J. Benefit Cost Anal. 2018; 9(2):285–304
doi:10.1017/bca.2017.17
© Society for Benefit-Cost Analysis, 2017

Richard Morgenstern

Retrospective Analysis of U.S. Federal Environmental Regulation

Abstract: Retrospective, or ex post, analysis of U.S. federal regulation aims to rigorously document regulatory outcomes using cost, benefit, and distributional metrics. This paper presents nine new case studies involving a total of 34 comparisons of ex ante and ex post estimates from a diverse group of environmentally oriented rules. Despite the potential for selection bias and other limitations of the case study approach, the results suggest a slight tendency to overestimate both costs and benefits (or effectiveness) of regulation. This paper considers various analytic issues relevant to developing credible baselines for comparison, and offers policy lessons regarding the design of emissions trading programs along with approaches for incorporating uncertainty into both preregulatory studies and policy designs. Recommendations to facilitate and support future retrospective analyses are also presented.

Keywords: benefit-cost analysis; regulatory review; retrospective analysis.

JEL classifications: D61; H1; L51.

1 Introduction

Retrospective, or ex post, analysis of U.S. federal environmental regulation aims to rigorously document the extent to which key policy objectives have been attained, at what cost, and with what distributional outcomes. Such analyses can help identify successful regulatory approaches, reform poorly performing rules, and inform the design of new ones. Although prospective, ex ante regulatory impact analyses (RIAs) are now routine for major new rules, retrospective analyses (RAs) are uncommon. We focus on analyses that measure the outcomes of regulation rather than efforts that simply update assumptions in ex ante studies.[1]

RAs have the advantage of looking at what actually happened, whereas RIAs are developed at the "point when the least is known and any analysis must rest on

Richard Morgenstern: Resources for the Future, 1616 P St., NW, Washington, District of Columbia, 20036, United States, e-mail: morgenst@rff.org

1 Natural experiments are also of interest, although they are generally difficult to carry out in the area of environmental protection where there is a great reluctance to discriminate among population groups.

Downloaded from https://www.cambridge.org/core. IP address: 207.241.233.138, on 19 Mar 2021 at 15:03:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/bca.2017.17

many unverifiable and potentially controversial assumptions" (Greenstone, 2009, 113). RAs also differ from the qualitative look-backs periodically conducted by executive branch agencies, which typically focus on paperwork reduction and "squeaky-wheel" concerns about specific rules (Aldy, 2014).

Despite encouragement from a growing number of academic experts and bipartisan groups, an Obama-era executive order (E.O. 13610) promoting RA by federal agencies, and recent Trump Administration regulatory reform initiatives, there are still many impediments to bringing rigorous performance assessment into the mainstream of federal environmental policy. As Hahn and Tetlock (2008, 77) noted almost a decade ago, RAs are rarely undertaken

> [begin excerpt]
> . . . because of data and funding limitations, and . . . little interest on the part of most governmental agencies . . . [When they are conducted], academics may select biased samples of regulations – for example, regulations where there is likely to be a publishable finding or applications that have a novel element . . . [And] . . . results from regulatory analyses could differ for several reasons including the author, data, model, key assumptions, and source of funding.
> [end excerpt]

Some question whether federal regulatory agencies could ever play "moneyball," i.e., completely base decisions on evidence and data (Bridgeland & Orszag, 2013). There are also the ever-present conceptual and technical challenges involved in estimating a counterfactual. That is, in the absence of a given regulation would environmental conditions have continued, worsened, or perhaps improved because of technological advances, economic circumstances, or other regulatory or voluntary initiatives? Although RIAs also require development of baseline scenarios, these must be revisited in RAs to assess whether external factors have significantly changed the costs and/or the benefits/effectiveness of regulation. It is no exaggeration to say that developing a credible counterfactual is the most demanding aspect of an RA (Kopits, McGartland, Morgan, Pasurka, Shadbegian, Simon & Wolverton, 2014).

This five-part paper presents key results of the Regulatory Performance Initiative (RPI) conducted by Resources for the Future (RFF), in conjunction with several university scholars, to examine retrospectively the performance of a broad group of environmentally oriented rules issued by the U.S. Environmental Protection Agency (EPA) and other federal agencies, plus certain delegated state authorities. In total, the RPI examined 34 regulatory outcomes across 24 regulations, employing alternative methods to establish counterfactual baselines, some based on complex models or unique data sets. In other cases, well-defined treatment and comparison groups were used to identify regulatory impacts. To preview some

Downloaded from https://www.cambridge.org/core. IP address: 207.241.233.138, on 19 Mar 2021 at 15:03:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/bca.2017.17

of the findings, we see a slight tendency to overestimate *both* benefits and costs, although this is by no means a uniform result. We also present a series of policy and methodology lessons learned in the RPI.

