Supporting Information

Olmstead et al. 10.1073/pnas.1213871110

SI Text

Choice of Dependent Variables

The dependent variables for the analysis, concentrations of chloride (Cl^-) and total suspended solids (TSS) at Pennsylvania water quality monitors, are drawn from the Storage and Retrieval Data Warehouse (STORET) database of the US Environmental Protection Agency (EPA). Although the observational surface water quality data available in the STORET database are rich, they are not comprehensive for the analysis of shale gas development impacts. For example, naturally occurring radioactive materials and heavy metals may be mobilized from deep shale formations during hydraulic fracturing, ending up in flowback and produced water, and eventually in surface water (1). However, the number of Pennsylvania water quality monitors that regularly sample for these contaminants is small. About 400 observations each of α- and β-particle concentrations are available for Pennsylvania monitors in the STORET database, 2000–2011, from 20 monitors. The only heavy metals with Pennsylvania monitors reporting in the STORET database are arsenic (115 monitors) and strontium (131 monitors); reported observations for these are about 33% and 15%, respectively, of the observations reported for Cl^-, which has the smaller sample size between the two contaminants we do analyze. Because we need sufficient observations before and after drilling and waste shipment to identify the effects, these samples are simply not large enough for econometric analysis like that performed here. Thus, we are limited by available data in the water quality parameters we can examine.

Second, we are limited by our ability to control for sources of variation in contaminant concentrations that could be both (*i*) correlated with shale gas development over time and space (wells or waste treatment) and (*ii*) insufficiently explained by the additional independent variables in our models, biasing coefficient estimates. An important example is total dissolved solids (TDSs). High TDS concentrations due to acid mine drainage are a legacy of coal mining in many Pennsylvania rivers and streams. If we were to use TDS concentration as a dependent variable (with almost 12,000 Pennsylvania observations in the STORET database, 2000–2011), we would face the potential concern of spatial or intertemporal correlation of abandoned and active mines with shale gas wells or waste treatment facilities. This is unlikely, but it is nonetheless an important reason why we do not model TDS concentrations. Further, the main constituent of TDS in mine drainage is sulfate, rather than Cl^-, supporting our choice to model Cl^- (2).

Finally, we obtained some additional data, not otherwise used in the analysis, to support our choice of dependent variables. Shale gas operators shipping waste to Pennsylvania treatment facilities are required by the state to submit a Chemical Analysis of Residual Waste Annual Report by the Generator, also known as Form 26R. Form 26R includes descriptive information about the waste generator, although only in rare cases is this information detailed enough to connect a waste sample to a particular shale gas well. It also includes a waste description, using standard codes for brine, fracturing fluids, and drill cuttings, for example, as well as a detailed chemical analysis performed by a state-accredited laboratory. For shale gas waste, the chemical analysis must include, at a minimum, a list of 52 analytes, including Cl^- and TSS.

Operators submit these forms to the Pennsylvania Department of Environmental Protection (PADEP) in hard copy. During August 2011–December 2011, we collected these data via personal visits to the four regional PADEP offices with significant shale gas development (Southwest, Northwest, Northcentral, and Northeast). All available Form26Rs for Marcellus Shale gas waste were scanned to portable document format, converted to spreadsheet format, cleaned, and then imported to a statistical software program for analysis.

The Form26R data contain Cl^- observations from 191 different shale gas waste samples, with an average concentration of just over 60,781 mg/L; TSS observations from 171 different shale gas waste samples have an average concentration of 663 mg/L. To the extent that these samples accurately characterize the waste shipments tracked by the PADEP that we use for the statistical analysis, this indicates that Cl^- concentration is, in fact, a good potential indicator for shale gas waste in Pennsylvania surface water and that waste disposal could also be associated with downstream TSS increases.

