overcome this problem together with unobserved heterogeneity discussed above, we also estimate the following standard Tobit model with a cluster effect:

$$PN_{gm} = \max(0, \beta_0 + \boldsymbol{x}'_{gm}\boldsymbol{\beta} + \boldsymbol{z}'_g\boldsymbol{\gamma} + \boldsymbol{w}'_g\boldsymbol{\delta} + v_{gm}) \tag{3}$$

$$v_{gm} \mid \boldsymbol{X}_g, \boldsymbol{z}_g, \boldsymbol{w}_g, c_g \sim Normal(0, \sigma_v^2) \tag{4}$$

where $\boldsymbol{x}_{gm}$ contains $severity_{gm}$, $newrule_{gm}$, $(severity \times newrule)_{gm}$, and rep_{gm}; $\boldsymbol{X}_g$ includes unity and $\boldsymbol{x}_{gm}$ for all m; σ_v^2 is the conditional variance, $\mathrm{Var}(v_{gm} \mid \boldsymbol{X}_g, \boldsymbol{z}_g, \boldsymbol{w}_g, c_g)$; and other variables are the same as above. Analogously to the linear model estimation, we estimate (3) and (4) by the pooled Tobit and the RE Tobit. A fully robust approach of RE can be applicable, in principle, in the Tobit case; however it is not common in standard statistical packages, such as STATA, yet and details still need to be done (Wooldridge, 2006).

The inconsistency of linear regression estimates does not necessarily mean that the results are of no use. Linear regression estimates can still approximate the impact of explanatory variables on the conditional expectation of a dependent variable when those covariates are near its population means (Wooldridge, 2002).

In order to make the Tobit coefficients comparable to the OLS estimates, one can use the adjustment factors derived from the following equations:

$$\mathrm{E}(PN \mid \boldsymbol{x}) = \Phi(\boldsymbol{x}'\boldsymbol{\beta}/\sigma_v)\boldsymbol{x}'\boldsymbol{\beta} + \sigma_v \phi(\boldsymbol{x}'\boldsymbol{\beta}/\sigma_v), \tag{5}$$

$$\partial \mathrm{E}(PN \mid \boldsymbol{x}) \big/ \partial x_j = \Phi(\boldsymbol{x}'\boldsymbol{\beta}/\sigma_v)\beta_j, \tag{6}$$

where $\boldsymbol{x}$ contains all explanatory variables; $\boldsymbol{\beta}$ is a vector of coefficients of $\boldsymbol{x}$; x_j is a variable in question; $\Phi(\cdot)$ is standard normal distribution; $\phi(\cdot)$ is standard normal density; and we assume x_j is not functionally related to other regressors, but this assumption can be relaxed easily. Derivation of equation (5) and (6) can be found in Wooldridge (2002).

The specification including an interaction term of *severity* and *newrule* is useful for estimating the impact of contamination level along with regulatory change on PN timing and testing if the impact is significant. In the linear model, the effect of contamination level on PN timing before the regulatory change is measured as β_1, whereas its impact after regulatory change is measured as $\beta_1 + \beta_3$. These coefficients can be written as the partial effect of each variable on $\mathrm{E}(PN \mid \boldsymbol{x})$ as follows:

$$\left.\frac{\partial \mathrm{E}(PN \mid \boldsymbol{x})}{\partial (severity)}\right|_{newrule=0} = \beta_1, \tag{7}$$

$$\left.\frac{\partial \mathrm{E}(PN \mid \boldsymbol{x})}{\partial (severity)}\right|_{newrule=1} = \beta_1 + \beta_3. \tag{8}$$