Following this introduction, Section 2 presents background information on RA. Section 3 introduces the case studies with a focus on the alternative approaches used to develop counterfactuals. Section 4 reviews the findings regarding the accuracy of preregulatory RIAs and presents salient policy and methodological lessons from the RPI. Section 5 offers concluding comments along with recommendation on how to enhance future RAs. This review is not intended to substitute for the many important issues raised in the individual RPI papers, including some nuances in interpretation that may not be fully captured here. Rather, the emphasis is on broad-scale observations from the group of studies, and possible future paths for RA.[2]

2 Background

It is tempting to equate ex ante and ex post estimates, respectively, with "forecasts" and "actuals." However, information on the ex post situation is often imperfect, while the quality of the ex ante estimates is limited by various uncertainties about (1) what firms or other regulated entities are currently doing; (2) what they will do in response to the regulation; and (3) what they would have done without the regulation. The first of these uncertainties is, in principle, knowable ex ante but in practice generally not known with much precision. The second and third items – and the resulting costs and benefits – are hypothetical and based on economic, engineering, and process-analysis models, along with discussions with industry experts; and sometimes, analogies from other industries. Problems are compounded when production and control technologies are changing rapidly.

Estimating costs and benefits in an RA involves dealing with the same issues as an RIA, albeit with more and presumably better data. The process of implementation and enforcement generates new information, not only about the responses of firms to the regulation but also about the situation prior to implementation, and the effect of external factors not anticipated in the RIA. Thus, notwithstanding any author-related differences, the ex post estimate will likely be much closer to the "truth."[3]

2 The full set of papers, including more complete descriptions of the underlying regulations, is available in discussion paper format on the RFF website. Most of the papers are under review for journal publication. See: http://www.rff.org/research/collection/regulatory-performance-initiative-rff

3 Of course, since different people with different biases, abilities, and available resources perform ex ante and ex post analyses there is no absolute guarantee of increased accuracy in the ex post studies.

Downloaded from https://www.cambridge.org/core. IP address: 207.241.233.138, on 19 Mar 2021 at 15:03:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/bca.2017.17

The relatively limited number of RAs that do exist often focus on inspections, audits, or other input measures. Although certain broad-scale, publicly available databases allow direct assessment of effectiveness, e.g., reduced air emissions and water discharges, relatively few studies involve welfare-based metrics such as net social benefits. Confidentiality and related concerns impede direct observation of actual costs.[4] Emissions trading and other economic incentive mechanisms are an exception, since the transparency of market prices and the tracking of changes in emissions can often provide valuable data for analysis.

The need to isolate the impact of the regulation from other factors that may affect emissions, environmental outcomes or costs is a major challenge for RA. Some RIAs and RAs extrapolate from historical data in constructing counterfactuals even though independent factors could affect both baseline behavior and regulatory outcomes. More sophisticated studies attempt to compare the performance of regulated versus unregulated (or differentially regulated) entities. Sometimes propensity score matching or other techniques are used to develop comparison groups. Yet even these comparisons can be problematic if the different firms participate in the same market, since the behavior of unregulated entities may be indirectly affected by the regulation – thereby violating the stable unit treatment value assumption.[5] Although it may raise other concerns, an alternative is to draw a comparison group from unregulated plants in other industries that have no plausible impacts on the regulated units. Difference-in-differences or other statistical models can help quantify regulatory outcomes. Existing (peer-reviewed) models or unique data sets can also be used to establish credible baselines.

Overall, no single method for RA dominates. The ability to conduct RAs depends on multiple, sometimes idiosyncratic factors, including the design of the rule, the available data on regulated, unregulated, and differentially regulated entities, and the existence of established models. The challenge for the researcher is to line up the major cost, benefit/effectiveness, or distributional metrics of a particular rule with the available data or models that can be used to identify the outcome(s) of the regulation.

4 Examples of high-quality cost- or benefit-oriented studies include Seong and Mendeloff (2004), Thompson, Sequi-Gomez and Graham (2002), Harrington, Morgenstern, Shih and Bell (2012), Dale, Camille, Michael, McMahon and Fujita (2009), Fowlie and Perloff (2013), and Deschenes, Greenstone and Shapiro (2016). Several survey papers have also attempted to draw broader generalizations from the limited literature on RAs of environmental, health, and safety regulation, such as Harrington, Morgenstern and Nelson (2000), Harrington (2006), OMB (2006), Simpson (2014) and Lutter (2012).

5 Another issue, explored by Elrod and Malik (2017), is the possibility of changes in the product mix of both regulated and unregulated firms in response to the new rule. Such changes could, in principle, confound attempts to estimate costs and benefits on both ex ante and ex post bases.