SI Data

Contaminant Concentrations. Between January 2000 and December 2011, 1,049 water quality monitors in Pennsylvania report at least one sample result in the STORET database for Cl^-, TSS, or both. The full sample for the Cl^- regressions comprises 8,364 observations from 860 monitors, with a mean observed Cl^- concentration, 2000–2011, of 19.1 mg/L (Table S1). The full sample for the TSS regressions comprises 11,919 observations from 644 monitors, with a mean observed TSS concentration, 2000–2011, of 20.4 mg/L (Table S2). Graphical depictions of Cl^- and TSS concentrations for monitors that have at least one upstream shale gas well by December 2011 (Fig. S1) or at least one upstream wastewater treatment facility treating shale gas waste through December 2011 (Fig. S2) show that concentrations fluctuate significantly over time, even when averaged at the half-year point. The graphs also suggest a weak correlation between the average observed TSS concentration and the average spatial density of upstream wells (Fig. S1). A potentially stronger correlation is observed between average Cl^- concentration and average density of upstream wastewater treatment facilities accepting shale gas waste (Fig. S2). The models we estimate exploit this observed variation and, unlike Figs. S1 and S2, control for confounding factors. These figures present average concentrations and densities across many monitors, whereas our regression estimates attribute to each monitor its monitor-specific upstream density.

Precipitation. Daily precipitation data from 515 Pennsylvania weather stations were obtained for 2000–2011 from the Global Historical Climatology Network of the National Oceanic and Atmospheric Administration's National Climate Data Center. We take the average sum (for all weather stations in a monitor's watershed) of precipitation on the day of each water quality monitor observation plus 3 d prior.

Shale Gas Well Locations and Dates. The shale gas well locations and entry dates (2003–2011) were obtained from two sources: the PADEP "Spud Data" dataset and the Pennsylvania Department of Conservation and Natural Resources (PADCNR) Wells Information System (IRIS/WIS) dataset. The PADEP Spud Data dataset includes records of 4,692 wells drilled in Pennsylvania, and the PADCNR IRIS/WIS dataset includes records of an additional 216 wells. Combined, these data sources gave us the locations of 4,908 wells through December 2011. A well enters our panel on the spud date or completion date, whichever is earlier. For 95% of the wells, we use the spud date. Only 1,815 wells have a completion date. We use this date only if the spud date is missing or if the reported spud date is after the completion date (1.5% of wells). In addition, we use the average length of time between spud and completion (80 d), where both dates are reported, to support our approach to investigating the differential impact of wells on Cl^- concentrations during well completion.

Facilities Accepting Shale Gas Waste. Data on shipments of waste from Marcellus Shale gas wells to wastewater treatment plants were obtained from the PADEP Oil and Gas Reporting website. The data also report shipments of drill cuttings to landfills, brine to injection wells in neighboring states, and conventional oil and gas brine spread on roads for deicing. The current study only examines shipments of Marcellus Shale waste that are indicated to have gone to treatment facilities in watersheds completely or partially in Pennsylvania (this includes shipments to facilities in neighboring states in watersheds with downstream portions in Pennsylvania). For the majority of shipments, the latitude and longitude of the waste treatment plants are recorded. Where location was missing from the PADEP data, we used other identifiers [the facility's National Pollutant Discharge Elimination System (NPDES) permit number, name, and/or street address] to determine location.

The first recorded waste shipment occurs in 2004, and shipments continue through the end of December 2011 (Fig. S2). Operators reported these data to the PADEP annually from 2004 to 2009 and twice per year in 2010–2011. From 2004 to 2011, 74 treatment facilities treated shale gas waste from Pennsylvania at some point; 50 of these are upstream of at least one monitor in our sample. We include any shipments sent to a "municipal sewage treatment plant" or a "brine or industrial waste treatment plant." We downloaded the data in early 2012. These have since been relabeled in the online data as "public sewage treatment plants" and "centralized treatment plant for recycle," respectively. It is important to note that recycling in this context includes treatment for both direct reuse and discharge to the environment (3).

To identify the number of facilities upstream in a water quality monitor's watershed treating shale gas waste on the date that a sample is drawn, because we do not know the exact dates of shipments, we assume that the facilities identified as receiving waste in a given report begin receiving waste on the date at the midpoint of the reporting period and stop accepting waste at the date at the midpoint of the last period in which they are observed accepting waste. For facilities that accepted waste in July 2011–December 2011 (the last reporting period in the data), we assume that they received waste for the entire period. There are no reported Marcellus Shale waste shipments in 2007; thus, we assume those facilities that were accepting waste in 2006 and in 2008 were also accepting waste in 2007 (and, likewise, that facilities not accepting waste in 2006 and 2008 were not accepting waste in 2007). We test the sensitivity of results to these assumptions in the robustness section below.

