The Effect of Reporting Thresholds on the Validity of TRI Data as Measures of Environmental Performance: Evidence from Massachusetts

By:
Lori Snyder Bennear*

Abstract: This paper examines the effect of reporting thresholds on the use of TRI data for both time-series and cross-sectional policy analysis. Facilities are only required to report *releases* to TRI if their *use* of a chemical exceeds some threshold. This creates an incentive for facilities to reduce their use of a listed chemical to a level just below the reporting threshold. However, this does not necessarily represent a real improvement in environmental performance, as the facility's release level may remain largely unchanged. The State of Massachusetts has expanded the disclosure requirements under TRI in ways that better allow for assessment of the measurement error induced by reporting thresholds. This paper utilizes data from the Massachusetts Toxics Use Reduction Act (TURA) to estimate bounds on the degree of bias introduced by the reporting thresholds in both time series and cross-sectional analyses for Massachusetts' facilities. In particular, this paper asks the questions: (1) How much of the observed decline in reported releases is artificially created by the existence of the reporting thresholds?; and (2) How much would our rankings of facilities change if we could account for releases by facilities that are below the reporting threshold?

The results of the analysis suggest that the bias introduced by the reporting thresholds may be significant. Up to 40 percent of the observed decline in TRI releases in Massachusetts may be attributed to truncation at the reporting thresholds. In addition, quartile rankings of facilities based on reported releases may be in error 45 percent of the time when behavior around the reporting thresholds is not taken into account. Because the TURA data are in many ways identical to the TRI data, the fact that truncation bias is a large concern in the TURA data suggests that this bias may be important for the national TRI data as well.

---

* Bennear is an Assistant Professor of Environmental Economics and Policy at the Nicholas School of the Environment and Earth Sciences, Duke University. She can be reached at lori.bennear@duke.edu or at 919-613-8083. The author is grateful to Cary Coglianese, Nolan H. Miller , Cody Rice, Robert N. Stavins and participants in the Harvard Environmental Economics Seminar and the Environmental Protection Agency's Corporate Environmental Behavior and the Effectiveness of Government Interventions Workshop.

The Effect of Reporting Thresholds on the Validity of TRI Data as Measures of Environmental Performance: Evidence from Massachusetts

1 Introduction

The old adage, "You can't manage what you don't measure," is the primary rationale for conducting systematic evaluations of the effectiveness of various environmental policy initiatives. Only if governments, non-profits, industries, and communities have good measures of environmental outcomes (changes in pollution levels, risk levels, etc.) can they evaluate what policies work, how well they work, and how to improve their effectiveness. If the metrics are not valid, then neither are the policy inferences drawn from these metrics. This paper examines one aspect of the validity of a frequently used measure of environmental performance—pollution releases reported under the federal Toxics Release Inventory (TRI) program.

The TRI data are the most comprehensive data available on facility-level releases of toxic chemicals. Facilities are required to disclose publicly their releases to all media—air, water, land, and underground injection (on-site releases) —as well as their transfers of chemicals off-site for recycling or disposal (off-site releases), for over 600 toxic chemicals. Because the TRI data capture releases to all media and include measures of the environmental impact of a facility's activities beyond the facility's fence (off-site releases), many have argued that the TRI data provide a more complete picture of facility-level environmental performance than other available metrics (Karkkainen, 2001).

In fact, the TRI data are used frequently for the purposes of comparing environmental performance across geographic areas and over time by government agencies, non-profits, and academic researchers. Environmental Defense uses the TRI data as one of the indicators of environmental performance in its Scorecard, an online database that allows the public to compare counties on a number of environmental metrics.[1] The Public Interest Research Group (PIRG) and its state affiliates frequently compile lists of the worst polluters in a state or region based on releases reported to the Toxics Release Inventory.[2] Similarly, EPA ranks states and industries by their total releases as reported in the TRI (U.S. EPA, 2002a). State environmental agencies also publish annual progress reports that measure changes in environmental performance based on changes in releases to the TRI, and label particular industries and facilities as the top polluters based on these releases.[3] Academic researchers have used the TRI data as outcome variables in evaluations designed to determine what factors affect environmental performance (Arora and Cason 1998; Grant and Jones 2003; Helland and Whitford 2003; Khanna and Anton 2002; King and Lenox 2000). Researchers have also used TRI data to evaluate whether requiring facilities to publicly disclose pollution leads facilities to decrease pollution. Fung and O'Rourke (2000) and Wolf (1996) argue that the observed 45 percent decline in overall TRI releases from 1988 to 1995 indicates that information disclosure is a valuable regulatory tool for reducing pollution.[4]

---

[1] The Environmental Defense scorecard can be accessed at www.scorecard.org.

[2] See, for example, U.S. Public Interest Research Group 1998.

[3] A full list of state TRI programs with links to the state annual reports can be found at www.epa.gov/tri/programs/state_programs.htm.

[4] See U.S. EPA 2003 for a list of detailed discussion of how TRI data have been used by government, business, academics, and citizen groups.

Despite the frequent use of TRI data for policy analyses, there are several known concerns about the validity of these data as measures of environmental performance. This paper defines the characteristics of a valid measure of environmental performance and outlines several known threats to the validity of the TRI data. The paper then focuses on estimating the magnitude of the measurement error created by the existence of arbitrary reporting thresholds. The potential for measurement error exists because facilities are only required to report *releases* to TRI if their *use* of a chemical exceeds some threshold. This creates an incentive for facilities to reduce their use of a listed chemical to a level just below the reporting threshold. However, this does not necessarily represent a real improvement in environmental performance, as the facility's release level may remain largely unchanged (or potentially could even increase). As a result, observed decreases in reported releases might overstate the true change in environmental performance. This paper asks the question: How much of any observed decline in reported releases could be artificially created by the existence of the reporting thresholds?

The TRI data are also used to rank facilities based on their pollution levels. Truncation at the reporting threshold may also have an effect of the validity of these cross-sectional rankings. This paper also asks the question: How much would our rankings of facilities change if we could account for releases by facilities that are below the reporting threshold?

While the potential for bias introduced by reporting thresholds is well known, there has previously been little that users of the data could do to ascertain the magnitude of this bias. The TRI data provide scant information that would allow a user of the data to ascertain whether a facility that ceases reporting did so because it went below the reporting threshold. However, the State of Massachusetts has expanded the disclosure requirements under TRI in ways that better allow for assessment of the reasons facilities cease reporting. This paper utilizes data from the Massachusetts Toxics Use Reduction Act (TURA) to estimate, for facilities in Massachusetts, bounds on the degree of bias introduced by the reporting thresholds in both time series and cross-sectional analyses.

The TURA data are similar to the TRI data. Indeed for Massachusetts facilities, data reported to TRI is a subset of the data reported to TURA. However, the TURA data include two additional features missing from the national data that make possible estimation of bounds on the truncation bias. First, the TURA reporting forms contain an optional question on why facilities are no longer reporting a chemical they had previously reported. Second, the TURA program requires facilities to report their chemical *use* in addition to their chemical *releases.* These two features of the data for Massachusetts allow estimation of the number of facilities that cease to report because they reduced use below the reporting threshold, but still use the chemical in positive quantities. These data are also used to estimate bounds on the amount of "missing" releases that result. Because the TURA data are in other ways identical to the TRI data, analysis of the TURA data provides some preliminary evidence on whether truncation bias is likely to be a large or small problem for the national TRI data.

The paper begins in Section 2.2, by describing the TRI and TURA data. Section 2.3 articulates a specific definition of "validity" of an environmental performance metric and highlights several threats to the validity of the TRI/TURA data as measures of environmental performance under this definition. This section defines truncation bias at the reporting thresholds and details how this bias may invalidate both time-series and cross-sectional comparisons. Section 2.4 estimates the magnitude of the truncation bias created by the existence of the reporting thresholds using the TURA data. The results suggest that truncation bias is indeed a serious threat to the validity of these data as measures of environmental performance, particularly in cross-sectional comparisons. Time-series estimates are off by roughly 40 percent in Massachusetts in the worst-case scenario. That is, 40 percent of the observed decrease in releases in Massachusetts may be artificial declines created by the reporting thresholds. For cross-sectional comparisons the results for Massachusetts indicate that quartile rankings of facilities may be wrong as often as 50 percent of the time when truncation bias is not accounted for. Section 2.4 ends with a discussion of the implications of the Massachusetts findings for the use of nationwide TRI data. Given the potential importance of these effects for policy analysis, Section 2.5 presents suggestions for adjusting policy analysis to account for truncation bias.

2 The U.S. Toxics Release Inventory and the Massachusetts Toxics Use Reduction Act

In December 1984, a Union Carbide pesticide plant in Bhopal, India accidentally released 40 metric tons of methyl isocyanate, killing an estimated 2,000 people and injuring 100,000 others. Shortly thereafter, a pesticide release in West Virginia hospitalized 150 people. Partly in response to concerns raised by these high-profile accidental releases, Congress passed the Emergency Planning and Community Right to Know Act (EPCRA) in 1986. EPCRA requires manufacturing firms to report releases of specific toxic chemicals on an annual basis, and to make these reports available to the public. The U.S. Environmental Protection Agency's (EPA) implementation of EPCRA resulted in the creation of the Toxic Release Inventory (TRI), which requires large manufacturing facilities publicly disclose total releases of listed toxic chemicals to all media on an annual basis.

There are three factors that determine whether a facility is subject to the disclosure requirements. The first is industrial sector. Originally, only manufacturing facilities were subject to the TRI reporting requirements. Subsequently, several other industrial sectors were added including: federal facilities, metal and coal mining, electrical generating facilities that combust coal or oil, chemical wholesale distributors, petroleum terminals and bulk storage facilities, and hazardous waste treatment storage and disposal facilities. The final two determinants of regulatory eligibility concern facility size. Only facilities with 10 or more full-time equivalent employees are required to report pollution levels. In addition, facilities are only required to disclose releases of listed chemicals for which they either: (1) manufacture or process more than 25,000 pounds or (2) otherwise use more than 10,000 pounds. Figure 2-1 provides a flow chart of TRI reporting requirements.

Several states have subsequently required additional public disclosure for plants in their jurisdictions. In 1989, Massachusetts passed the Toxics Use Reduction Act which expanded the reporting requirements for facilities in that state (Massachusetts Department of Environmental Protection 2003). There are three main differences between the TURA and the TRI requirements. First, the list of chemicals for which facilities are required to report is larger in Massachusetts. Massachusetts facilities must report releases of all chemicals required by TRI and also any chemical listed under the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA), commonly known as Superfund.