On the other hand, with the pooled Tobit model the partial effects become:

$$\left.\frac{\partial \mathrm{E}(PN \mid \boldsymbol{x})}{\partial (severity)}\right|_{newrule=0} = \beta_1 \Phi(\boldsymbol{x}'\boldsymbol{\beta}/\sigma_v)\big|_{newrule=0}, \tag{9}$$

$$\left.\frac{\partial \mathrm{E}(PN \mid \boldsymbol{x})}{\partial (severity)}\right|_{newrule=1} = (\beta_1 + \beta_3)\Phi(\boldsymbol{x'\beta}/\sigma_v)\big|_{newrule=1}. \tag{10}$$

Unlike the linear model, these estimates depend on $\boldsymbol{x}$. Further, given unobserved heterogeneity with RE Tobit, one can obtain the average partial effects (APEs) in order to examine the impact of contamination level on PN timing. Following Wooldridge (2002, 2006), it is convenient to define $m(z,\sigma^2) \equiv \Phi(z/\sigma)z + \sigma\phi(z/\sigma)$, so that $\mathrm{E}(PN \mid \boldsymbol{x}, c) = m(\boldsymbol{x'\beta} + c, \sigma_v^2)$. Then the APEs can be estimated by taking the numerical partial derivatives of the expression:

$$G^{-1}\sum_{g=1}^{G} m(\hat{\beta}_0 + \boldsymbol{x}^{\circ}\hat{\boldsymbol{\beta}} + \boldsymbol{z}'_g\hat{\boldsymbol{\gamma}} + \boldsymbol{w}'_g\hat{\boldsymbol{\delta}}, \hat{\sigma}_v^2), \tag{11}$$

with respect to elements of $\boldsymbol{x}^{\circ}$ (*severity* and *newrule* for the study), where $\boldsymbol{x}^{\circ}$ is a given value of some of the explanatory variables. Detailed derivation of equation (11) can be also found in Wooldridge (2002, 2006).

IV. Results

4.1 Pooled OLS and Fully robust RE

The coefficient estimates of the linear model (1) using pooled OLS and fully robust RE are reported in column 1 and 2 of Table 4, respectively. The coefficient estimates from both models have similar signs and statistical significance levels, except for some of contaminants' dummy variables. As one might expect from Figure 1, the effect of *severity* on PN timing before the change ($\beta_1 = -2.47$ for OLS and -2.14 for RE) is not significantly different from zero. On the other hand, the *severity*'s impact after the new rule ($\beta_1 + \beta_3 = 8.21$ for OLS and 9.12 for RE) is positive and statistically significant ($F[1, 306] = 4.91$, $p < 0.03$) for both models. It indicates that the *severity* of the contamination incidence has affirmative impact on PN delay after the new rule enforcement. In other words, after the new rule, consumers are notified later when contamination incidences are severer and thus they should be notified sooner. Most contaminant dummies are insignificant, but those are jointly significant at less than 1% level for both models. There are several contaminants other than above nine contaminants in the dataset; however, the numbers of observation for those contaminants are too few and make estimation with Tobit-type models below instable. Thus, for comparison purpose, we omitted those contaminants from our estimation.