Waste Shipment Quantities. Like the waste facility destination data, the waste quantity data are reported to the PADEP as annual quantities from wells to facilities from 2004 to 2009 and biannually in 2010–2011. However, the waste quantities reported before 2010 are problematic. Fifty percent of observations before 2010 report the exact same quantity of waste sent by operators from one well to multiple treatment facilities in the same period, indicating that operators often do not specify waste quantities shipped to individual facilities. The PADEP Oil and Gas Reporting website notes that reporting an average quantity of waste shipped to all facilities is an acceptable practice for operators. There are no other publicly available data on the quantity of shale gas waste shipped to individual treatment facilities.

Estimating Density of Wells and Waste Treatment. We divide the number of shale gas wells upstream and the number of waste treatment facilities upstream by the area of the watershed that is upstream of each monitor to obtain the shale gas development density variables for the statistical analysis. To obtain the denominator in these variables, the watershed area that is upstream of each monitor, we used the flow length tool in the ArcGIS surface hydrological toolset of the Geographic Information System (GIS). Flow length is the distance traveled from any cell along the surface flow network to an outlet. For each cell, we calculated the flow length to the closest headwater location or stream outlet. If the flow length of the monitor's cell is smaller than the flow length of the cell a shale gas well or waste treatment facility is in, the monitor is considered downstream of the well or waste treatment facility. We then sum the area (square kilometers) of all the cells in the watershed with flow lengths greater than the monitor's cell to estimate upstream area.

Main Econometric Models

To identify the impact of shale gas development on Cl^- and TSS concentrations, we must adequately control for other contaminant sources. The primary sources of Cl^- in surface water include natural deposition from dissolution of geological salt deposits, irrigation drainage, sea spray and seawater intrusion in coastal areas, sewage and industrial effluent, land development in a watershed, and road salt (4, 5). Higher stream flow generally reduces Cl^- concentrations through dilution. The primary anthropogenic source, and the primary source in many regional watersheds, is road salt (6). Important sources of TSS include heavy precipitation and soil erosion from land disturbance, treated conventional wastewater, and septic system leakage.

Rather than controlling individually for each potential source, we take the standard econometric approach of controlling comprehensively for potential confounders using sets of fixed effects (FEs). The basic estimating equation is Eq. **S1**, in which C_{it} is the contaminant concentration (milligrams per liter) measured at monitor i on day t, $monitor_i$ is a monitor FE, wm_{it} is a watershed-calendar-month FE, ym_t is a FE for each of the 132 mo between January 2000 and December 2011, p_{it} is the 4-d precipitation variable, and ε_{it} is a standard econometric error term:

$$C_{it} = monitor_i + wm_{it} + ym_t + \alpha p_{it} + \gamma W_{it} + \beta T_{it} + \varepsilon_{it}. \qquad \textbf{[S1]}$$

Variables indicating shale gas wells upstream in the watershed are denoted by W_{it}, and those describing shale gas waste treatment upstream in the watershed are denoted by T_{it}. The intertemporal variation in W_{it} comes from the fact that wells are being added continuously in some watersheds and not others between 2003 and 2011. In contrast, the intertemporal variation in T_{it} does not come from the construction and permitting of new treatment facilities (the stock of waste treatment facilities is constant between 2000 and 2011) but from the fact that facilities start and stop accepting shale gas waste over time.