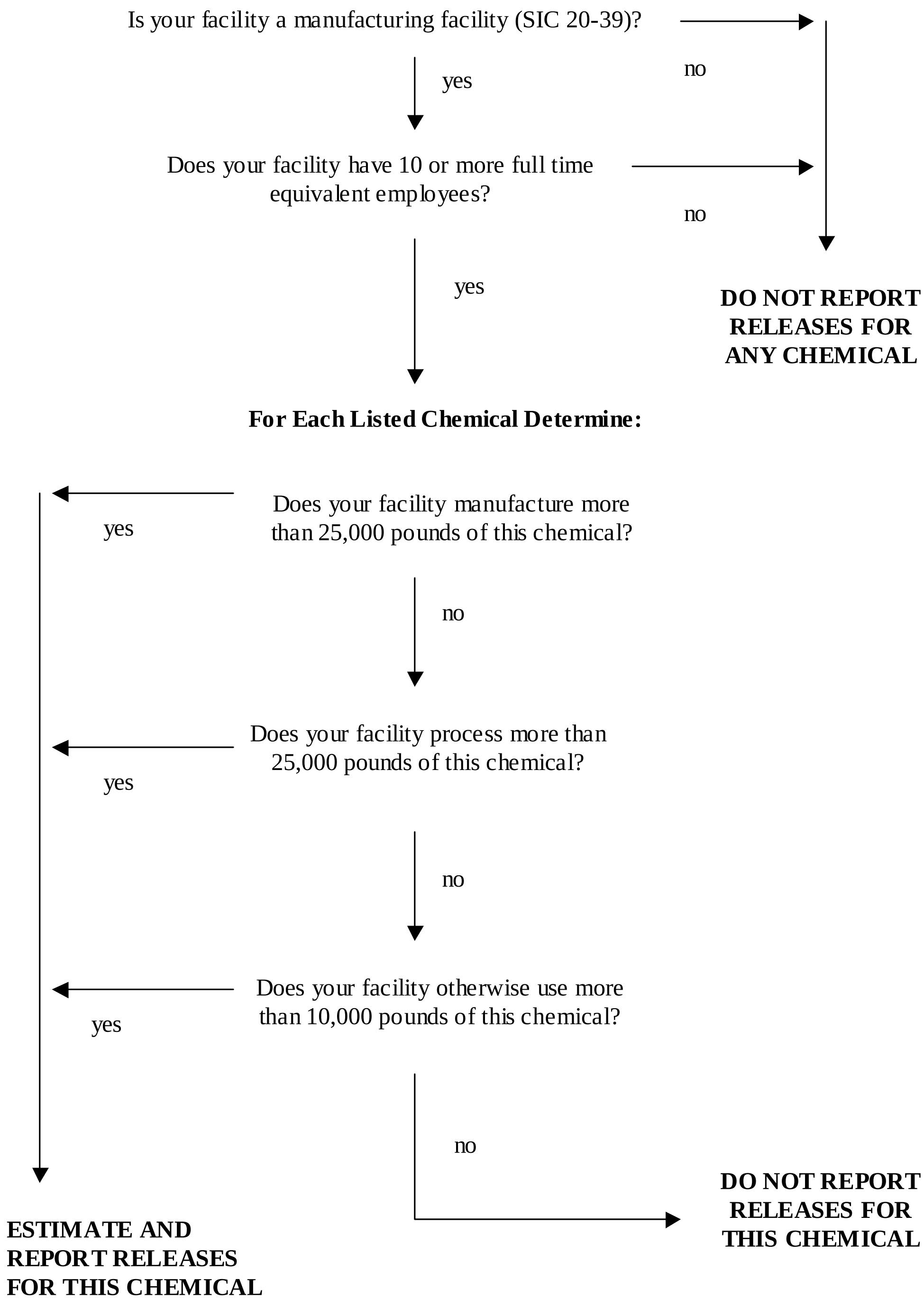

Figure 2-1: Flow Chart of TRI Reporting Requirements

The second difference is that the reporting thresholds are weakly lower in Massachusetts. If a facility triggers the federal reporting threshold for at least one listed chemical (i.e., manufacturers or processes more than 25,000 pounds or otherwise uses more than 10,000 pounds), the facility must report for all listed chemicals for which total manufacture, process, and use is greater than 10,000 pounds. Therefore, for most facilities in Massachusetts there are not three separate reporting thresholds, but one binding threshold at 10,000 pounds of total use for each listed chemical.

The final difference is that facilities in Massachusetts are required to report chemical *use* in addition to chemical *release*. The reporting thresholds are based on chemical use, but the federal program does not require public disclosure of use levels. In Massachusetts facilities must report both total use of the chemical and total releases of the chemical. Figure 2-2 provides a flow chart for the TURA reporting requirements.

This paper is concerned about the effects of the reporting thresholds present in both the TRI and the TURA designs, on the valid uses of these data for policy analysis. Before specifically examining the TRI and TURA data and determining whether or not these data are valid measures of environmental performance, it is worthwhile to articulate a definition of valid measurement of environmental performance. That is the subject of the next section.

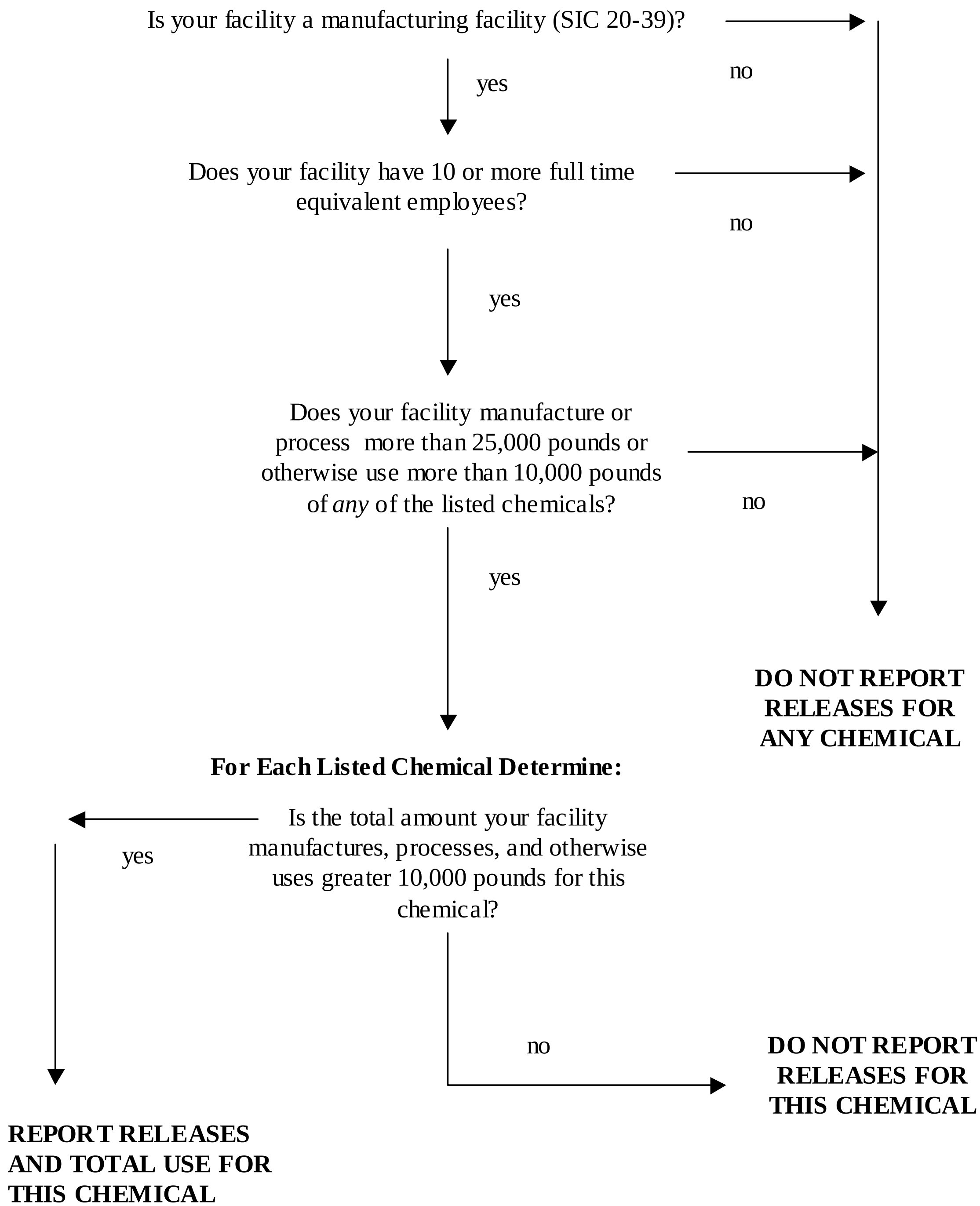

Figure 2-2: Flow Chart of TURA Reporting Requirements

3 The TRI/TURA Data as Measures of Environmental Performance

3.1 Defining a valid measure of environmental performance

There are a number of potential characteristics of a valid measure of environmental performance (National Academy of Engineering, 1999; White and Zinkl, 1997; U.S. EPA, 1992). The most frequently discussed characteristics of a valid performance metric are:

1. Significance: The measure has a clear relationship to environmental performance.
2. Precision: The stochastic component of the measure is small relative to the deterministic component.
3. Verifiability: The measure is publicly available and can be validated by a third party.
4. Comparability: The measure allows for comparisons across facilities, industries, countries, and across time.

Focusing on the last of these, comparability, there are potentially two standards of comparability that could be required--cardinal comparability or ordinal comparability. Cardinal comparability implies that the measure can be compared in magnitude across different entities. Define releases by facility *i* to be X_i. If X is cardinally comparable, $X_1 > X_2$ implies that facility 1 is doing better than facility 2 and is, in fact, doing precisely $X_1 - X_2$ better.[5]

[5] Cardinal comparability is often present in outcome measures used to evaluate program effectiveness by many other federal and state agencies such as the Department of Health and Human Services and the Labor Department. If participants in a pilot job training program earn an average of

A weaker comparability standard would be ordinal comparability. Ordinal comparability implies that if $X_1 > X_2$ then facility 1 is doing better than facility 2. However, it is not the case that the magnitude of the difference in performance is captured by the magnitude of the difference in the performance metric, X. Ordinal comparability is the minimum comparability standard for a valid performance metric. The whole point of program or policy evaluation is to ascertain whether a program or policy made society better off. If the metric used cannot be relied upon to order outcomes from better to worse, then the metric cannot be relied upon to measure whether a program resulted in a better outcome. Cardinal comparability is not necessary if the goal of the evaluation is to determine whether a specific initiative improved outcomes. However, cardinal comparability is necessary if the goal of the evaluation is to determine how much of an improvement was obtained by the firm, industry, state, or policy initiative. For example, cardinal comparability is necessary for cost-effectiveness analyses that compare policies on the basis of the amount of risk reduction per dollar spent.

The remainder of this paper evaluates the Toxics Release Inventory data as measures of plant-level environmental performance in relationship to these four characteristics of validity. In particular the paper highlights the potential threat to validity presented by the existence of minimum reporting thresholds and attempts to quantify the degree of the bias created by these thresholds.

3.2 Validity of TRI Data as a Measure of Environmental Performance

There are several known threats to the validity of the TRI data. The first is that data are only collected for certain industries and chemicals. Originally data were collected for

---

$25,000 per year and non-participants earns an average of $20,000 per year, then trainees are better off than non-trainees by exactly $5,000 per year, on average.

manufacturing facilities and for approximately 300 chemicals. The fact that the TRI data are not comprehensive may reduce the *significance* of these data as a measure of environmental performance. For the measure to be significant in needs to have a clear relationship to environmental performance. If the TRI data only capture releases of a small subset of total chemicals, the significance of these data may be compromised.

This concern was recognized early in the program (U.S. Government Accounting Office 1991) and the program has since been expanded to more than double the number of chemicals and to include federal facilities, metal and coal mining facilities, electric generating facilities that combust coal or oil, chemical wholesale distributors, petroleum terminals and storage facilities, and hazardous waste treatment storage and disposal facilities. Obviously the TRI data cannot be used as a valid measure of environmental performance for plants that are not obligated to report these data. Thus, these data will necessarily by limited to use only in evaluations that affect these industrial sectors.