Table 4. Estimation Results

	(1)	(2)	(3)	(4)
Independent variables	Pooled OLS	Fully robust RE	Pooled Tobit	RE Tobit
severity	−2.47	−2.143	−19.263	−18.764
	(1.19)	(1.04)	(1.92) *	(1.89) *
newrule	24.914	21.464	130.833	129.545
	(2.22) **	(1.77) *	(8.85) ***	(8.81) ***
severity × newrule	10.675	11.264	28.044	27.904
	(2.44) **	(2.45) **	(2.55) **	(2.56) **
rep	−28.636	−28.502	−45.678	−45.918
	(5.42) ***	(5.70) ***	(5.86) ***	(5.93) ***
chem1	−26.645	−1.159	−89.344	−88.066
	(0.53)	(0.03)	(1.29)	(1.28)
chem2	−63.2	−37.545	−300.999	−296.617
	(1.23)	(1.10)	(3.72) ***	(3.69) ***
chem3	−32.578	−4.572	−65.44	−62.226
	(0.61)	(0.12)	(1.04)	(1.00)
chem4	−89.182	−62.344	−100.987	−96.698
	(1.71) *	(1.78) *	(1.15)	(1.11)
chem5	−10.594	14.31	121.062	120.474
	(0.20)	(0.41)	(1.79) *	(1.79) *
chem6	−16.114	5.617	−20.319	−23.362
	(0.30)	(0.15)	(0.31)	(0.36)
chem7	−20.058	5.961	−25.547	−27.025
	(0.37)	(0.16)	(0.41)	(0.43)
chem8	−20.164	5.457	−43.288	−39.874
	(0.39)	(0.16)	(0.74)	(0.68)
chem9	−13.448	12.485	−40.732	−36.815
	(0.27)	(0.37)	(0.69)	(0.63)
gw	17.404	17.279	56.842	54.609
	(1.53)	(1.51)	(3.45) ***	(3.33) ***
gov	16.294	16.037	16.268	16.423
	(1.46)	(1.59)	(1.36)	(1.37)
pop	−9.621	−10.506	−15.512	−15.964
	(2.75) ***	(3.09) ***	(4.89) ***	(5.00) ***
white	−120.113	−123.38	-191.145	-196.949
	(2.88) ***	(2.93) ***	(4.04) ***	(4.14) ***
age5	−384.67	−251.135	−638.498	−639.501
	(1.65)	(1.18)	(2.37) **	(2.36) **
educ	−88.135	−81.302	−126.711	−124.7
	(1.85) *	(1.78) *	(2.16) **	(2.11) **
income	−0.339	−0.298	−0.557	−0.551
	(2.04) **	(1.96) *	(2.44) **	(2.41) **
Observations	2062	2062	2062	2062
R-squared	0.08	--	0.06	0.06
Number of cluster		307		307

Note: In parentheses are t-statistics in absolute value; * significant at 10%; ** significant at 5%; *** significant at 1%.

The results on other regression coefficients are also interesting. Population size, the fraction of white population, education level, and income level of the community have all negative and significant impacts on PN timing. These results make intuitive sense. The water managers of PWSs are likely to be aware that if they serve a large population, a number of consumers who might get sick (= cancer risk coefficients x the size of the population) becomes large. The water managers may also believe that the communities with high income, high education, and high white population can exercise political pressures on their job security if they are accused of issuing a public notice too late. These factors may affect private incentives of the managers to decide when to issue a public notice. The results suggest that *pop*, *educ*, *white*, and *income* influence private valuation of PN timing in a positive manner to the consumers while *severity* influence it in a negative manner, implying that high contamination incidence gives the water managers incentives to hide or delay notifying the public of that incidence (for whatever reason!).

4.2 Pooled Tobit and RE Tobit

Pooled Tobit and RE Tobit coefficient estimates are reported in column 3 and 4 of Table 4. As shown in the table, virtually all coefficients have the same signs as the pooled OLS or RE estimates at the similar statistical significance levels. In fact, the Tobit results reinforce the results from the linear models. The coefficients on *severity* (before new rule), *gw*, and *age5* become statistically significant at least 10% level. Many contaminant chemical dummies are insignificant, these are jointly significant again at less than 1% level for both pooled and RE Tobit model. As discussed in the previous section, the coefficients of Tobit-type models on continuous explanatory variables can be adjusted in order to make them comparable to the OLS estimates by multiplying the adjustment factors in (6). The factor in (6), evaluated at the mean values of $\boldsymbol{x}$'s, is about 0.482 for RE Tobit. In most cases the estimated Tobit impacts at the mean values are below the corresponding OLS (or RE) estimates. For example, the Tobit effect on household income is about $0.482\times(-0.551)\approx-0.266$, which is smaller than the OLS (RE) estimate of –0.339 (–0.298). One thousand dollars of median household income in the community is estimated to decrease expected PN timing by about 0.27 days. Based on the likelihood ratio test comparing the pooled Tobit to RE Tobit estimators, RE Tobit estimator is preferred to the pooled Tobit estimator [$\chi_1^2=144.5$, $p < 0.01$].