Additional Descriptive Models

By including the large set of FEs in the contaminant regressions, we reduce potential omitted variables bias in the estimates of shale gas development impacts but we lose descriptive power and make it difficult to compare results with water quality analyses outside of the economics literature. To address this situation, after estimating the models reported in Tables 1 and 2, we recover the estimated FEs $(\widehat{monitor}_i, \widehat{wm}_{it}, \widehat{ym}_t)$ and regress their sum on descriptive variables that might intuitively be included in the contaminant regressions but cannot be because they are subsumed by one or more of the FEs, which control for all non–time-varying characteristics of water quality at the monitor, seasonality in the watershed, and time trends. The basic estimating equation is Eq. **S2**, in which Z is a vector of descriptive characteristics (different for Cl^- and TSS models) and v_i is a standard econometric error term:

$$\widehat{monitor}_i + \widehat{wm}_{it} + \widehat{ym}_t = Z_i^{'}\delta + v_i. \qquad \textbf{[S2]}$$

In the case of Cl^-, Z includes the count of all NPDES-permitted waste treatment facilities (whether or not they accept shale gas waste) upstream in the monitor's watershed, road kilometers × lanes, and a single modeled estimate of stream flow at the monitor (Table S3). In the case of TSS, Z includes land cover variables and modeled stream flow (Table S4). The results of these models must account for the fact that the dependent variable in Eq. **S2** is estimated; thus, we bootstrap to obtain consistent SEs. Each of the 100 bootstraps involves resampling, with replacement, STORET monitor clusters, and reestimating both Eqs. **S1** and **S2**. The reported SE is the SD of the sample of bootstrapped coefficient estimates.

Additional data were required to implement this descriptive analysis. We obtained the location of NPDES facilities from the EPA's Permit Compliance System and Integrated Compliance Information System, with the help of EPA staff. When an NPDES facility had multiple pipes, we used the location of the pipe closest to the water quality monitor. We calculated the kilometers × lanes of road in each watershed using the state road shape file from the Pennsylvania Department of Transportation. The density of roads for each watershed was calculated based on the kilometers of road in the watershed, multiplied by the number of lanes per road, divided by watershed area. The land cover data for the TSS model, and stream flow estimates for both models, were obtained from the National Hydrography Dataset.

The results in Table S3 suggest that the spatial density of roadways in a monitor's watershed increases the portion of average Cl^- concentrations explained by the set of FEs in our main models (an expected impact related to road salt). Increases in the non–time-varying portion of stream flow do not contribute a statistically significant portion of the variation in Cl^- concentrations explained by the FEs (recall that we control for precipitation in the main models). The density of NPDES-permitted facilities in a monitor's watershed makes a weakly significant positive contribution to the variation in Cl^- concentrations represented in the set of FEs (Table S3).

The non–time-varying portion of stream flow has no statistically significant impact on the portion of TSS concentrations explained by the FEs (Table S4). Relative to agricultural land (the excluded land use category), the percentage of both forested and urban land upstream in the watershed decreases the portion of TSS concentrations explained by the FEs.

We do not include these models in the main results, because they impart no additional information about the links between Cl^- and TSS concentrations in Pennsylvania surface water and shale gas development. We include them here to demonstrate that the main models do, in fact, control for important alternative sources of the contaminants we analyze, even though our comprehensive approach to reducing omitted variables bias through FE regressions means that the impacts of these specific controls cannot be interpreted from the main model results.

Waste Treatment Regulatory Changes During the Period of Analysis

The capacity of wastewater treatment plants to treat shale gas waste adequately has received regulatory attention. In August 2010, Pennsylvania announced newly instituted regulations for wastewater from shale gas operations. Treatment facilities built or expanded after August 21, 2010, or those planning to accept greater quantities of shale gas waste than they were currently accepting on that date were required to meet standards, beginning in January 2011, of 250 mg/L Cl^-, as well as 500 mg/L TDSs, 10 mg/L barium, and 10 mg/L strontium. The regulation also prevented future shipments to publicly owned treatment works (POTWs) unless the waste was pretreated at a centralized waste treatment (CWT) facility that met the new standards. However, 27 facilities were grandfathered under the old standard (allowing disposal of treated wastewater with TDS concentrations up to 2,000 mg/L). In April 2011, the PADEP asked all operators in the Marcellus Shale to stop shipping waste to the 14 grandfathered facilities still thought to be accepting waste, effective May 19, 2011.

According to the PADEP waste shipment data used for this analysis, in the period July 2011–December 2011, which begins after the voluntary request to operators on May 19, 2011, a total of 36 treatment facilities in Pennsylvania were accepting shale gas waste, 8 of which were formerly grandfathered facilities (including two POTWs) specifically named in the PADEP-EPA correspondence regarding the voluntary ban on shipments. Our analysis ends in December 2011. PADEP data for a later period, January 2012–July 2012, indicate no shipments to POTWs. However, shipments to CWT facilities continue.