A second threat to the validity of the TRI data as a measure of environmental performance stems from the fact that TRI data are reported in total pounds of a chemical released, but it is widely acknowledged that pounds of different chemicals present widely different levels of risk. Similarly, a pound of a chemical released in rural Oklahoma may have a different impact than a pound of the same chemical released in downtown Manhattan. Some of the factors that determine the overall environmental risk presented by a facility's activities include the total pounds of releases, the toxicity of these releases, the exposure level in the population, the duration of exposure, and the sensitivity of the population. Simply adding up total pounds of different chemicals released and comparing across facilities may provide an incorrect ordering of facilities according to relative risk. For example, a facility that releases a few pounds of a very toxic, persistent, and bio-accumulating toxin may appear to have much better "environmental performance" than a facility that releases more pounds of a less hazardous and more degradable substance.

Even if the point estimate of environmental performance provided by TRI is biased by lack of information on toxicity, one might hope that *changes* in TRI releases still provide a measure of *changes* in environmental performance. That is, perhaps time series comparability is maintained even if cross-sectional comparability is compromised. For example, even if pounds of different chemicals are not equal in their risk profile, one might argue that fewer pounds is better than more pounds, holding all else constant. But the key is that all else must be held constant. The assumption that the trend in TRI releases is a valid indicator of the trend in environmental performance is likely to hold if, among other things, the facility consistently uses and reports on the same chemicals in every year. One change in reporting that may threaten the validity of changes in TRI releases as measures of changes in environmental performance is if a facility substitutes from a less toxic to a more toxic chemical (or vice versa). Even if the facility reduces releases by a substantial amount, the total risk presented by the facility may remain the same, decrease by less, or even increase. Thus, not adjusting for toxicity compromises the *significance* and the *comparability* of the TRI data both across facilities in a given year and within a facility over time.

EPA and academic researchers have been working to develop indicators that take the TRI data as an input, combine it with toxicity, exposure, and degradability information, and then calculate a risk-based metric. EPA has recently released its Risk Screening Environmental Indicators (RSEI) model, which combines TRI data and risk information to develop a risk-adjusted pollution measure for peer review (U.S. EPA, 2004). These changes will doubtless improve the quality of estimates of environmental performance using TRI data.

A third threat to the validity of the TRI data stems from the fact that these data are self-reported by facilities. This affects both the *precision* and the *verifiability* of the data as measures of environmental performance. Changes in how a facility estimates releases or even how it classifies releases may lead to changes in reported releases that are not reflective of real changes in environmental performance (Graham and Miller 2001). In the initial years following the establishment of TRI, the accuracy of the self-reported releases was thought to be poor (U.S. Government Accounting Office 1991); however, most observers believe it has improved over time. EPA has issued updated guidance on estimating releases both for different industries and chemicals,[6] and it audits a small number of facilities' TRI reports each year. Despite these efforts, concerns over the precision and verifiability of the self-reported data remain.

The final threat to the validity of the TRI data is due to truncation in reported releases caused by the existence of reporting thresholds. It is this threat that this paper addresses. Reporting thresholds set a minimum level of chemical use at a facility, below which the facility is not required to report releases. For the TRI data, the reporting thresholds are 25,000 pounds for manufacture and processing of a listed chemical and 10,000 pounds for any other use of the chemical. Similarly, the TURA data require reporting only if total use (manufacturing plus processing plus otherwise use) is greater than 10,000 pounds. These reporting thresholds create incidental truncation in the TRI data. This truncation threatens the *significance* and *comparability* of the TRI data for both ordinal and cardinal analyses. To illustrate the potential bias, define:

y_{ict} = actual releases of chemical c, by facility i, in time t

[6] For example, U.S. EPA (2000) provides detailed guidance for the Textile Processing Industry. The set of all industry and chemical guidance can be obtained online at www.epa.gov/tri/guided_docs.

M_{ict} = amount of chemical c manufactured by facility i in time t

P_{ict} = amount of chemical c processed by facility i in time t

U_{ict} = amount of chemical c otherwise used by facility i in time t

ς = set of all chemicals, *c*, that are reportable to TRI

For each facility there exists a true aggregate measure of releases, Y, such that

$$Y_{it} = \sum_{c \in \varsigma} y_{ict} .$$

However, we do not observe y_{ict} or Y_{it}. Instead, we observe:

$$\tilde{Y}_{it} = \sum_{c \in \varsigma} y_{ict} r_{ict} ,$$

where

$$\begin{aligned} r_{ict} &= 1 \text{ if } (M_{ict} > 25{,}000) \text{ or } (P_{ict} > 25{,}000) \text{ or } (U_{ict} > 10{,}000) \\ &= 0 \text{ otherwise} \end{aligned} .$$

Further define:

r_{it} = vector of r_{ict} ($r_{i1t}, r_{i2t}, \ldots, r_{iCt}$) for all c in ς ,

$$\begin{aligned} u_{ict} &= 1 \text{ if } (M_{ict} > 0) \text{ or } (P_{ict} > 0) \text{ or } (U_{ict} > 0) \\ &= 0 \text{ otherwise} \end{aligned}$$, and

u_{it} = vector of u_{ict} ($u_{i1t}, u_{i2t}, \ldots, u_{iCt}$) for all c in ς .

Under what circumstances can we use observed data on reported releases, $\tilde{Y}_{it}$, to make valid cross-section or time series comparisons? Begin with a cross-sectional comparison. If we observe: $\tilde{Y}_{1t} > \tilde{Y}_{2t}$, under what circumstances can we be confident that

$Y_{1t} > Y_{2t}$? This inference will be valid if both facilities report for all chemicals for which they use any positive quantity.[7] That is if:

$$r_{1t} = u_{1t} \text{ and } r_{2t} = u_{2t}.$$

The key point is that if there are some chemicals which a facility uses in positive quantities, that is $u_{ict}>0$, but for which the facility is not required to report because u_{ict} is below the reporting threshold, then $\tilde{Y}_{it}$ does not necessarily equal Y_{it} and $\tilde{Y}_{it}$ is not necessarily ordinally comparable across facilities.[8]

Turning to time-series evaluation, under what circumstances do the trends in reported releases provide valid information on the trends in actual releases? That is, under what circumstances does $\tilde{Y}_{it} - \tilde{Y}_{it-1} = Y_{it} - Y_{it-1}$? As with the cross-sectional comparison, time series comparisons based on reported releases will only provide valid inferences on true releases if the facility reports for every chemical that it uses in any positive quantity.[9] That is, if:

$$r_{1t} = u_{1t} \text{ and } r_{1t-1} = u_{1t-1}$$

Notice that the facility does not need to report for the same set of chemicals in both years. It just must report for all chemicals it uses in each year. If a facility stops using a chemical and, hence, stops releasing this chemical, that is a real change in environmental

[7] This is the only condition under which this inference is generally valid. However, this inference may be valid in certain special cases. For example, if every facility does not report for a chemical, c, but releases of that chemical are the same for all facilities. There is no reason to think that this will hold.

[8] If ordinal comparability is not preserved, then cardinal comparability is also not preserved.

[9] This condition allows for valid time series inference in all cases. There may be special cases in which require less restrictive conditions. For example, if the amount of non-reported releases is constant across time, then the absence of these releases does not bias the time series comparison. However, there is no *a priori* reason for assuming non-reported releases are constant across time.

performance. But if a facility simply uses less of the chemical and is no longer required to report releases, this is not necessarily a real change in environmental performance.

The bias that results from truncation at the reporting threshold is referred to as truncation bias. The term truncation bias here has a slightly different meaning than the classic econometric definition. The truncation bias in this paper most closely resembles incidental truncation bias, which arises when facilities or individuals are observed on the basis of the outcome of another decision (Wooldridge, 2002: 552). For example, imagine one only observes wages for the employed, so observing wages is the result of another decision, in this case the labor market participation decision. The truncation bias in the TRI data is incidental truncation bias at the unit for which the data are collected, which is the facility-chemical-year level. One only observes a facility-chemical-year record if the facility triggers the reporting threshold for that chemical in that year. So observing data is the result of the chemical use decision. This incidental truncation bias is further aggravated by aggregation to other levels of analysis, such as the facility, firm, industry, or state level. The aggregation essentially treats all unobserved data as zero, which further invalidates comparisons. Perhaps a more specific name for this bias is "truncation and aggregation bias," but it will be referred to here as truncation bias for short.

While the first three threats to the validity of TRI data are well recognized and measures have been taken to address the threats, little has been done to assess or reduce the problems associated with truncation bias. This lack of attention is not the result of a lack of understanding of the problem, rather it is largely due to the fact that data have not been available that would allow analysts to estimate the extent of the problem and correct it. In contrast, the bias introduced by not weighting TRI pounds by some measure of toxicity is conceptually easier to address because risk factors can be determined with available data and the TRI data can be corrected using these risk factors. However, since facilities are generally not required to report chemical use levels, one cannot observe whether a facility is reporting for all chemicals it uses in positive quantities. Thus, the truncation bias cannot be corrected for in a systematic way using available data.

While systematic correction of the truncation bias cannot be obtained, this paper focuses on estimating bounds on the degree of bias presented by truncation of the data at the reporting thresholds. A reasonable question might be: How much of the observed decrease in TRI releases is potentially due to truncation bias? Similarly, for a cross-section one can ask how much the relative rankings of facilities could change if truncation bias were to be accounted for. A better understanding of the approximate magnitude of the bias introduced by the reporting thresholds can help ascertain whether this bias is a practical, rather than purely theoretical, threat to the validity of the TRI data as a measure of environmental performance.

4 Estimating Bounds on the Truncation Bias

There are several limitations in the TRI data that inhibit systematic estimation of the degree of potential truncation bias. First, the researcher only observes whether a facility reports releases of a chemical in a given year. If a facility does not report for a chemical in year t for which it reported in year t-1, the researcher cannot determine if the facility eliminated use of the chemical, substituted to a different chemical, or was below the reporting threshold.[10]

[10] Facilities that report in one year and then cease reporting in future years represent only a fraction of the truncated observations. There may also be facilities that use a chemical in every year below the reporting threshold, have releases of these chemicals, but are never legally required to report. This non-reporting is equally problematic for policy analysis, but little can be done to identify facilities for which this may be true.

In contrast, the TURA data provide two sources of variation that better allow for the identification of bounds on the truncation bias. First, if a facility does not report for a chemical in year *t* for which it had reported in year *t-1*, the TURA reporting forms include an optional question that asks the facility to explain the change. Approximately, one-third of facilities that cease reporting for a chemical answer the optional question. I check to see if the facilities that respond to this optional question are representative of the set of facilities that cease reporting and find mixed evidence on whether responders are systematically different from non-responders. Responses to the optional question are used to gauge the degree to which facilities are not reporting for chemicals they use in positive quantities and sensitivity analysis is done to determine how responsive the results are to changes in this imputation method.

The second source of variation in the TURA data that can be used to help estimate bounds on the truncation bias is that facilities are required to report how much chemical they *use* in addition to how much they *release*. The federal TRI has reporting thresholds based on use, but only requires reporting on releases. The use data combined with the responses to the optional question about why facilities ceased reporting reveal that the distance from the reporting threshold is a good predictor of whether a facility ceases reporting because it went below the reporting threshold, but still used the chemical in positive quantities. I use this relationship to predict for non-responders the reason why they ceased reporting.