Downloaded from https://www.cambridge.org/core. IP address: 207.241.233.138, on 19 Mar 2021 at 15:03:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/bca.2017.17

3 Review of the retrospective analysis case studies

In an attempt to reduce selection bias, the initial strategy for the RPI was to examine entire classes of regulations, such as all air toxics rules establishing maximum achievable control (MACT) levels issued under Section 112 of the 1990 Clean Air Act Amendments, or all "best available technology" effluent guidelines issued under authority of the Clean Water Act of 1972. Unfortunately, this proved to be infeasible due to unanticipated data problems, although several RPI studies come close to this "class" approach.

The design ultimately adopted for case selection was principally based on two different competitive solicitations, one focused on RFF staff and the other aimed at outside researchers, each resulting in the selection of four cases.[6] The outside competition drew a particularly large response – almost two dozen high-quality proposals from university scholars. A ninth case was added later, based on the discovery of an ongoing research effort that closely matched the requirements of the RPI. Selection of all the cases, as stated in the request for proposals posted on the RFF website, was based on "the originality and scientific merit of the proposed research, compatibility with the goals of the RPI, and the demonstrated research competence of the applicants." The final group of RAs included in the RPI includes both market-based and traditional regulations issued by multiple federal and state agencies. Outcome metrics are based on costs (or prices), behavioral change, and environmental benefits (both physical and monetized). Some of the cases have a clear health focus, while others are oriented to ecological or energy issues. Several cases involve the development of new technologies.

3.1 The case studies

To orient the reader to the case studies and to provide context for the summary observations presented below, we begin with a brief description of the case studies

3.1.1 Acid rain program (EPA)

Chan, Chupp, Cropper and Muller (2017) studied Phase II of the EPA Acid Rain Program (ARP). To quantify cost savings, the authors compared compliance costs

6 Originally, Fraas and Egorenkov prepared separate papers on the control of industrial air toxics and the control of industrial water toxics. These were posted on the RFF website as DPs 15–23 and 15–41. Recently, the two studies were integrated into a single paper, now forthcoming in the *Journal of Benefit-Cost Analysis.*

Downloaded from https://www.cambridge.org/core. IP address: 207.241.233.138, on 19 Mar 2021 at 15:03:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/bca.2017.17

under the ARP for coal fired electricity units not subject to New Source Performance Standards (NSPS) with the costs for the same units under a uniform performance standard that achieves an identical aggregate emissions target for the single year 2002. Overall, Chan et. al. find that more scrubbers than necessary were actually installed. In total, almost one fourth of the generating units were not using the least-cost compliance method, a result the authors attribute to state-level institutional and political constraints not considered in the ex ante analyses.

Chan et al. (2017) also examine the effect of emissions trading on health damages and net economic benefits. The key to this analysis is the existence of a stochastic integrated assessment model, AP2, which the authors use to develop base case emissions projections. They estimate the health damages associated with observed emissions from the non-NSPS coal fired generating units regulated under the ARP and compare them with damages from a no-trade counterfactual in which each unit emits sulfur dioxide (SO_2) at a rate equal to its 2002 allocation of permits, plus any drawdown of its allowance bank. Annual damages under the no-trade scenario are several billion dollars lower than under the ARP for the single year studied, reflecting the transfer of allowances from units west of the Mississippi River to those with higher exposed populations in the eastern United States.

3.1.2 Renewable fuels standard (EPA)

EPA's Renewable Fuel Standard (RFS2) was studied by Lade, Cynthia Lin and Smith (2015). Using the observed prices for biofuel credits (known as RINs or renewable identification numbers) as an upper-bound measure of the RFS2-induced increase in wholesale gasoline and diesel prices, the authors find these prices exceeded EPA's ex ante cost estimates by as much as 1,400% in 2013 and again in 2014, although they subsequently dropped off in response to agency announcements to reduce mandated biofuel quantities. Even though the RIN prices likely reflected some speculative fervor, mean or median measures reveal regulatory costs well in excess of EPA estimates. The authors attribute EPA's cost underestimates to (1) the failure to consider delays in development of the advanced biofuel industries as well as the alternative fueling infrastructure; and (2) the failure to account for the uncertainty inherent in predicting the demand for and the relative prices of both oil and biofuels.