Robustness Checks

We implement numerous robustness checks of the main results. First, we consider the possibility that our econometric approach suffers from a fundamental mismatch between typical variation in Cl^- and TSS concentrations (subdaily) and the observed variation in shale gas wells and waste treatment (daily for wells and semiannually or annually for waste shipments). In Table S5, column 1 reports results from the main Cl^- model in Table 1 (column 2) replacing daily Cl^- observations on the left-hand side with quarterly average Cl^- concentrations by monitor. Column 2 of Table S5 does the same, but for semiannual average Cl^- concentrations by monitor. The results are very similar to those in Table 1 (column 2): Wells have no significant impact on downstream average Cl^- concentration, and waste treatment increases downstream average Cl^- concentration at both levels of aggregation. The amount of variation in coarse Cl^- concentration averages explained by the models (as measured by R^2) is larger than for the daily models. However, by collapsing the rich intertemporal variation in well location from daily to quarterly or semiannual averages, these models perform a weaker test of the hypothesis that wells affect downstream Cl^- concentration than the models reported in Table 1 (although the variable of wells is insignificant in both, the interpretation of this negative result is much stronger in the main models because they exploit significant additional variation to obtain the result). Because the waste shipment data are reported only every 6 to 12 mo, the effect on the waste treatment variable estimate is negligible.

The same tests are implemented for TSS (Table S6, columns 1 and 2), where collapsing the intertemporal variation in well location has a much stronger impact, because this is the main "positive" result for TSS. Without exploiting this variation, we no longer estimate a significant impact of well pads on downstream TSS concentrations (in comparison to Table 2, column 3). Of 4,908 wells in the data, 4,400 were drilled in 2009–2011; when we average well counts even at the quarterly level for just these 3 y (1,095 daily observations become 12 quarterly observations), it is not surprising that the TSS effects for wells are no longer significant. As in the case of Cl^-, R^2 is higher for these aggregate models, but the coefficient estimates mask an important impact from wells by moving to a coarser temporal analysis. Because the main purpose of this analysis is to obtain unbiased estimates of the impacts of wells and waste treatment on downstream water quality, the models reported in Tables 1 and 2 are preferable to those in Tables S5 and S6 (columns 1 and 2).

A second pair of robustness checks considers the possibility that our estimates overweight water quality observations from non-Marcellus Shale areas in southern and southeastern Pennsylvania (Fig. 1). We reestimate the Table 1 models, using only the 3,489

Cl$^-$ concentration observations (42% of the sample in Table 1) from monitors overlying the Marcellus Formation. This omits any Cl$^-$ impacts at monitors outside of the region, but it ensures that we are making an "apples to apples" comparison of Cl$^-$ concentrations at monitors within the Marcellus Formation, but with differing proximity to wells and waste treatment facilities. Results are robust to this test (Table S5, column 3). We do the same for TSS, shrinking the sample size from almost 12,000 to 5,900 observations (Table S6, column 3). The point estimates from the main specification and from the Marcellus Formation-only subsample are very similar to those in Table 2, column 3.

Next, we test the robustness of the models to a more stringent assumption about which waste treatment facilities are releasing treated shale gas waste to surface water, rather than exclusively recycling waste for reuse by shale gas operators. Waste facility location data and other descriptive information for treatment facilities came from multiple sources and are not complete for all facilities. To ensure that the waste treatment effects we estimate are identified only off of variation in shipments of shale gas waste to NPDES-permitted facilities releasing treated waste to surface water, we reestimate the main Cl$^-$ and TSS models, using only waste shipped to facilities for which we have full information, including a valid NPDES permit number. The results for Cl$^-$ (Table S5, column 4) and TSS (Table S6, column 4) are robust to this test.