For facilities that directly reveal that they ceased reporting because they went below the reporting threshold, and for facilities that are predicted to have ceased reporting because they went below the reporting threshold, I then estimate the effect of these missing releases on trends in releases over time and on cross-sectional comparisons of facilities within a given year. Missing releases are estimated using three different procedures. The first is a lower bound estimate on total releases at the facility. The lower bound estimate assumes that when a facility ceases reporting its true releases are zero. This is the implicit assumption currently made by government agencies, non-profits, and academic researchers when aggregating releases data to the facility level (or higher levels). The second estimate of missing releases can be considered an upper bound estimate. In the upper bound scenario, if a facility ceases reporting a chemical because it went below the reporting threshold, releases are set equal to the most recent level of releases reported for that chemical at that facility. Moreover, the facility is assumed to continue releasing the chemical at the same level in perpetuity. The final estimate of missing releases is one that extrapolates the value of non-reported releases based upon linear trends in reported releases. These three scenarios present an upper and lower bound and an intermediate estimate of the degree of bias introduced by the reporting thresholds.

Section 4.1 examines why facilities claim they ceased reporting in Massachusetts using the optional question from the TURA form. This section also discusses the estimation used to predict reasons for non-reporting for facilities that did not answer the optional question. Section 4.2 discusses the estimation of the magnitude of missing releases in Massachusetts based on three estimation methods. Section 4.3 discusses the implications of the Massachusetts results for national analyses.

4.1 Why do facilities cease reporting?

Facilities in Massachusetts were required to disclose pollution and chemical use data to the TURA program beginning in 1990. From 1990 to 1999 there were a total of 23,200 chemical reports filed by 1,092 facilities. During this same time period there were 3,758 cases where a facility reported for a chemical in one year, but did not report the chemical in the following year.[11] For these facilities, the TURA form provides an optional question where the facility can explain why it is no longer reporting for this chemical. The question is multiple-choice with the following six possible responses: (1) chemical use is below the reporting threshold but greater that zero, (2) chemical was not used this year, (3) substituted a different chemical, (4) chemical use eliminated without substitution, (5) decline in business, and (6) other. If the facility answers other, they are given the option to fill in a reason. This question is answered for a total of 1,271 (or 33.8 percent) of the cases where reporting ceases. Table 2-1 provides the distribution of responses.

Table 2-1: Explanation for Non-Reporting Among Respondents to the Option Question in Massachusetts

Reason for No Longer Reporting	*Number of Respondents*	*Percentage of All Non-missing Responses*
Chemical use is below the reporting threshold but greater than zero	844	66.40
Chemical was not used this year	96	7.55
Substituted a different chemical	87	6.85
Chemical use eliminated without substitution	60	4.72
Decline in business	31	2.44
Other	153	12.04

Approximately two-thirds of all respondents to the optional question answered that they were no longer reporting because their chemical use was below the reporting threshold, but greater than zero. Of the respondents that answered "other" the most

[11] It is possible that the facility reported that chemical again in future years. This occurs 379 times.

frequent explanation was that the chemical in question had been delisted by the state and reporting was no longer legally required.

The high percentage of respondents that state they ceased reporting because their chemical use was positive, but lower than the reporting threshold, raises concern that the degree of truncation bias may not be trivial. However, before estimating bounds on the truncation bias, it is necessary to determine whether the facilities that responded to the optional question are systematically different from those that did not respond. If responders are systematically different from non-responders in ways that may also be correlated with their reason for not reporting, then the sample of responders cannot be used to impute explanations for non-reporting. To see whether the responders to the optional question are a representative sample, I compare the distribution of the data for the two groups for three key variables—two-digit SIC code, year reporting stopped, and total releases to the environment in year before reporting stopped. Table 2-2 presents the distribution of SIC codes, Table 2-3 presents the distribution of years, and Table 2-4 presents the mean and standard deviation for total pounds released.

Table 2-2: Distribution of Standard Industrial Codes by Optional Question Responders and Non-Responders

	Percentage of Total	
SIC code	***Respondents to Optional Question***	***Non-Respondents to Optional Question***
17	0.24	0.04
20	1.83	1.04
22	4.54	4.23
23	0.48	0.41
24	0.08	0.29
25	1.27	1.41
26	5.74	4.39
27	1.83	1.49
28	21.83	21.72
29	0.08	0.17
30	**5.10**	**7.21**
31	**0.80**	**1.66**
32	0.88	0.99
33	6.22	7.54
34	15.30	14.17
35	**2.15**	**3.69**
36	11.24	10.48
37	3.03	2.61
38	6.06	4.77
39	3.51	2.69
45	**0.00**	**0.21**
47	0.00	0.08
49	**2.79**	**4.85**
51	**4.38**	**2.90**
72	0.56	0.70
73	0.00	0.04
75	**0.00**	**0.21**
76	0.08	0.04

Bold SIC codes indicated that the difference in percentages is statistically significant at the 5% level. The t-test assumes unequal variances across groups.

Table 2-3: Distribution of Years When Reporting Ceased by Optional Question Responders and Non-Responders

	Percent of Total	
Year Reporting Ceased	Respondents to Optional Question	Non-Respondents to Optional Question
1990	10.96	11.76
1991	**11.36**	**8.62**
1992	15.06	13.05
1993	14.67	9.51
1994	**13.56**	**15.31**
1995	8.60	10.68
1996	**8.99**	**8.90**
1997	7.57	7.98
1998	**9.23**	**14.18**

Bold SIC codes indicated that the difference in percentages is statistically significant at the 5% level. The t-test assumes unequal variances across groups.

Table 2-4: Distribution of Reported Pollution Releases by Optional Question Responders and Non-Responders

	Mean	Standard Error
Respondents to Optional Question	16,749	4,333
Non-Respondents to Optional Question	18,102	1,443
Difference between Respondents and Non-Respondents	1,353	4,567

A t-test on the difference in average pollution releases between the two groups cannot reject the null hypothesis that this difference equals zero. The t-statistic is 0.30 allowing the variance between the two groups to be unequal.

The data suggest that responders to the optional question are not systematically different from non-responders at least on total pounds of chemicals released. The difference in the average release levels is 1,353 pounds with a standard error of 4,567 pounds. There are some systematic differences in industry and year reporting stopped. Of the 28 SIC codes, the two groups—responders and non-responders—are statistically different for 7 of them.[12] There are also some differences among responders and non-responders in the years for which reported ceased. Of the nine years, the distribution of responders and non-responders differs in four years.[13] In addition, one may be concerned that there are unobservable differences between the responders and the non-responders that also are correlated with whether the facility goes below the reporting threshold but still uses the chemical in positive quantities. In the absence of a valid instrument that can explain response to the optional question, but not explain going below the threshold, specific sample selection correction models cannot be employed. Rather in this section, I impute reasons for non-reporting for those that did not answer the optional question based on data from the sample of facilities that answered the optional question assuming that responders are reasonably representative on non-responders. I then conduct sensitivity analysis on these results which is presented in Section 2.4.3.

There are several ways one might impute these data. A common method is known as "hot deck" (Ford 1983; Little and Rubin 1987). Essentially, hot deck is a matching strategy—find a facility that responded to the question that looks like a facility that did not respond and assign the matching responders value to the non-responder. A variant of hot deck is to use regression to estimate a relationship between facility characteristics and the explanation for ceasing reporting for those that respond to the optional question, and then using this regression function, predict for non-responders what their explanation would have been.[14] This regression-based imputation strategy is employed here.

---

[12] A Pearson's Chi-squared test rejects equality of the industry distributions across the two groups.

[13] A Pearson's Chi-squared test rejects equality of the year distributions across the two groups.

[14] These two strategies differ in the degree that matching is "enforced" and the functional form assumption. Hot deck is a non-parametric strategy that does not impose a specific functional form on the relationship between the covariates and the response variable. Regression is loose matching, but

To construct the prediction relationship, I first create a new dummy variable that takes a value of one if the facility ceased reporting because it went below the reporting threshold and zero otherwise. I then use this variable as the dependent variable in a logit estimation on observable facility characteristics. The task is then to compile a set of observable characteristics that explain whether a facility will go below the reporting threshold.

One characteristic that may explain the propensity to go below the reporting threshold is the facility's distance from the reporting threshold. Degeorge, Patel and Zeckhauser (1999) demonstrate that the existence of performance thresholds for managers induces specific changes in their earnings management, with managers managing to the thresholds. For example, empirically there appears to be a concentration of profits just above zero for managers that are compensated based on whether their unit earns positive profits. Degeorge, Patel and Zeckhauser refer to this as threshold-regarding behavior. Threshold-regarding behavior is particularly pronounced among units that are very close to the profit threshold. In other words, if your unit is quite far from earning positive profits, you do not try to manipulate earnings much, because no amount of manipulation will cause your unit to earn positive profits. But, if your unit is very close to the positive profit threshold, manipulation of earnings is more valuable.

In the context of the TURA reporting thresholds, we might expect to see similar threshold-regarding behavior. Facilities that are very close to reporting threshold in year t have a greater incentive to manage their chemical use so that they fall below the reporting threshold in $t+1$. To measure the distance from the reporting threshold, I construct a

---

does impose a functional form. The regression-based matching will perform well when the assumed parametric specification is a good approximation to the average response function.

variable that measures, for each chemical, how far the facility is from each of the reporting threshold. In Massachusetts, this process is simplified by the fact that once the facility triggers the reporting threshold for one chemical, the reporting threshold for all chemicals is 10,000 pounds of combined manufacture, process, and other use amounts. So in Massachusetts there are not three separate reporting thresholds, but one binding threshold at 10,000 pounds of total use. The facility's distance from the reporting threshold is then given by total use minus 10,000 pounds. The hypothesis is that the greater the distance from the reporting threshold, the less likely a facility is to cease reporting that chemical because its use of the chemical went below the reporting threshold.

Similarly, if there is a relationship between chemical use and chemical release, then total releases of the chemical in time *t* may predict whether the facility goes below the reporting threshold in year *t+1*. Thus, total releases are also used as an explanatory variable. Other potential explanatory variables include industry dummy variables, and year dummy variables that proxy for changes in industry best practices and exogenous technological change. The relationship estimated is then given by:

$$\text{below\_threshold}\begin{pmatrix}0\\1\end{pmatrix} = \alpha + \beta_1 \text{Distance from Threshold} + \beta_2 \text{releases} + \beta_{\text{sic}} SIC + \beta_t Year + \varepsilon$$

The results of this estimation are present in Table 2-5.

Table 2-5: Results for Estimation of Threshold Logit for Respondents to Optional Question

	Coefficients	Percent Increase in Baseline Probability Resulting from a 10% Decrease from Mean
Constant	1.02 (1.19)	NA
Distance from Threshold	-0.00001 *** (0.000003)	6.5%
Releases	-0.00003 *** (0.000006)	3.1%
SIC dummies	Yes +	NA
Year dummies	Yes +	NA
Number of observations	1,251	
Pseudo R-squared	0.17	
Baseline Probability	60.0 %	

Baseline Probability is the probability evaluated at the mean value of all continuous explanatory variables and at zero for all binary variables.
*** Significant at the 1% level
+ An F-test shows variables are jointly significant at the 1% level

Because the logit estimation is non-linear, the coefficients presented do not convey information about the marginal effect of a change in one of the explanatory variables on the dependent variable. To determine what the magnitude of the effect of each explanatory variable is on the probability that a facility ceases reporting because it goes below the reporting threshold, I first calculate the baseline probability—that is the probability of going below the threshold predicted by the logit equation when all of the continuous variables are set at their mean value and all of the binary variables are set at zero. The baseline probability evaluated at the mean is 60.0%. I then decrease each covariate in turn by 10% from its mean value and report the percentage point change from the baseline. Using this method, a 10 percent decrease from the average distance from the reporting threshold results in a 6.5 percentage point increase in the probability of going below the reporting threshold. Similarly a 10 percent decrease in total releases to the environment results in a 3.1 percentage point increase in the probability of going below the reporting threshold.