3.1.3 Industrial air toxics (EPA)

Two related papers evaluated EPA's regulation of industrial air and water toxics. Gray and Shadbegian (2015) focus on the agency's Cluster Rule, which

Downloaded from https://www.cambridge.org/core. IP address: 207.241.233.138, on 19 Mar 2021 at 15:03:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/bca.2017.17

simultaneously controlled air and water toxics releases from pulp and paper mills. Using as a comparison group a set of pulp and paper mills not covered by the regulations, and targeting chloroform as a marker of toxic releases, they find significant reductions, nearly identical to the ex ante prediction. Other outcomes are mixed, as Gray and Shadbegian see reductions in benzene and other volatile organic compounds (VOCs) of about half of EPA's ex ante prediction, and not always statistically significant. They find no significant reduction in particulate matter (PM_{10}) emissions, contrary to the ex ante prediction. Using a broader set of toxics and a group of unregulated plants in other industries as a comparison, Fraas and Egorenkov (forthcoming *Journal of Benefit-Cost Analysis*) also find mixed results, with some observed outcomes meeting and others falling (far) short of EPA's ex ante projections.[7] The key to conducting these RAs is the availability of emissions data from the Toxics Release Inventory (TRI), a large, publicly available database on toxic releases from all plants meeting a minimum size threshold, including both regulated and unregulated facilities.

3.1.4 Salmonella enteriditis contamination of eggs (Food and Drug Administration)

A Food and Drug Administration (FDA) rule requiring egg farmers to adopt control measures to limit possible contamination by *Salmonella enteriditis* (SE), a pathogen occurring almost exclusively in eggs, was studied by Lutter (2015). The author's identification strategy involves designating as a comparison group the observed cases of *Salmonella* serotypes other than SE that occur commonly in poultry (but not eggs) and are influenced by the same risk factors.[8] Using the FoodNet database on human cases of foodborne illness[9] and a difference-in difference model to compare the incidence of illness from SE and other pathogens, before and after implementation of the FDA rule, Lutter is unable to reject the null hypothesis that the rule had no effect at all.

7 Interestingly, in three other industries examined (petroleum refining, pharmaceuticals, and wood furniture) Fraas and Egorenkov (forthcoming) also find little or no toxics reductions in response to MACT rules.

8 As Lutter (2015, 3) notes, "The FDA reported that for outbreaks where a vehicle of transmission of SE was identified, 81 percent of outbreaks and 79 percent of illnesses . . . were attributed to eggs . . . infection controls in egg farms are likely to affect predominantly, and perhaps exclusively, the incidence of disease from SE but not from non-SE."

9 FoodNet is maintained by the U.S. Centers for Disease Control and Prevention, commonly known as CDC.

Downloaded from https://www.cambridge.org/core. IP address: 207.241.233.138, on 19 Mar 2021 at 15:03:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/bca.2017.17

3.1.5 Critical habitat designation under the endangered species act (department of interior)

Nelson, Withey, Pennington and Lawler (2017) examined difficult-to-study changes in land cover induced by the Department of Interior's designation of critical habitat (CH) on individual parcels under the Endangered Species Act (ESA).[10] The authors use propensity score matching methods to compare development patterns in the CH areas with patterns in areas subject to ESA regulation (under a prior listing decision) but not designated as CH. Despite the expectation of more early development in CH-designated areas, these effects are not strong enough to differentiate the average rates of developed or agricultural land change in CH areas versus those areas subject to ESA regulations without CH designation. At the same time, the authors do find that CH designation has increased landowners' uncertainty and that conversion to developed and agricultural use in CH areas does require an increase in the rate of return. Acknowledging the potential for selection bias in their sample, and the limited number of high-value areas examined, Nelson et al. recommend that future CH designations continue to highlight potential land management impacts, especially for higher value land.[11]

3.1.6 Minimum efficiency performance for appliances (Department of Energy)

Taylor, Spurlock and Yang (2015) assessed minimum efficiency performance standards (MEPS) set by the Department of Energy (DOE) for five appliance types – room air conditioners, refrigerators–freezers, dishwashers, clothes washers, and clothes dryers. The authors compare the observed ex post appliance prices and performance with the ex ante engineering-cost analyses developed in the RIAs. Overall, the authors find that all five MEPS result in the purchase of energy-efficient appliances at lower-than-expected prices and, with the exception of front-loading clothes washers, the various quality metrics suggest consumers are not worse off

10 This analysis by Nelson et al. (2017) jointly considered the impacts of multiple critical habitat regulations, not a single rule.

11 Specifically, they "cannot conclude that all opportunity costs of CH designation above and beyond the opportunity costs of ESA listing are minimal. However, it is very unlikely that such costs will typically be very high" (Nelson et al., 2017). The authors also note that their results may provide some support for the existence of a "shoot, shovel, and shut up" effect or a preemptive development dynamic, wherein newly established CH areas may have been developed more quickly pre-CH designation than otherwise, possibly to reduce landowners' uncertainty and avoid future conflict with the CH regulations.