Finally, given the temporal coarseness of the waste shipment reporting described above, we also test the robustness of our results to different assumptions about the dates within annual and biannual reporting periods on which waste shipments from wells to particular treatment facilities commenced and stopped. Results are robust to two alternative assumptions about the commencement of shipments: that they begin on the period's first day (Tables S5 and S6, column 5) and on its last day (Tables S5 and S6, column 6). Results are also robust to assuming that the facility stops accepting waste at the end of the period, rather than at the midpoint. The waste shipment results for both Cl$^-$ and TSS are also robust to excluding all shipments in 2007, which has missing waste shipment data. Results from each of these last two tests are available from the authors on request.

1. Entrekin S, Evans-White M, Johnson B, Hagenbuch E (2011) Rapid expansion of natural gas development poses a threat to surface waters. *Front Ecol Environ* 9(9):503–511.
2. Johnson DB, Hallberg KB (2005) Acid mine drainage remediation options: A review. *Sci Total Environ* 338(1-2):3–14.
3. Slutz J, Anderson J, Broderick R, Horner P (2012) Key shale gas water management strategies: An economic assessment tool. SPE/APPEA International Conference on Health, Safety, and Environment in Oil and Gas Exploration and Production, Perth, Australia, September 11–13, 2012. Available at: http://spe.org/atce/2012/pages/schedule/technical_program/documents/spe157532-page1.pdf. Accessed February 21, 2013.
4. Jin L, Whitehead P, Siegel DI, Findlay S (2011) Salting our landscape: An integrated catchment model using readily accessible data to assess emerging road salt contamination to streams. *Environ Pollut* 159(5):1257–1265.
5. Shaw S, Marjerison R, Bouldin D, Parlange J-Y, Todd W (2012) Simple model of changes in stream chloride levels attributable to road salt applications. *J Environ Eng (New York)* 138(1):112–118.
6. Trowbridge PR, Kahl JS, Sassan DA, Heath DL, Walsh EM (2010) Relating road salt to exceedances of the water quality standard for chloride in New Hampshire streams. *Environ Sci Technol* 44(13):4903–4909.

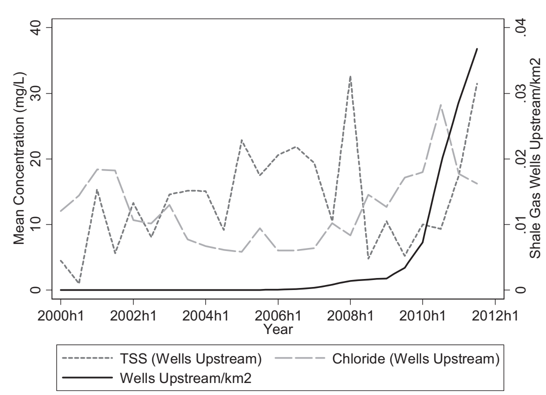

Fig. S1. Average contaminant concentrations and density of upstream shale gas wells for monitors with at least one upstream well in their watershed (2000–2011).

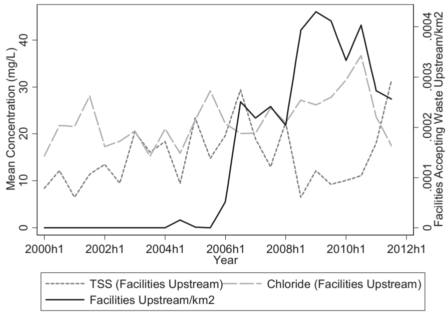

Fig. S2. Average contaminant concentrations and density of upstream treatment facilities accepting shale gas waste for monitors with at least one accepting facility in their watershed (2000–2011).