This probability function can then be used to predict whether facilities that ceased reporting for a chemical, but did not explain why, were likely to have ceased reporting because they went below the reporting threshold. The above equation was used to predict the probability of going below the threshold for the 2,482 observations with no answer to the optional question. Observations with a predicted probability greater than 50 percent were coded as going below the threshold. Table 2-6 provides a breakdown of the observations that ceased reporting by explanation. Of the 2,482 observations for which no explanation for non-reporting was provided, 1,786 (72.0 percent) were predicted to have stopped reporting because the facility went below the reporting threshold for that chemical, but still use the chemical in positive quantities.

Table 2-6: Distribution of Observations (Facility-Chemical-Year) that Cease Reporting by Explanation

	Explanation Provided by Facility	Explanation Predicted	Total
Below the Reporting Threshold	844	1,786	2,630
All Other Reasons	424	696	1,120
Total	1,268	2,482	3,750

In summary, analysis of the Massachusetts TURA data indicates that a substantial percentage of facilities that cease reporting a chemical do so because they go below the reporting threshold for that chemical, but still use it and may still have positive releases of these chemicals. While this frequency of threshold-regarding behavior seems to present some concern about the validity of the TRI data for making comparisons among facilities or across time, the level of concern may still be low if the total amount of releases that disappear from the registry is small. That is, perhaps the percentage of observations affected by the reporting thresholds is large, but the total share of releases represented by these observations is small. The next section addresses the question of the likely magnitude of the truncation bias.

4.2 What is the Magnitude of Missing Releases in Massachusetts?

The difficulty in assessing the effect of truncation at the reporting thresholds on the validity of the TRI data is that it is impossible to know how large a problem non-reportable releases are, precisely because these releases are no longer reported. The best we can do is assess how large this problem might reasonably be, given observable information. To that end, this section focuses on estimating the magnitude of "missing" releases by estimating bounds on the possible size of these releases—a lower bound estimate, an upper bound estimate, and a linear projection estimate.

To make these results meaningful, I restrict the analysis to chemicals and industries that have been subject to the TURA reporting requirements since its inception—the so-called core group. This ensures that we are examining variation in reported releases due to facility behavior around the thresholds and not changes in releases due to changes in the regulatory requirements themselves. The analysis is also done for two different measures of pollution from the TRI. The first is on-site releases. On-site releases are releases of the pollutant to air, water, land, or under-ground injection that occur at the facility's location. The second measure is total on- and off-site releases. Total releases include on-site releases, but add transfers of waste to offsite locations for disposal or recycling. In general, it is thought that total releases is over-inclusive as a measure of environmental performance because some of the transfers are not pollution, but are transfers for recycling and reuse. On-site releases, however, are under-inclusive as a measure of environmental performance because some waste that is transferred off-site is pollution attributable to the facility. Researchers have used both on-site and total releases as measures of environmental performance, so it important to see if the reporting thresholds have different effects for the two measures.

The lower bound estimate of the value of non-reported releases is that these releases are zero. This assumes that once a facility drops below the reporting threshold for a given chemical, they no longer release any of that chemical. A conservative upper bound estimate of the value of non-reported releases is that they equal the last reported value of releases for that chemical at that facility. Thus, if a facility reports releases of 500 pounds of a chemical in year t and then drops below the reporting threshold, the upper bound estimate of missing releases is that this facility releases 500 pounds a year of that chemical in perpetuity.[15]

One might argue that assuming unobserved releases are either zero or set at their most recent value in perpetuity are extreme assumptions. There is some evidence that suggests that, on average, setting releases equal to the last reported value is not as extreme an assumption as it may first appear. There are 287 observations (1 percent of the total) where the facility stops reporting for a chemical in one year because it went below the reporting threshold and then in the future the facility begins reporting for that chemical again. How do releases of the chemical in future years compare to the reported releases in the last reported year? On average, the future reported releases are 796 pounds *greater* than

[15] Occasionally a facility will report for a chemical for a few years, then stop reporting for that chemical because they are below the reporting threshold, and then begin reporting for the chemical again in later years. In this case, the upper bound estimate of releases equals observed releases in any year in which the facility actually reports releases and in non-reporting years are set equal to the releases in the most recent year in the past for which the facility reported releases for that chemical.

the last reported releases for that chemical.[16] Of course, this does not rule out the possibility that assuming releases are constant for non-reporting facilities is a conservative upper bound. Facilities that waver above and below the reporting threshold are distinct from facilities that go and remain below the reporting threshold. Thus, we might not expect the release behavior of the facilities that oscillate around the reporting threshold to be indicative of the release behavior of facilities that go below the reporting threshold and stay there forever. But this evidence does suggest that using the last reported releases as an estimate of future reported releases might be a reasonable upper bound.

Providing upper and lower bounds on the potential bias in reported releases induced by truncation at the reporting thresholds provides useful information on how large the bias may be. However, it does not provide any information on the probability distribution of the true bias within that range. While this range is useful, perhaps, we would wish to also have an estimate of bias under a more probable scenario. An alternative assumption about the behavior of releases among non-reporting facilities is to assume that facilities do not fundamentally behave in different ways with respect to their release decisions based on whether they are above or below the reporting threshold. If this is the case, then the trends in non-reported releases might be expected to increase or decrease at the same rate as the trends in observed releases. Under this assumption, one can project trends in releases for non-reportable data based on observed trends in releases among reported chemicals.

How can we predict the trend in reported releases so that it can be extrapolated to non-reporting facilities? One might hypothesize that the amount of the chemical used, the

[16] The largest future drop in reported releases is by 63,571 pounds and the largest future increase in reported releases is 105,600 pounds.

type of chemical, the industrial sector, and other similar factors. would all be good predictors of how much chemical a facility will release in a given year. While all of these factors do explain chemical releases, taken together they only explain about seven percent of the variation in reported releases.[17] The best predictor of reported releases turns out to be a distributed lag of past releases. A simple one period distributed lag, where current releases of each chemical are regressed on releases of that chemical from the previous year, explains 80% of the variation in releases. Increasing the number of lags increase the predictive power slightly, but makes the equation less useful for prediction because a smaller number of observations have multiple lags. The results for one-, two-, and three-period distributed lag models are presented in Table 2-7. Using the one-period distributed lag model, current on- and off-site releases are, on average, 95 percent of previous years' releases. Similarly, using a one-period distribution lag model, current on-site releases are 80 percent of previous years' releases.

[17] This estimate comes from a regression of releases on total chemical use with two-digit SIC code, chemical, and year fixed effects. Including facility-chemical fixed effects increases the predictive power of the regression to 65 percent. However, the distributed lag model, which yields a higher predictive power, also is preferred because all of the data are observable. That is, it relies only on past releases, which we observe for all facilities that cease reporting.

Table 2-7: Distributed Lag Models of Reported Releases

	Total Releases (On- and Off-site)			On-site Releases		
Specification	1	2	3	1	2	3
1 year lag	0.947 ***	0.865 ***	0.872 ***	0.798 ***	0.740 ***	0.737 ***
	(0.054)	(0.131)	(0.149)	(0.036)	(0.084)	(0.115)
2 year lag		0.105	0.086		0.087	0.130
		(0.134)	(0.143)		(0.068)	(0.089)
3 year lag			0.017			-0.021
			(0.083)			(0.046)
Observations	8586	6572	4964	8927	6952	5330
Adjusted R-squared	0.779	0.795	0.804	0.782	0.807	0.808

*** indicates the coefficient is statistically significant at the one percent level.

Using the coefficients from the one-period distributed lag model, I predict releases for each facility that ceases reporting. This generates trends for facilities that cease reporting. These projected trends are downward sloping for all facilities that cease to report for a chemical, and the rate of decrease is larger for on-site releases than for total releases. Figure 2-3 diagrams the process of estimating lower, upper, and linear projected releases for a facility and Table 2-8 provides data on the magnitude of missing releases, both total releases (on and off-site) and on-site releases, using both the upper bound assumption and the linear projection.

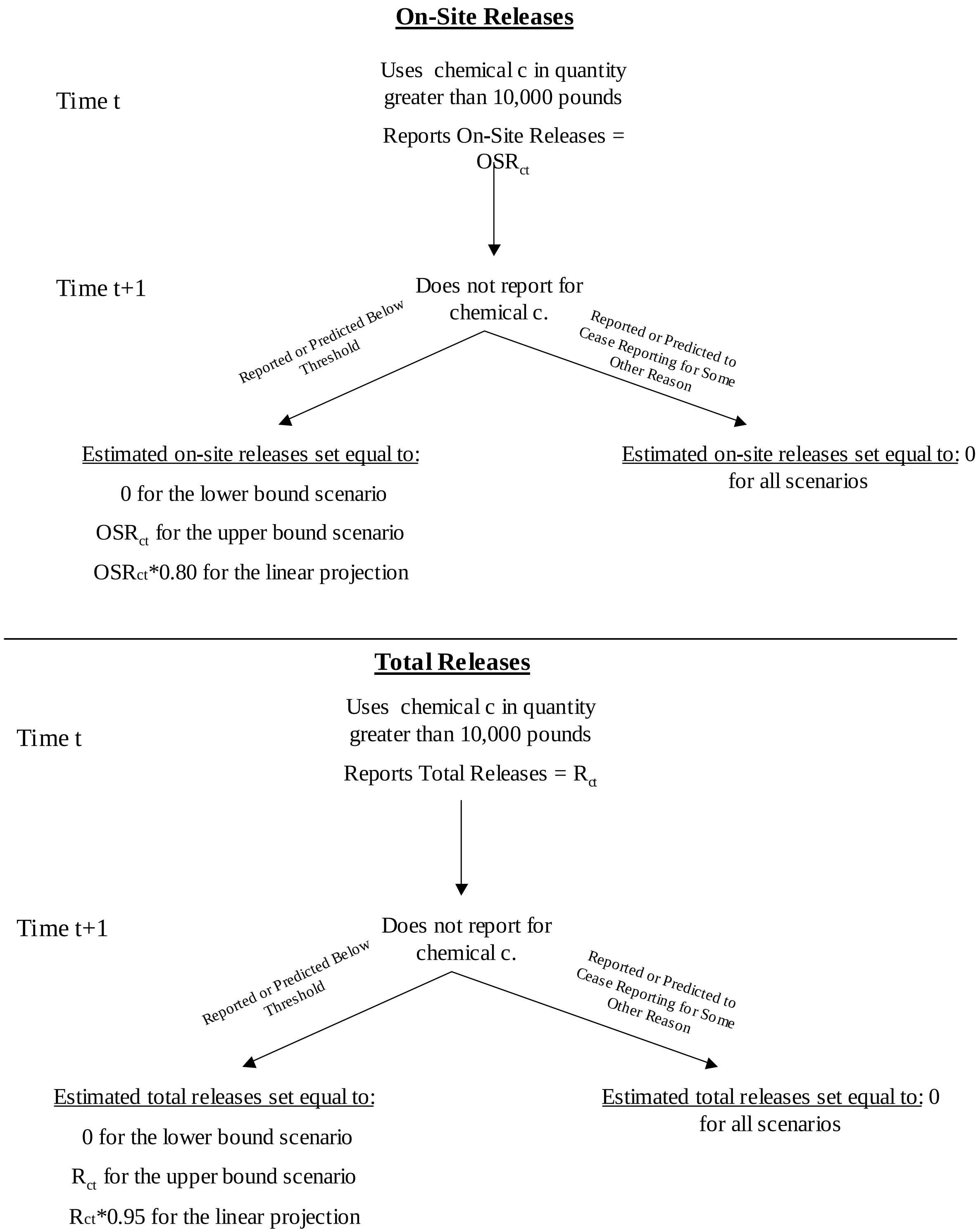

Figure 2-3: Flow Chart for Imputation of "Missing" Releases

Table 2-8: Estimates of TURA Releases in Massachusetts (in 1,000s of Pounds)