The impact of the contamination level on PN timing of the Tobit models can be evaluated with equations (9), (10), and (11). The results are summarized and compared to the ones with the linear models in Table 5. The results clearly show that before the new rule was enforced, the contamination level has had negative and relatively insignificant impact on the expected PN timing, whereas it has positive and significant impact after the new rule. This result is particularly alerting, as it implies that consumers are informed later when they face greater health risks. Because the chemical contaminants covered in this analysis pose latent, chronic health risks, the water managers may have incentives to delay PN timing.

The estimated expected PN timing estimated with Tobit-type models are smaller than those estimated with the linear models.

Table 5. Impact of contamination level on expected PN timing

	Before new rule	After new rule
Pooled OLS	–2.47	8.21
	(1.19)[a]	(4.91)**[b]
Fully robust RE	–2.14	9.12
	(1.04)[a]	(4.91)**[b]
Pooled Tobit	–4.33	5.51
	(3.68)*[c]	(3.57)*[c]
RE Tobit	–4.67	5.72
	(3.56)*[c]	(3.88)** [c]

Notes: [a] t-statistics (absolute value), [b] F-statistics, and [c] Wald statistics calculated by Delta methods; Asterisks * and ** indicate significance at the levels 10% and 5%, respectively.

V. Concluding Remarks

Providing quality information on health-threatening contamination incidences can improve consumer welfare, because the consumers can, and often do, take self-protective actions given the information provided. Unfortunately, important contamination information often comes from private or quasi-private sources and the distinction between private and public information is also sometimes ambiguous. Thus, the regulatory authority may need to manage “private” incentives of providing publicly useful information. There is no lack of literature arguing that supply of public goods is smaller than optimal when provision of those goods are in the hands of private agents seeking their own economic rents. The same argument seems to apply here --- private provision of publicly useful information is less or later than optimal.

Though the argument makes an intuitive sense, there appears to be lack of empirical research investigating this issue. We provide one such study, examining a unique EPA dataset on the public notification enforcement and compliance for the Safe Drinking Water Act. The data contain 2,062 observations of maximum contaminant level violation incidences in Illinois from 1980 to 2006. The dataset allows us to estimate the determinants of the observed public notification timings. More importantly, we also estimate the impact of the new public notification rule on the coefficient for the contamination severity level. The new rule, issued May 2000 and fully enforced May 2002, (i) extended the PN compliance deadline (for tier 2 violations) of 14 days to 30 days and (ii) gives the primacy agency discretion to extend the PN deadline in some appropriate circumstances. Our results are disconcerting and indicate that the new rule has had welfare-decreasing effects. With linear models, one point increase in the severity level of a contamination incidence is significantly associated with an increase in PN delay of 8-9 days after the new rule but is not statistically significantly associated with PN delay before the new rule.

Moreover, the new rule per se increases the PN timing by 21-24 days, which is approximately the simple effect of the extension rule. These results were robust to other specifications, in particular, to the Tobit specifications to account for censoring issues.

Our empirical results seem to suggest that even in the case of mandated public notification, private or quasi-private managers of the public water systems still make private calculation of the benefits and costs of making information public --- the PN timing is significantly associated with various covariates, which one should not observe if the managers were truly welfare-maximizing agents. These findings should be taken seriously in lieu of guiding future information disclosure/public notification policies.