Downloaded from https://www.cambridge.org/core. IP address: 207.241.233.138, on 19 Mar 2021 at 15:03:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/bca.2017.17

in nonregulated aspects of product performance. Indeed, consumers appear to be better off in several important product quality dimensions.[12]

Unlike the other RPI studies, the Taylor et al. approach does not directly control for unforeseen technology advances and essentially attributes to the regulation all gains beyond those estimated by the DOE models. If, as the authors argue, those models are based on overly conservative definitions of "technical feasibility," they will tend to understate performance of the regulated appliances and overstate costs (prices). Further, their *Consumer Reports*-based quality metrics may not be an ideal source for discovery of operational problems, since it relies so heavily on laboratory test results rather than in-use performance. Notwithstanding these limitations, the study is based on a broad set of price and performance metrics and does demonstrate a method for conducting RA on a set of difficult-to-evaluate rules.

3.1.7 Clean water state revolving fund (EPA funded, state operated)

Harrington and Malinovskaya (2015) evaluated EPA's Clean Water State Revolving Fund (CWSRF), a federally supported state-level program that provides funding to states for publicly owned treatment works (POTWs). They focused on both the effectiveness and the efficiency of the federally subsidized loan program in improving water quality over the eight-year period 2007–2014. The key to the analysis is the authors' ability to assemble a comprehensive data set on hundreds of individual facilities in four states (Iowa, Indiana, Maryland, and Texas) from multiple federal and state information sources, including the Clean Watershed Needs Survey, EPA's Clean Water Benefits Reporting System, and EPA's Integrated Compliance Information System, along with POTW-specific information on effluent quality at the beginning and end of the study period.

Using as a comparison group POTWs in the same state that requested but were not awarded loans, the authors find that the typical loan-supported plant significantly improved the quality of effluent discharges of both biochemical oxygen demand and organic nitrogen in all four states. They also found, however, that plants receiving loans tended to have cleaner pre-2007 effluent discharges than facilities

12 As evidence of the quality issues with front-loading washers, a nationwide class-action settlement agreement overseen by the Federal District Court of Northern Ohio mandated compensation to owners of front-loading washers produced in 2001–2010 for mold-related damages. This suggests that Taylor et al. may not have adequately accounted for operation and maintenance costs associated with the regulation-induced design changes – in this case the inconvenience and cost of cleaning the washer after use, the energy and water costs of additional hot water flushes, and the inconvenience and cost of service calls. See http://www.walb.com/story/32328197/customers-who-complain-of-mold-in-front-load-washers-could-get-payout-in-lawsuit-settlement.

Downloaded from https://www.cambridge.org/core. IP address: 207.241.233.138, on 19 Mar 2021 at 15:03:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/bca.2017.17

not receiving loans prior to the study period. Thus, although loans were effective in improving water quality, it appears the funded plants were already among the cleaner ones in terms of pollutant discharge concentrations. The authors were not able to assess the demographic characteristics of the communities served by these cleaner POTWs or the economic benefits of cleaning up water discharges at the individual facilities. These study results clearly raise concerns about the efficiency of the program in achieving maximum water quality improvements with the available funds.

4 Summary of case study results

Table 1 summarizes all the ex ante and ex post comparisons of the federal rules studied in the RPI, i.e., excluding the state-level results from the CWSRF, which are shown separately in Table 2. For simplification, these comparisons are labeled as "accurate" if an RIA estimate falls roughly within +/−25% of the ex post observation.[13] Outside these bounds, the results are deemed "underestimates" or "overestimates." Taking all the Table 1 results together, we see that of the 8 regulations for which we are able to make ex ante and ex post cost comparisons, 6 involve overestimates, 2 are underestimated, and none is deemed accurate. At the same time, some of the classifications seem more problematic than others. If one removes the MEPS for clothes washers on the basis of the noted issues, then the cost of only five rules would be classified as overestimated.

Of the 18 regulatory requirements for which we are able to compare benefits or effectiveness estimates on an ex ante and ex post basis, we find that 10 involve overestimates, 6 are underestimates, and 2 are assessed as relatively accurate. Removing the MEPS for clothes washers reduces the total to 9 overestimates of benefits/effectiveness.

Looking at the federally funded but state-operated CWSRF as a separate category (Table 2), we seen an additional 4 cases of cost overestimation and 4 more cases of accurate benefits estimation. Combining the federal and state estimates, we see a total of 10 cost overestimates, 2 underestimates, and 0 accurate ones. In terms of benefits/effectiveness, we see a total of 10 overestimates, 6 underestimates, and 6 accurate estimates.

Binning the results of all the federal and state actions into the three categories of under-, over-, and accurate estimates indicates a slight tendency for overestimation of both the costs and the benefits/effectiveness of the regulations, a conclusion

13 The convention of defining accuracy as +/−25% was used in Harrington et al. (2000) and subsequently adopted by other authors, including OMB (2006).

Downloaded from https://www.cambridge.org/core. IP address: 207.241.233.138, on 19 Mar 2021 at 15:03:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/bca.2017.17

Table 1 Summary of Ex Ante–Ex Post Comparisons of Federally Issued Rules.