	Total Reported Releases (On- and Off-Site)				On-Site Releases			
year	Lower Bound	Linear Projection	Upper Bound	% Difference Between Lower and Upper Bounds	Lower Bound	Linear Projection	Upper Bound	% Difference Between Lower and Upper Bounds
1991	43,372	45,291	45,392	4.7%	17,653	18,915	19,230	8.9%
1992	44,870	48,495	48,777	8.7%	15,573	17,517	18,288	17.4%
1993	38,199	43,696	44,245	15.8%	11,862	14,598	16,015	35.0%
1994	38,833	45,141	45,998	18.5%	9,300	12,100	14,170	52.4%
1995	36,557	43,493	44,692	22.3%	8,715	11,238	13,886	59.3%
1996	36,984	44,308	45,864	24.0%	7,451	9,855	13,061	75.3%
1997	32,449	39,848	41,775	28.7%	6,367	8,553	12,271	92.7%
1998	33,048	40,898	43,228	30.8%	5,621	7,697	11,905	111.8%
1999	19,967	28,282	31,032	55.4%	5,126	7,113	11,786	129.9%

The results indicate that the degree of missing releases in the early years is relatively modest as a percentage of total releases. In 1991, missing releases are only about five percent of total (on- and off-site) releases and between seven and nine percent of reported on-site releases. However, over time, missing releases as a percentage of total reported releases rise dramatically. There are two reasons for this. First, missing releases are cumulative. In every year, about two to four percent of the previous year's releases are not reported due to facilities going below the reporting threshold. But in each year the total amount of missing releases are all those releases that are no longer being reported by all facilities, these include the releases from facilities that stopped reporting for the chemical this year, as well as the missing releases from facilities that went below the reporting threshold two years ago, three years ago, and so forth. Particularly in the upper bound scenario, when facilities that drop below the reporting threshold are assumed to continue releasing at the same level in perpetuity, the cumulative effect can be quite large.

The second explanation is that reported releases fall considerably over time, even for facilities that continue to report for all chemicals. For the upper bound case, where non-reporting facilities are assumed to continue to release at the same level forever, the relative importance of these releases increases over time—the missing releases stay the same, but the total reported releases decrease—resulting in a sharp increase in missing releases as a percentage of total reported releases. In fact, for on-site releases the difference between the upper and lower bound estimates of releases differs by 130 percent in 1999.

The data in Table 2-8 illustrate that the magnitude of missing releases generated by the existence of the reporting threshold is non-trival. But does this result in significantly biased estimates in the trend in environmental performance over time? Figures 2-4 graphs the trend in total releases over time for each of three scenarios: (1) only reported releases (lower bound estimated assumes non-reported releases equal zero), (2) reported releases plus the upper bound estimate of missing releases (sets missing releases equal to their last reported value for all years), and (3) reported releases plus linearly extrapolated estimates of missing releases.

Looking at total releases first, it is clear that the estimate of the trend in environmental performance is substantially affected by the exclusion or inclusion of the estimated releases for non-reporting facilities. Using a lower bound assumption, that all non-reported releases are zero, the change in total releases from 1990 to 1999 is 36.7 percent. However, if one instead uses the upper bound assumption, that facilities that fall below the reporting threshold continue to release the same amount forever, the change in total releases over the same ten year period was only 0.6 percent. Using the more moderate assumption that facilities that are no longer reporting continue to decrease releases over time at the same rate as the average reporting facility, the change in total releases over the ten-year period is 10.1 percent.

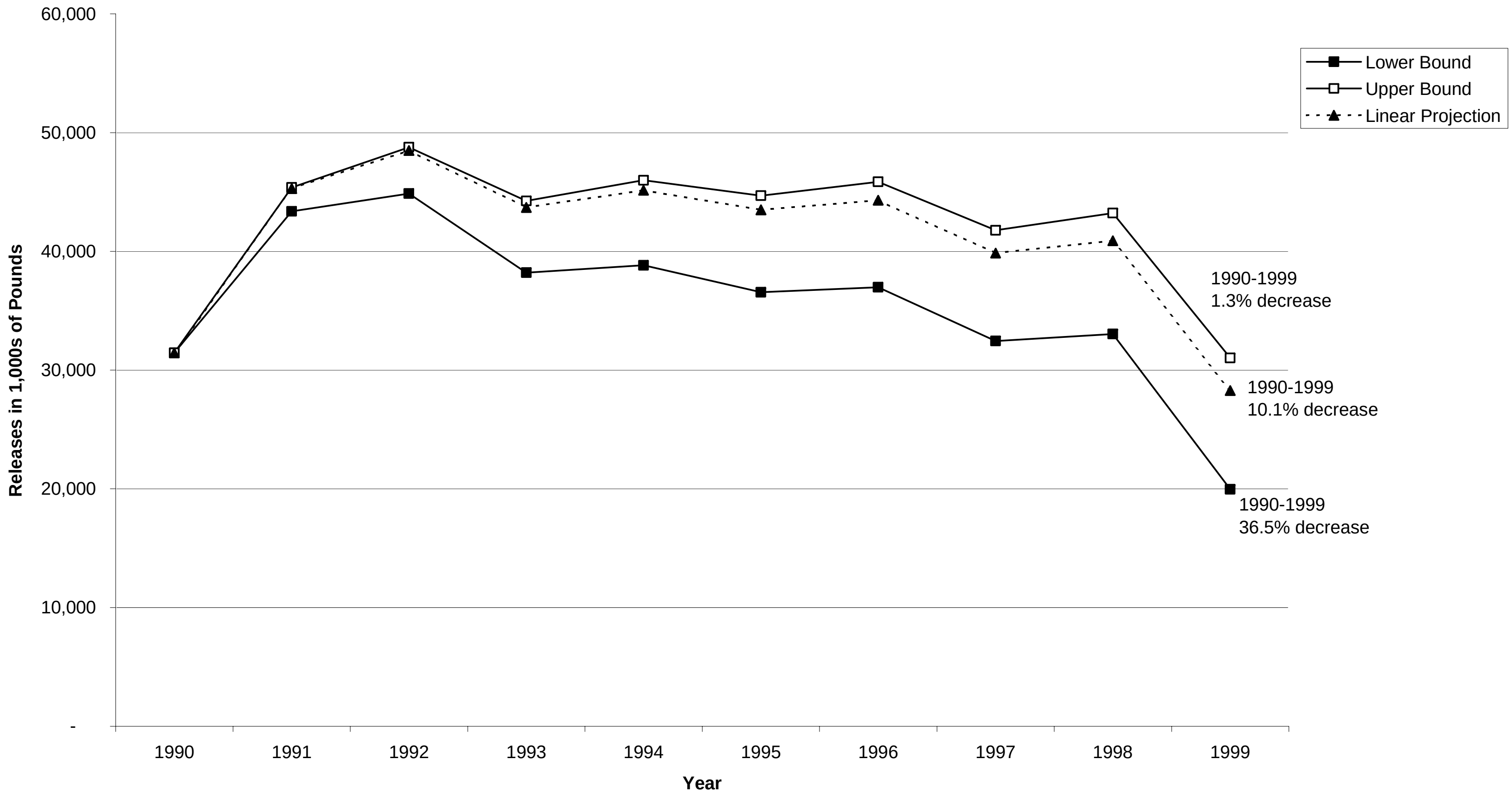

Figure 2-4: Trends in Total (On- and Off-Site) Releases in Massachusetts

The story is somewhat less stark, but still significant, if one ignores the upward trend in reported off-site releases between 1990 and 1991 and instead examines trends only for the years 1991 to 1999. For this period, the lower bound assumption leads to the conclusion that total releases have fallen by 54.0 percent, the upper bound assumption estimates the decline at 31.6 percent, and the linear projection leads to an estimate of a 37.6 percent decrease. Using the difference in trends from 1991 to 1999, it appears that in the worst case, non-reported releases by facilities that fall below the reporting threshold may account for approximately 22 percentage points of the total 54 percentage point decrease in reported releases, or roughly 40 percent.

The potential degree of bias introduced by the reporting thresholds in the trends for on-site releases is less pronounced, although still sizeable. Figure 2-5 provides the estimated trends in on-site releases. Using the lower bound assumption, the decrease in on-site releases from 1990-1999 was 75.7 percent. Using the upper bound assumption, the decrease over this period was 44.0 percent. Finally using the assumption of linearly projected decreases in non-reported releases, the change over the decade was 66.2 percent. Thus, of the observed 75 percent decline in reported releases, from 1990-1999 as much as 31.6 percentage points (or 40 percent) of this decrease may be accounted for by non-reported releases.