Table S1. Summary statistics for Cl^- models

Variable	Mean	SD	Minimum	Maximum
Concentration, mg/L	19.1	37.7	0	1,065
Cumulative precipitation (4 d), mm	171	218	0	3,353
Gas wells upstream/km^2	0.0076	0.0326	0	0.365
Gas wells upstream (0–90 d)/km^2	0.0012	0.0054	0	0.0798
Gas wells upstream (90–180 d)/km^2	0.0011	0.0051	0	0.0798
Facilities accepting waste upstream/km^2	0.0002	0.0009	0	0.0152
Waste quantity treated upstream, MMbbl/km^2	0.0002	0.0020	0	0.033
Affected facilities accepting waste/km^2	0.0001	0.0006	0	0.0152
Nonaffected facilities accepting waste/km^2	0.0001	0.0005	0	0.0152
NPDES upstream/km^2	0.0033	0.0041	0	0.049
Roads (lanes × km)/km^2	1.32	0.749	0.056	4.68
Modeled stream flow, ft^3/s	80.7	145	0.462	1,119
Area of watershed upstream, km^2	1,516	1,330	0.465	5,965
Facilities accepting waste upstream	0.128	0.431	0	3
Waste quantity treated upstream, MMbbl	0.15	1.02	0	12
Annual no. of waste shipments upstream/km^2	0.028	0.168	0	3.32
Gas wells upstream	14.3	82.5	0	1,199
Gas wells upstream (0–90 d)	2.07	11.8	0	155
Gas wells upstream (90–180 d)	1.95	11.6	0	153
NPDES upstream	4.72	5.23	0	25
Above Marcellus Shale (indicator)	0.417	0.493	0	1

There were 8,364 observations, 860 monitors, and an average of 98 d between observations. The sample is all Pennsylvania surface water quality monitors in the STORET database measuring Cl^- concentration, 2000–2011. MMbbl, million barrels.

Table S2. Summary statistics for TSS models

Variable	Mean	SD	Minimum	Maximum
Concentration, mg/L	20.4	81.2	0	2,862
Cumulative precipitation (4 d), mm	179	233	0	3,353
Facilities accepting waste upstream/km^2	0.0001	0.0007	0	0.0141
Gas wells upstream/km^2	0.0041	0.0232	0	0.365
Well pads upstream/km^2	0.0022	0.0107	0	0.135
Well pads permitted, prespud/km^2	0.0002	0.0010	0	0.0145
Well pads upstream/km^2 × precipitation (4d)	0.356	2.93	0	109
Modeled stream flow, ft^3/s	140	193	0.462	1,119
% Urban upstream in watershed	6.64	8.6	0.184	36.4
% Forest upstream in watershed	62.1	19.1	17.8	97.2
Area of watershed upstream, km^2	1,698	1,367	0.465	5,965
Facilities accepting waste upstream	0.0986	0.387	0	3
Gas wells upstream	8.16	62.6	0	1,199
Well pads upstream	4.25	28.9	0	487
Well pads permitted, prespud	0.384	1.84	0	33
Above Marcellus Shale (indicator)	0.495	0.5	0	1

There were 11,919 observations, 644 monitors, and an average of 55 d between observations. The sample is all Pennsylvania surface water quality monitors in the STORET database measuring TSS, 2000–2011.

Table S3. Regressions of Cl^- model water quality monitor FEs on descriptive variables

Variable	(1)	(2)	(3)	(4)	(5)	(6)
NPDES upstream/km^2	2,059.173*	2,107.199*	2,091.231*	2,104.504*	2,093.871*	2,121.993*
	(1,076.917)	(1,117.263)	(1,110.31)	(1,120.117)	(1,120.125)	(1,120.116)
Roads (lanes × km)/km^2	7.776**	7.793**	7.764**	7.778**	7.756**	7.810**
	(3.111)	(3.081)	(3.087)	(3.081)	(3.082)	(3.082)
Modeled stream flow, ft^3/s	0.011	0.010	0.010	0.010	0.011	0.010
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
N	5,809	5,809	5,809	5,809	5,809	5,809
Mean $\widehat{monitor}_i + \widehat{wm}_{it} + \widehat{ym}_t$	17.788	17.511	17.591	17.588	17.441	17.441

Table S4. Regressions of TSS model water quality monitor FEs on descriptive variables

Variable	(1)	(2)	(3)	(4)	(5)
Modeled stream flow, ft^3/s	−0.002 (0.008)	−0.002 (0.007)	−0.002 (0.007)	−0.002 (0.007)	−0.002 (0.008)
% Urban upstream	−0.800** (0.373)	−0.793** (0.335)	−0.794** (0.328)	−0.796** (0.357)	−0.795** (0.377)
% Forest upstream	−0.622*** (0.188)	−0.626*** (0.162)	−0.626*** (0.160)	−0.626*** (0.180)	−0.625*** (0.182)
N	7,676	7,676	7,676	7,676	7,676
Mean $\widehat{monitor}_i + \widehat{wm}_{it} + \widehat{ym}_t$	4.036	3.869	3.873	4.019	3.885