RULE	COSTS	BENEFITS OR EFFECTIVENESS	METHODS
ACID RAIN PROGRAM PHASE II TRADING	Underestimated	Overestimated	Model estimation, simulation
CLUSTER RULE			
Air: chloroform		Accurate	Comparison with unregulated sources, diff-in-diff, propensity score matching
Water: chloroform		Accurate	
Air: VOCs		Overestimated	
Air: toxics		Overestimated	
Air: PM_{10}		Overestimated	
Pulp and paper effluent guideline		Overestimated	Comparison with unregulated sources, diff-in-diff estimation, time series analysis (OLS)
AIR TOXICS			
Printing and publishing		Underestimated	Comparison with unregulated sources, diff-in-diff estimation, time series analysis (OLS)
Pulp and paper		Overestimated	
Petroleum refining		Overestimated	
Pharmaceuticals		Overestimated	
Wood furniture		Overestimated	
RENEWABLE FUEL STANDARD	Underestimated		Use of credit prices as proxies for costs
EGG RULE *(SALMONELLA)*		Overestimated	Comparison with unregulated entities, diff-in-diff estimation
CRITICAL HABITAT DESIGNATION UNDER ENDANGERED SPECIES ACT	Overestimated		Comparison with unregulated land, propensity score matching
MINIMUM EFFICIENCY PERFORMANCE STANDARDS			
Room air conditioners	Overestimated	Underestimated	Direct comparison with RIA estimates, OLS on multiple data sets
Refrigerators	Overestimated	Underestimated	
Dishwashers	Overestimated	Underestimated	
Clothes washers	Overestimated	Underestimated	
Clothes dryers	Overestimated	Underestimated	

Downloaded from https://www.cambridge.org/core. IP address: 207.241.233.138, on 19 Mar 2021 at 15:03:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/bca.2017.17

Table 2 Summary of Ex Ante–Ex Post Comparisons for Clean Water State Revolving Fund.

RULE	COSTS	BENEFITS OR EFFECTIVENESS	METHODS
CLEAN WATER STATE REVOLVING FUND			
Indiana	Overestimated	Accurate	Comparison with uncontrolled sources, event study/time series analysis
Iowa	Overestimated	Accurate	
Maryland	Overestimated	Accurate	
Texas	Overestimated	Accurate	

supported by a number of earlier studies.[14] At the same time, there is considerable variation across the cases. In light of the relatively small number of cases and the diversity of methods and baselines in the underlying studies it is not possible to draw statistical inferences from these results.

5 Lessons learned

The results of a single project such as the RPI, even one based on a diverse set of cases, cannot possibly fill the enormous knowledge gaps about RA of federal environmental regulation. Admonitions against generalizing from a limited number of case studies are well known and clearly applicable here. Still, a number of observations emerge from the individual cases that merit emphasis.

5.1 RA methods

The choice of a method for identifying baseline conditions against which the regulation can be compared is critical to the design of an RA. As noted, the preferred method depends on multiple factors, including the design of the rule, the available data on regulated, unregulated, and differentially regulated entities, the existence of peer-reviewed models, and other factors. Four methods were used in the RPI to measure baselines: identification of a specific comparison group, use of an established model, observed prices of tradable allowances, and direct comparison with the baseline used in the RIA.

14 For example, Harrington et al. (2000), Harrington (2006), and Simpson (2014).

Downloaded from https://www.cambridge.org/core. IP address: 207.241.233.138, on 19 Mar 2021 at 15:03:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/bca.2017.17

Among the RPI cases, the most popular method involved development of a comparison group against which to measure the regulated (treated) entities. The authors of the three industrial toxics studies relied on the Toxics Release Inventory, which covers both regulated and unregulated facilities, to identify a suitable comparison group. The author of the food safety study was also able to use a single data set, FoodNet, which reports food-based illness for the entire population, to develop a comparison group. In contrast, the authors of the Clean Water State Revolving Fund study had to assemble their own data set from more than half a dozen different federal and state sources that covered all public water treatment plants in the four states studied. The authors of the critical habitat designation study used propensity score matching to select a comparison group from a large government-sponsored mapping system. For all these cases, difference-in-differences or other statistical models were used to estimate the effect of the regulations on the treated versus the comparison groups.

The authors of the Acid Rain Program study relied on AP2, a well-developed model, to establish a counterfactual for the analysis of the benefits, and they estimated their own logit model based on ex post data for the cost portion of the RA. Post-regulatory information was integrated into these models to simulate emissions trading versus a uniform performance standard. The authors of the Renewable Fuel Standard paper based their RA on observed biofuel credit prices, which they treated as an upper-bound estimate of regulatory costs. The Minimum Efficiency Performance Standards study authors compared observed product prices and quality metrics, painstakingly extracted from *Consumer Reports*, product manuals, and other sources, with the baseline estimates developed in the RIAs.