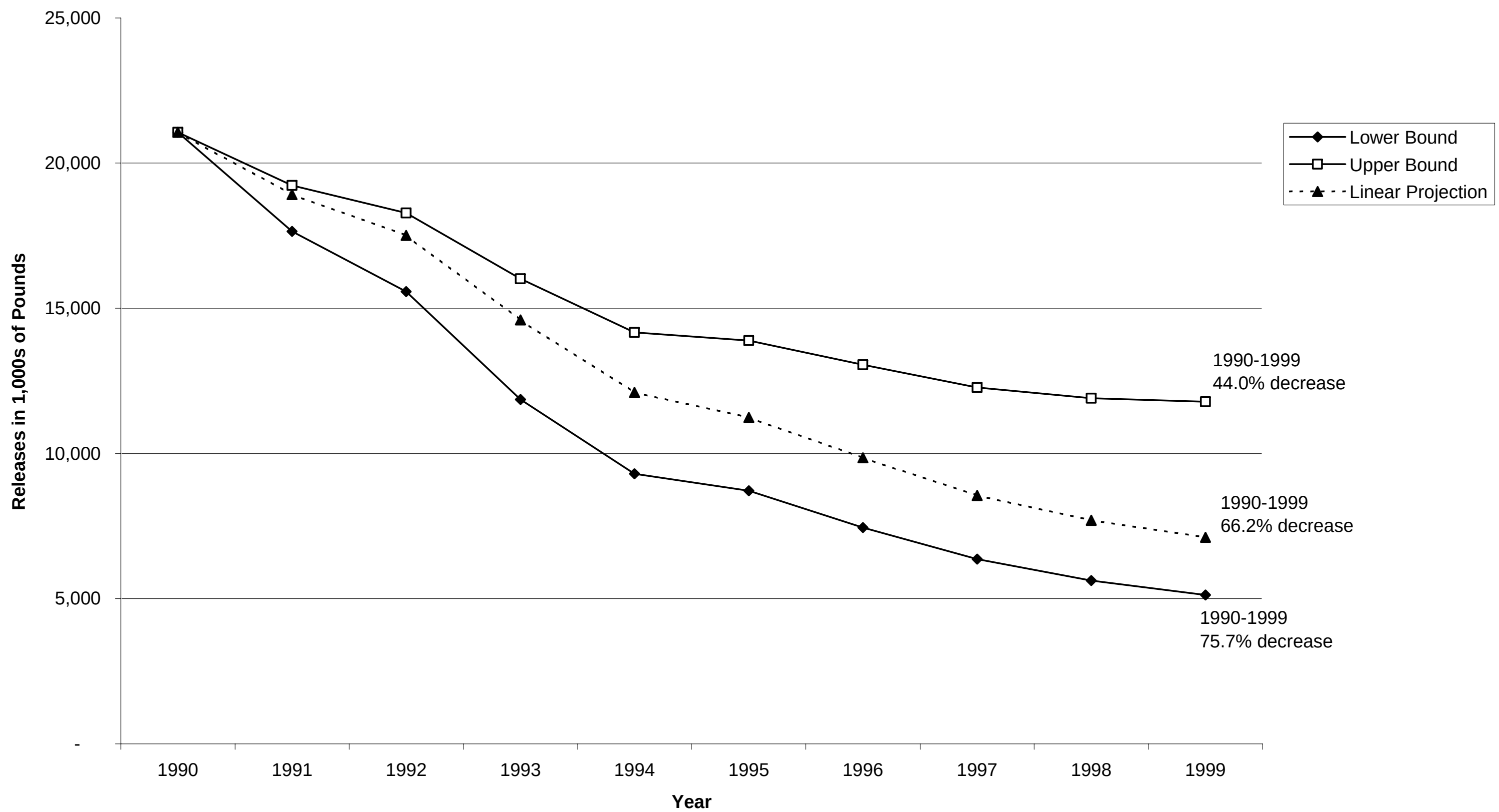

Figure 2-5: Trends in On--Site Releases in Massachusetts

As evidenced by these data, our understanding of the magnitude of the change in environmental performance is substantially affected by the truncation bias introduced by the reporting threshold. Making different assumptions about what happens to releases when facilities cease to report because they fall below the reporting threshold makes a big difference in our understanding of the level of environmental improvement. Having said that, it is somewhat reassuring that even in the upper-bound scenario, on-site releases of toxic chemicals have fallen by 44 percent in Massachusetts. In this case, using TURA data to conduct policy analysis is very likely to result in erroneous cardinal estimates of the magnitude of improvement, but unlikely to result in erroneous ordinal estimates—things have improved by somewhere between 44 and 76 percent.

A separate concern is whether the reporting thresholds induce bias in cross-sectional comparisons of facilities. Imagine that we rank facilities from lowest releases to highest releases. If the degree of non-reported releases is small relative to any given facility's total reported releases, adjusting for these non-reported releases will change our estimate of how much a facility releases, but may not change the facility's rank much. If on the other hand, the magnitude of missing releases is large relative to total reported releases for some set of facilities, then adjusting facility totals to account for this may substantially changing the rankings of facilities.

To investigate the degree of cross-sectional bias potentially introduced by the reporting thresholds, I first divide facilities into quartiles based on their reported releases. Thus, facilities whose reported releases are in the lowest 25 percent of all facilities in that year are assigned to quartile number one. Facilities whose releases are among the highest 25 percent of all facilities are assigned to quartile number four. These are the quartile rankings associated with the lower bound assumption that all non-reported releases are equal to zero. I then recalculate the facility's quartile ranking based on the upper bound assumption and the linear projection. If the relative importance of the non-reported releases is small, we would expect to see few facility quartile rankings change. In this case we would not be terribly concerned about the potential cross-sectional bias introduced by the reporting thresholds. However, if the relative importance of missing releases is large, we might expect to see more facility quartile rankings change dramatically. An alternative way to think about this is to imagine a policy maker assigning a facility a grade based on its pollution levels relative to other facilities in a given year. A facility can get a grade of excellent, good, fair, or poor. If we adjust for missing releases resulting from facilities going below the reporting threshold, do we change our grades for only a few facilities or for a substantial number of facilities? The results of the quartile distribution comparison are provided in Tables 2-9 and 2-10 for total and on-site releases, respectively.

Table 2-9: Quartile Ranking Comparisons for Total Releases

		Upper Bound Ranking			
		1	2	3	4
Lower	1	**1,245**	885	799	275
Bound	2	297	**367**	25	4
Ranking	3	260	481	**659**	102
	4	0	64	317	**1,413**

Table 2-10: Quartile Ranking Comparisons for On-Site Releases

		Upper Bound Ranking			
		1	2	3	4
Lower	1	**1,430**	629	864	515
Bound	2	233	**321**	30	6
Ranking	3	170	640	**491**	70
	4	0	176	415	**1,230**

In each table, the diagonal elements (in bold) represent facilities whose quartile ranking is unaffected by inclusion of estimates of their missing releases. All off-diagonal entries are facilities whose quartile rankings differ under the upper and lower bound assessments. For total releases, 31 percent of facilities have different quartile rankings under the upper bound assessment than under the lower bound assessment. For on-site releases 52 percent of facilities have different quartile rankings under the two different scenarios.

For the purposes of using TURA or TRI rankings for regulatory targeting purposes, perhaps one is most concerned about errors in the top and bottom quartile. That is, one is concerned most with mislabeling a facility as "poor" or "excellent." Focusing on these quartiles, we can see from the data in Table 2-9 that 21 percent of the facilities that are labeled "poor" based on their reported releases would have been labeled either "fair" or "good" once reported releases are adjusted to include an upper bound assessment of missing releases. In addition, 61 percent of facilities that were labeled "excellent" using only reported releases would have been labeled "good," "fair," or "poor" if missing releases are included. The results for on-site releases are quite similar (Table 2-10). For on-site releases 33 percent of facilities labeled "poor" using reported releases should have received a higher grade and 58 percent of facilities labeled "excellent" should have received a lower grade.

This potential variation in ordinal rankings of facilities within a given year is large enough to be of substantial concern. Indeed, the degree of variation in the cross-section rankings of facilities seems more troubling than the variation in trends over time. In the trends, we were confident in the direction of the change (there had been an improvement), but not in the magnitude of the change. With the cross-sectional rankings we are potentially assigning the wrong grade two-thirds of the time in the best quartile and a quarter of the time within the worst quartile. If regulatory or enforcement resources are targeted based on rankings of reported releases, these resources are likely to be misallocated.

4.3 Sensitivity Analysis

One of the key components of the above analysis was imputing whether or not a facility ceased reporting because it went below the reporting threshold if the facility did not answer the optional question. As discussed in Section 2.4.1, there were some observable differences between facilities that responded to the optional question and those facilities that did not respond. How sensitive are the findings to changes in the imputation method? I try three alternative imputations and provide information on the impact of these changes on the estimates of truncation bias both for trend and cross-sectional analyses.

The first sensitivity analysis is to change the probability cutoff from the logit that determines whether a facility that does not answer the optional question stopped reporting because it went below the reporting threshold. In the main analysis, facilities with a predicted probability of going below the threshold greater than 50 percent from the logit were assumed to have ceased reporting because they went below the reporting threshold. The first test is to increase this probability cutoff to 75 percent.

The second sensitivity analysis is to only examine missing releases for facilities that actually respond to the optional question and state that they cease reporting because they went below the reporting threshold, but still use the chemical in positive quantities. No imputation is made for facilities that do not answer the optional question. This is essentially the same as increasing the probability cutoff from the logit to 100 percent.

The third analysis does not designate a cutoff, but rather weights releases by the predicted probability that the facility went below the threshold from the logit. Thus if the logit yields a predicted probability that a facility went below the reporting threshold, but still uses the chemical of 20 percent, then the upper bound estimate of that facility's releases is 20 percent of the last reported value in perpetuity.

Table 2-11 presents the results of the sensitivity analysis for estimates of the effect on the trend in total and on-site releases. For comparison, the first column contains the original estimates (where the probability cutoff equals 50 percent). Under the most conservative scenario, which only used data from facilities that actually answered the optional question, up to 13 percent of the observed decrease in total releases and up to 15 percent of the observed decrease in on-site releases from 1991-1999 could be due to missing releases.

Table 2-11: Maximum Percent of Observed Decline in Releases from 1991-1999 That Could be Explained By Threshold Regarding

	50%	75%	100% (responders only)	Weighted by Probability from Logit
Total Releases	41.4	22.3	13.1	37.3
On-Site Releases	45.4	25.9	14.6	40.1

Sensitivity analysis was also done on the cross-sectional rankings. Table 2-12 presents the results for the differences between the upper bound and lower bound quartile rankings for all three modifications to the probability cutoff (75%, 100%, and expected value). The first column of the table provides the main findings based on a 50 percent cutoff for comparison. In the most conservative scenario, the quartile rankings are wrong about 28 percent of the time. The rankings are only wrong about 14 percent in the bottom quartile. That is only about 14 percent of facilities that would have received a label of "poor" using the reported data should have received a higher grade if missing data were incorporated. In the upper quartile the percent error is still higher than average. Approximately 30 percent of facilities that would have been labeled "excellent" based on reported releases would have received a lower ranking if missing releases had been accounted for.

Overall the sensitivity analysis supports the general findings presented in the main analysis. First, a potentially significant share of the decrease in observed releases may be due to facilities no longer being legally obligated to report releases because their use of the chemical is below the reporting threshold. Second, the reporting thresholds may also skew the cross-sectional rankings of facilities. In particular, facilities that appear to be low releasers or good environmental performers based on TRI releases may actually not be better than other facilities with higher releases. The error in cross-sectional ranking diminishes as one moves down the distribution. The rankings are considerably less wrong about identifying the worst facilities.

Table 2-12: Maximum Potential Error in On-Site Cross-Sectional Rankings Due to Threshold Regarding

	50%	75%	100% (responders only)	Weighted by Probability from Logit
Total	52%	42%	28%	51%
Bottom Quartile	33%	21%	14%	30%
Top Quartile	58%	48%	30%	59%

4.4 Extrapolating to the National Level

So far we have examined the degree of bias potentially introduced by the reporting thresholds into both trend and cross-sectional measures of environmental performance for Massachusetts' facilities only. The reason for focusing on Massachusetts was one of data availability. Data from TURA are critical in estimating the magnitude of the truncation bias.

Can we extrapolate the findings from Massachusetts to estimate the degree of truncation bias in national trends and cross-sectional comparisons? Unfortunately, any such precise extrapolation would be shaky, at best, and outright misleading at worst. While the results for Massachusetts do give a strong reason to be concerned about truncation bias affecting the validity of national TRI data, the actual bias at the national level could be lower or higher than in Massachusetts. The set of Massachusetts' facilities is far from a representative sample of national facilities reporting to the TRI. On average, facilities in Massachusetts have reported releases that are an order of magnitude smaller than average releases for facilities in other states. The average total releases between 1990 and 1999 for Massachusetts' facilities were 22,971 pounds while the average for all other facilities were 166,368 pounds.[18] These substantial differences in the only observable variable that might be used to both predict which facilities cease

[18] This difference is not only due to a smaller variance among Massachusetts facilities. If one graphs the distribution of both total and on-site releases for Massachusetts and all other states, the entire distribution for Massachusetts' facilities is shifted to left. A Kolmogorov-Smirnov test rejects the null hypothesis of equality of the distribution at the 1% level.

reporting because they went below the reporting threshold and to estimate the magnitude of missing releases, make valid extrapolation infeasible.