Table S5. Robustness checks for Cl^- models

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Total precipitation (4 d), mm	−0.006* (0.003)	−0.010 (0.007)	−0.005** (0.002)	−0.003*** (0.001)	−0.003*** (0.001)	−0.003*** (0.001)
Gas wells upstream/km^2	27.805 (26.767)	208.430 (325.141)	23.411 (17.844)	23.624 (19.486)	25.633 (20.88)	19.392 (17.235)
Facilities accepting waste upstream/km^2	1,674.332** (661.356)	2,471.092*** (552.899)	2,046.894*** (681.930)	1,576.075*** (219.387)	1,485.522*** (371.52)	1,947.051*** (458.089)
N	4,246	2,745	3,489	8,364	8,364	8,364
Mean Cl^-, mg/L	18.959	19.456	16.982	19.074	19.074	19.074
R^2	0.589	0.859	0.711	0.499	0.499	0.499

Each column reestimates column 2 from Table 1, with the following changes. In column 1, the dependent variable is average Cl^- concentration at a monitor by quarter (four observations per year), and we include FEs for monitors, year-quarters, and watershed-quarters. In column 2, the dependent variable is average Cl^- concentration at a monitor by half-year (two observations per year), and we include FEs for monitors, year-half-years, and watershed-half-years. In column 3, the sample includes only daily Cl^- observations from monitors overlying the Marcellus Formation. In columns 4–6, we use the full sample of daily Cl^- observations from Table 1 but change our original waste shipment specification as follows: Column 4 includes only waste shipments to facilities with known NPDES permit numbers; column 5 assumes shipments begin at the start of each waste shipment reporting period, rather than the midpoint; and column 6 assumes shipments begin at the end of each waste shipment reporting period, rather than the midpoint. Columns 3–6 include FEs for monitor, year-month, and watershed-calendar month. Variables divided by square kilometers are divided by the area of the watershed that is upstream of the monitor. Reported SEs are robust and clustered by watershed. Statistically significant at the *10% level; ** 5% level; ***1% level.

Table S6. Robustness checks for TSS models

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Total precipitation (4 d), mm	0.054*** (0.015)	0.030** (0.013)	0.089** (0.034)	0.086*** (0.025)	0.086*** (0.025)	0.086*** (0.025)
Facilities accepting waste upstream/km^2	240.643 (437.578)	−262.589 (233.433)	490.287 (418.292)	650.042 (398.715)	680.116 (509.192)	576.04 (430.797)
Well pads upstream/km^2	23.886 (45.283)	37.77 (69.324)	98.098* (50.907)	98.829* (54.183)	98.143* (54.083)	97.594* (54.418)
N	5,913	3,463	5,900	11,919	11,919	11,919
Mean TSS, mg/L	17.848	17.513	16.859	20.392	20.392	20.392
R^2	0.519	0.829	0.405	0.283	0.283	0.283

Each column reestimates column 3 from Table 2, with the following changes. In column 1, the dependent variable is average TSS concentration at a monitor by quarter (four observations per year), and we include FEs for monitors, year-quarters, and watershed-quarters. In column 2, the dependent variable is average TSS concentration at a monitor by half-year (two observations per year), and we include FEs for monitors, year-half-years, and watershed-half-years. In column 3, the sample includes only daily TSS observations from monitors overlying the Marcellus Formation. In columns 4–6, we use the full sample of daily TSS observations from Table 2 but change our original waste shipment specification as follows: Column 4 includes only waste shipments to facilities with known NPDES permit numbers; column 5 assumes shipments begin at the start of each waste shipment reporting period, rather than the midpoint; and column 6 assumes shipments begin at the end of each waste shipment reporting period, rather than the midpoint. Columns 3–6 include FEs for monitor, year-month, and watershed-calendar month. Variables divided by square kilometers are divided by the area of the watershed that is upstream of the monitor. Reported SEs are robust and clustered by watershed. Statistically significant at the *10% level; ** 5% level; ***1% level.