Clearly, the counterfactual no-regulation baseline can never be observed for regulated entities, although in some instances it can be inferred where a comparison group of similar, unregulated, or differentially regulated firms is available. As noted, the key is to line up critical cost, effectiveness, or distributional metrics of a particular rule with the available data or models that can be used to identify the regulatory outcome(s).

5.2 RIA shortcomings

At the risk of oversimplification, our analysis reveals several shortcomings in the RIAs reviewed for the RPI, including errors in forecasting the development and deployment of new technologies, what might be labeled basic analytic mistakes, and a general failure to account for the often-heterogeneous nature of regulated

Downloaded from https://www.cambridge.org/core. IP address: 207.241.233.138, on 19 Mar 2021 at 15:03:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/bca.2017.17

entities. Although none of these shortcomings is novel, their continued appearance in million-dollar-plus studies is disheartening.

Errors in forecasting technology development are an important issue in several case studies. Some might suggest a tendency to conservatism in forecasting the cost and performance characteristics of new technologies. In fact, the errors go both ways in the RPI. The Minimum Efficiency Performance Standards program appears to have overestimated costs and underestimated effectiveness, while the Renewable Fuel Standard program underestimated costs.

In terms of basic analytic issues, the authors of the Renewable Fuel Standard study note that beyond the technology forecasting errors, both the RIA and the basic regulatory design failed to consider the potential variability in oil and biofuel prices or delays in infrastructure investments. There was also a single-minded focus on the long run without regard for near-term "start-up" problems. Especially for transformative regulations of this type, the authors suggest that future RIAs should assess a variety of scenarios and not focus on a single best estimate.

A further theme of the cases relates to the often ignored heterogeneity of impacts along several dimensions. The authors of the paper on critical habitat designation note the possibility of adverse effects of designation for high-value land, even though these effects do not show up in their statistical analysis of average effects across all land types. The Acid Rain Program case study authors find that failure to account for differences in population density surrounding generating plants can result in perverse impacts from emissions trading, at least in the single year studied (2002). Similarly, the Cluster Rule study authors point to the large differences in compliance costs among pulp and paper plants that were largely ignored in the actual rule design. The author of the study on the egg rule for *Salmonella* observes quite different levels of management sophistication across egg farms. All are examples of potential opportunities to subcategorize regulated entities to achieve greater cost-effectiveness.

We also note the importance of institutional constraints that are sometimes overlooked in RIAs. This is most apparent in the Acid Rain Program, where state-level political factors limited the attainment of the modeled least-cost solution. Permitting delays and other institutional constraints may also have been a factor in the cost underestimates of the Renewable Fuel Standard regulation, since the production ramp-ups were constrained in various ways not anticipated in the preregulatory analysis. Of course, there is no magic formula to determine how much additional analysis is enough. However, there are reasons to believe that the instances identified here would pass reasonable value-of-information criteria.

Finally, at the risk of emphasizing a small point, several of the RPI studies indicate that regulatory agencies relied on outdated baseline information on firms'

Downloaded from https://www.cambridge.org/core. IP address: 207.241.233.138, on 19 Mar 2021 at 15:03:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/bca.2017.17

compliance efforts, which, in turn, led to significant overestimation of both the expected costs and the benefits/effectiveness of new regulations. In at least one case, EPA failed to account for early reductions taken by firms in response to state requirements or federal voluntary programs. To avoid such errors in the future, greater effort is needed on the part of regulatory agencies to obtain more up-to-date information on emissions and on abatement equipment in place.

5.3 Policy lessons

Policy lessons that may help shape future regulations are, potentially the most valuable type of learning from RAs. Although it is tempting to dig deeply into the individual cases to extract specific policy lessons, recognizing the limits of the case study approach, we choose instead to focus on a few broad thematic issues for future policy development.

We highlight the findings of the Chan et al. (2017) study of Phase II of the SO_2 emissions trading program, which point to an underestimate of the costs compared with the modeled least-cost solution and the estimated negative benefits associated with trading for the single year studied (2002), as emissions moved from areas of low to high population density in the eastern United States.[15] The policy remedies here are straightforward. Chan et al. recommend that regulators treat the modeled least-cost solution as the lower bound of costs rather than the best estimate. Even though this begs the question of how to model the most likely case, it calls out the need to carefully think through the amount and nature of trading likely to occur under a given program design rather than make the blanket assumption of perfect markets. On the benefits side, the remedies are also well known: when there is reason to expect adverse spatial effects, non–1:1 trading ratios or zoned trading should be considered. Here again, the need is for regulators to assess the probability that unrestricted trading will generate "hot spots" and/or significantly reduce overall program benefits. Far from a condemnation of emissions trading, this is a call for thoughtful assessment by regulators of the specific characteristics of the market created by the rule and the potential for inefficiencies of one type or another.