Even on an intuitive level it is difficult to predict how the precise degree of truncation bias nationally will compare to the estimated degree of bias in Massachusetts. On the one hand, the reporting thresholds are lower in Massachusetts than for the federal TRI program. In Massachusetts, once a facility triggers a reporting threshold for a single chemical, the facility must report for all chemicals for which manufacturing plus processing plus otherwise use amounts are greater than 10,000 pounds. For the federal program reportability is calculated separately for each chemical based on the 25,000 pounds manufacture or process and 10,000 pounds otherwise use thresholds. This might imply that truncation bias is likely to be a larger problem for the federal TRI data than for the Massachusetts TURA data.

On the other hand, Massachusetts has an aggressive state-level pollution prevention and reporting program that provide additional incentives for facilities in that state to reduce use of their chemicals below the regulated level. Thus, we might expect to see more threshold-regarding behavior in Massachusetts than in the rest of the country. Similarly, Massachusetts' facilities do have lower average releases. This may also indicate that Massachusetts' facilities are, on average, closer to the reporting threshold and we would expect to see more threshold–regarding behavior in that state than in the nation as a whole.

While I cannot provide any specific estimate of the degree of truncation bias in the national trends or cross-section comparisons using TRI data, the experience in Massachusetts does suggest that concern over truncation bias is well-founded. In addition, some preliminary evidence from decreases in the federal reporting thresholds for lead provide further evidence that the reporting thresholds may affect inferences from the federal TRI data.

In 2001 the reporting threshold for lead was lowered to 100 pounds at the federal level. For the 2000 reporting year, there were 1,997 facilities reporting for lead and lead compounds and total reported releases of lead were 374 million pounds. In 2001, the first year of reporting under the lower threshold, there were 8,444 facilities reporting a total of 443 million pounds. This represents a net increase of 69 million pounds (19 percent). About half of this increase--33.5 million pounds--is attributable to facilities that did not report on lead and lead compounds in 2000.[19]

Given these results, what implications should we draw about how to use TRI data as a measure of environmental performance for policy analysis? That is the topic of the final section.

5 Policy Implications for Using TRI as a Measure of Environmental Performance

The above analysis indicates that the existence of the reporting thresholds may introduce substantial bias in both the trend and cross-sectional estimates of environmental performance using reported TRI releases. However, the TRI data are currently one of the best sources of facility-level pollution levels nationwide. What is the policy analyst to do? The following recommendations are likely to enhance the validity of studies that use TRI data for policy analysis.

[19] Personal communication, Cody Rice, Office of Pollution Prevention and Toxics, U.S. Environmental Protection Agency, April 26th, 2004.

First, the number of chemicals reported by a facility should be treated as an additional policy outcome.[20] If EPA is investigating the effectiveness of a new regulation, it is not sufficient to examine only the effects on total TRI releases (even if these releases are adjusted for toxicity). If the policy also has an effect on the number of facilities that reduce chemical use below the reporting threshold, then estimates of the policy's effect on TRI releases are likely to be biased upward—that is, the effect of the policy will be overstated. If however, one estimates that the policy does not have an effect on the number of chemicals reported, but does have an effect on total releases, one can feel more confident that the policy has actually improved environmental performance and not simply reduced reporting.[21]

If one is only concerned about whether a policy had a positive effect on environmental performance, then a worst-case estimate of the effect may be appropriate. For this worst-case estimate one would assume that any facility that ceases reporting for a chemical did so because it went below the reporting threshold (as opposed to eliminating the chemical, going out of business, or so forth). The facility's releases could then be set to their last reported level in perpetuity. This clearly overestimates total releases, but if the policy is still found to lower releases even under this extreme

20 It may be tempting to normalize total releases by how many chemicals the facility reports. For example, one could compare facilities both cross-sectionally and over time based on their average releases per report. However, this correction does not remove the truncation bias. For example, imagine a facility releases uses three chemicals and in the first reporting years reports releases of 200 pounds of Chemical A, 200 pounds of Chemical B, and 150 pounds of Chemical C. In that year the average releases per chemical reported is 183 pounds. In the second reporting year the facility only reports for two chemicals. It still releases 200 pounds of chemical B and 150 pounds of Chemical C. Now the average releases per reporting year are lower at 175 pounds. Releases per chemical reported declined, but the releases for that facility's reported chemicals do not change across those two years.

[21] See the first chapter of this dissertation for an example of a policy analysis that estimates the effect of the policy on both releases and number of chemicals reported.

assumption, then one can feel confident in the policy's effectiveness. The magnitude of the effect will be wrong, but the direction, if positive, will be correct.

One could also do sensitivity analysis on the directional effect of the policy (although again not on the magnitude of the effect) by comparing the results for the whole sample to the results for a sample only of facilities that report for the same chemicals over the relevant time frame. If one estimates positive effects of the policy in both samples, then the effects are not due only to decreases in reporting.

Such sensitivity analysis would be greatly improved by the addition of a question on the federal reporting form that asks facilities why they are not reporting for a chemical in the current year for which they reported in previous years. This question, similar to the one used on the Massachusetts TURA form, could help EPA and others assess the potential for truncation bias.

Unfortunately, the only fail-proof way to ensure that truncation bias will not affect the results, and the only way to ensure the magnitude of policy estimates is accurate, is to eliminate the reporting thresholds. EPA has the regulatory authority to change the reporting thresholds, and has done so on two occasions. In 1999, EPA reduced the reporting threshold for persistent and bio-accumulating toxins (the threshold was reduced to 10 or 100 pounds depending on the chemical). In 2000, EPA reduced the reporting threshold for lead to 100 pounds. Obviously, there are costs associated with lowering-reporting thresholds. These costs include administrative costs for promulgating a series of notice-and-comment rulemakings that are likely to be contentious. There are also substantial paperwork compliance costs for facilities that will be required to report for chemicals that were previously unreportable. These costs were estimated at 80 million dollars for the first year of reporting under the lower threshold for lead (U.S. EPA, 2001).

Despite the costs, there may be important benefits from reducing or eliminating the reporting thresholds. For example, EPA argues that responsible use of TRI data allows "Federal, state, and local governments to compare facilities or geographic areas, identify hot spots, evaluate existing environmental programs, and track pollution control and waste reduction progress" (EPA, 2002b). This statement is only correct with the caveat that analysis must also include an examination of the effect of the policy on the number of chemicals reported, or in some other way address the potentially serious issue of truncation bias. EPA has spent considerable resources developing a series of risk-based weights for the TRI data in an effort to enhance the validity of these data as a measure of environmental performance. Based on the results presented in this paper, it is worth asking whether a similar effort on reducing truncation bias is warranted.

Bibliography

Arora, Seema and Timonth N. Cason (1998), "Do Community Characteristics Influence Environmental Outcomes? Evidence from the Toxics Release Inventory," *Journal of Applied Economics* 1(2): 413-53.

Degeorge, Francois, Jayendu Patel, and Richard Zeckhauser (1999), "Earnings Management to Exceed Thresholds," *The Journal of Business* 72(1): 1-33.

Ford, B.L. (1983) "An Overview of Hot-Deck Procedures," in: Madow, W.G., I. Olin and D. B. Rubin, eds., *Incomplete Data in Sample Surveys* (New York, New York: Academic Press): 185-207.

Fung, Archon and Dara O'Rourke (2000), "Reinventing Environmental Regulation from the Grassroots Up: Explaining and Expanding the Success of the Toxics Release Inventory, *Environmental Management* 25(2): 115-127.

Graham, Mary and Catherine Miller (2001) "Disclosure of Toxic Releases in the United States," *Environment* 43(8): 9-20.

Grant, Don and Andrew W. Jones (2003), "Are Subsidiaries More Prone to Pollute? New Evidence from the EPA's Toxics Release Inventory," *Social Science Quarterly* 84(1): 162-73

Helland, Eric and Andrew B. Whitford (2003), "Pollution Incidence and Political Jurisdiction: Evidence from TRI," *Journal of Environmental Economics and Management* 46 (3): 403-24.

Karkkainen, Bradley (2001), "Information as Environmental Regulation: TRI, Performance Benchmarking, Precursors to a New Paradigm?," *Georgetown Law Journal*, 89: 257-370.

King, Andrew A. and Michael J. Lenox (2000), "Industry Self-Regulation Without Sanctions: The Chemical Industry's Responsible Care Program, " *Academy of Management Journal* 3(4): 689-716.

Khanna, Madhu and Wilman R. Q. Anton (2002), "EPA's Voluntary 33/50 Program: Impact on Toxic Releases and Economic Performance of Firms," *Journal of Environmental Economics and Management* 37(1):1-25.

Little, R. J. A. and Rubin D.B. (1987), *Statistical Analysis with Missing Data* (New York, New York: John Wiley Press).

Massachusetts Department of Environmental Protection, Bureau of Waste Prevention (2003), *Toxics Use Reporting Instruction* (Boston, Massachusetts: Massachusetts Department of Environmental Protection)

National Research Council, National Academy of Engineering (1999), *Industrial Environmental Performance Metrics: Challenges and Opportunities* (Washington, D.C.: National Academy Press).

White, A. and D. Zinkl (1997), *Green Metrics: A Status Report on Standardized Corporate Environmental Reporting* (Boston, MA: Tellus Institute).

Wooldridge, Jeffrey M. (2002), *Econometric Analysis of Cross Section and Panel Data* (Cambridge, Massachusetts: The MIT Press).

Wolf, Sidney M. (1996), "Fear and Loathing About the Public Right To Know: The Surprising Success of the Emergency Planning and Community Right-To-Know Act," *Journal of Land Use and Environmental Law* 11(2): 217-324.

United States Environmental Protection Agency, Office of Pollution Prevention and Toxics (2004), *EPA's Risk-Screening Environmental Indicators (RSEI) Chronic Human Health Methodology* (Washington, DC: United States Environmental Protection Agency).

United States Environmental Protection Agency, Office of Environmental Information (2003) *How are the Toxics Release Inventory Data Use? – Government, Business, Academic and Citizen Uses* (Washington, DC: United States Environmental Protection Agency, Document number EPA 260-R-002-004).

United States Environmental Protection Agency, Office of Environmental Information (2002a) *2001 Toxics Release Inventory, Executive Summary* (Washington, DC: United States Environmental Protection Agency, Document number EPA 2060-S-03-001).

United States Environmental Protection Agency, Office of Environmental Information (2002b) *Toxics Release Inventory (TRI) Factors to Consider When Using TRI Data* (Washington, DC: United States Environmental Protection Agency, Document number EPA 260-F-02-017).

United States Environmental Protection Agency, (January 17, 2001), "Lead and Lead Compounds; Lowering of Reporting Thresholds; Community Right-to- Know Toxic Chemical Release Reporting; Final Rule" *Federal Register 40 CFR Part 372*, 66(11): 4499-4547.

United States Environmental Protection Agency (2000), *Emergency Planning and Community Right-to-Know Act Section 313 Reporting Guidance for the Textile Processing Industry* (Washington, D.C.: United States Environmental Protection Agency, Document number EPA 745-B-00-008).

United States Environmental Protection Agency (1992), *Guidance for Data Useability in Risk Assessment (Part A)* (Washington, D.C.: United States Environmental Protection Agency, Document number 9285.7-09A): 31-34.

United States Government Accounting Office (1991), *Toxic Chemicals: EPA's Toxic Release Inventory in Useful but Can Be Improved* (Washington, D.C.: United States General Accounting Office, Document number GAO/RCED-91-121).

U.S. Public Interest Research Group (1998), Troubled Waters: A Report on Toxic Releases Into America's Waterways (Washington, D.C.: U.S. Public Interest Research Group).