However, there are several caveats to our results. First, our results are solely based on observations from Illinois. We have had access to the full data set for all states in Region 5, namely, Illinois, Indiana, Michigan, Minnesota, Ohio, and Wisconsin. However, we have had difficulty using these data because of either insufficient observations or high irregularities in reported data. Moreover, even for the Illinois data, we have had difficulty constructing the "PN timing" variable. Thus, it should be noted that there are apparent data problems, as with any empirical analyses. Second, though we have run several robustness checks, we have not tried duration models, which might have produced different results. The main reason for this is because we have had too many censored observations and we were unable to conceptually justify the use of duration models for our dataset. Extension of our analyses using duration models is, therefore, left for future research.

Reference

Cameron, A.C., Trivedi, P.K., 2005. Microeconometrics: methods and application. Cambridge University Press, New York.

Christensen, P.O., Feltham, G.A., 2001. Efficient Timing of Communication in Multiperiod Agencies. *Management Science* 40(2), 280-294.

Colantoni, C.S., Davis, O.A., Swaminuthan, M., 1976. Imperfect Consumers and Welfare Comparisons of Policies Concerning Information and Regulation. *Bell Journal of Economics* 7(2), 602-615.

Gayer, T., Hamilton, J.T., Viscusi, W.K., 2000. Private Values of Risk Tradeoffs at Superfund Sites: Housing Market Evidence on Learning about Risk. *Review of Economics and Statistics* 82(3), 439-451.

Ippolito, P.M., Mathios, A.D., 1990. Information, Advertising and Health Choices: A Study of the Cereal Market. *RAND Journal of Economics* 21(3), 459-480.

Melumad, N.D., Reichelstein, S., 1987. Centralization versus Delegation and the Value of Communication. *Journal of Accounting Research* 25, 1-18.

Melumad, N.D., Shibano, T., 1991. Communication in Settings with No Transfers. *RAND Journal of Economics* 22(2), 173-198.

Madajewicz, M., Pfaff, A., Geen, A., Graziano, J., Hussein, I., Momotaj, H., Sylvi, R., Ahsan, H., 2006. Can Information Alone Change Behavior? Response to Arsenic Contamination of Groundwater in Bangladesh. *Journal of Development Economics*, doi: 10.1016/j.jdeveco.2006.12.002.

Pitchik, C., Schotter, A., 1987. Honesty in a Model of Strategic Information Transmission. *American Economic Review* 77(5), 1032-1036.

Smith, V.K., Desvousges, W.H., 1990. Risk Communication and the Value of Information: Radon as a Case Study. *Review of Economics and Statistics* 72(1), 137-142.

Smith, V.K., Desvousges, W.H., Fisher, A., Johnson, F.R., 1988. Learning about Radon's Risk. Journal of *Risk and Uncertainty* 1, 233-258.

Smith, V.K., Johnson, F.R., 1988. How Do Risk Perceptions Respond to Information? The Case of Radon. Review of Economics and Statistics 70(1), 1-8.

U.S. Environmental Protection Agency, 1999a. Cancer Risk Coefficients for Environmental Exposure to Radionuclides. Federal Guidance Report No. 13.

U.S. Environmental Protection Agency, 1999b. Response to Public Comments on EPA’s Proposed Public Notification Rule.

U.S. Environmental Protection Agency, 2000a. National Primary Drinking Water Regulations, Public Notification Rules, Final Rule. Federal Register, 40 CFR Part 9 *et al.*

U.S. Environmental Protection Agency, 2000b. Public Notification Handbook.

U.S. General Accounting Office, 1992. Drinking Water: Consumers Often Not Well-informed of Potentially Serious Violations, Report # GAO/RCED-92-135.

Viscusi, W.K., Magat, W.A., Huber, J., 1986. Informational Regulation of Consumer Health Risks: An Empirical Evaluation of Hazard Warnings. *RAND Journal of Economics* 17, 351-365.

Wooldridge, J.M., 2002. Econometric Analysis of Cross Section and Panel Data. The MIT Press: Cambridge, MA.

Wooldridge, J.M., 2006. Cluster-sample methods in applied econometrics: an extended analysis. Michigan State University Department of Economics, Working Paper.