Big spikes in observed biofuel credit prices, a measure of compliance cost for the Renewable Fuel Standard, ultimately led EPA to relax program requirements, albeit not without a lag that was costly to consumers and oil refiners, to the benefit of ethanol and biodiesel producers. The need for these changes raises the obvious question of regulatory design: even with the stipulated congressional mandates,

15 Since the authors do not present evidence about other years, it is not possible to determine whether the trading program had a positive or negative effect on net benefits overall.

Downloaded from https://www.cambridge.org/core. IP address: 207.241.233.138, on 19 Mar 2021 at 15:03:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/bca.2017.17

how could EPA have built greater flexibility into the program in the face of considerable uncertainty about future oil prices and the costs and pace of development of renewable fuels? Although technology-forcing regulation has scored some successes in the past, there needs to be a reasonable plan B if the optimistic scenario does not work out. As Lade et al. (2015) note, uncertainty analysis, or what they call "stress tests," can help identify potential vulnerabilities. Clearly, multiple options exist to address uncertainties, including use of a safety valve or other flexible mechanism that automatically relaxes the requirements in an orderly way under specified market conditions. The lesson here is simple: regulators should be sensitive to the possibility of major miscalculations, especially for transformative programs like the Renewable Fuel Standard, and should build in more flexibility at the outset.

A third relevant issue is drawn from the RA on the Clean Water State Revolving Fund, a federally funded state program that has not been subject to extensive retrospective analysis. We note that this RA required assembling and linking half a dozen diverse data sets that measure loan requests, loans granted, water quality discharges, and other metrics, in each of the four states studied. Given the potential strategic nature of the loan requests, one might argue they are an imperfect measure of costs. At a minimum, however, the study's water quality-related findings call into question the effectiveness of a program long considered a bedrock of municipal water pollution control. Going forward, more analysis is warranted to assess the performance of this federally financed, state-operated program. To facilitate such analysis, the type of datasets so painstakingly assembled by Harrington and Malinovskaya need to be made more accessible to the public.

6 Concluding observations

This review of the RPI findings suggests opportunities for advancing the conduct of RA and enhancing public discussion of the performance of individual regulations and entire regulatory programs. It also demonstrates the inherent difficulty of conducting RAs on environmental regulations. Clearly, a truly systematic RA of the effects of environmental regulation, based on standard measures of economic efficiency (e.g., cost-effectiveness or net benefits) is a ways off. At the same time, as political conflicts over environmental regulation increase, evidence-based research can play an important role in both prioritizing and legitimizing future regulatory actions.

Cass Sunstein, then administrator of the Office of Information and Regulatory Affairs (OIRA) (in the U.S. Office of Management and Budget, OMB) emphasized the "culture of retrospective review and analysis," (Sunstein, 2011*a*,*b*). In fact, there

Downloaded from https://www.cambridge.org/core. IP address: 207.241.233.138, on 19 Mar 2021 at 15:03:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/bca.2017.17

is an evolving consensus among analysts of all political stripes that, where practicable, future regulations should be designed in ways that facilitate evaluation of their consequences and thus promote retrospective analyses. Writing before the 2016 Presidential election, a bipartisan group of former OIRA administrators recommended that "at the time of initial rulemaking, [agencies] prospectively develop plans to reevaluate their regulatory choices" (Institute of Policy Integrity, 2016, 1). In its early days the Trump Administration has signaled a strong interest in regulatory reform that could clearly benefit from rigorous RA.

Following the recommendations of the Administrative Conference of the United States (2014), recent papers by Cropper, Fraas and Morgenstern (2017), and Fraas and Egorenkov (forthcoming), call for specific steps to formalize requirements for RA of a selective set of rules within the federal government. Arguably, these requirements should stipulate the types of measurable outcomes to be targeted in the RAs, including a relevant comparison group; the associated data requirements for both compliance costs and effectiveness or benefits, possibly coordinating with existing data collection efforts by the U.S. Census Bureau or other relevant agencies; a power calculation or other assessment of a minimum sample size to identify regulatory outcomes; and the period of evaluation as well as the time necessary to complete the RA. The government also needs to stipulate appropriate processes for conducting and peer reviewing these studies, and for the resources to support this work.

OMB could also do more to streamline data collection under the Paperwork Reduction Act. Similar to earlier agency actions to expedite approval for research-related information collection (Sunstein, 2011*a*), OMB could explicitly extend this streamlined process for data collection to RAs of federal regulation.[16]

In terms of future research, four areas seem ripe for additional work:

- Additional case studies of individual rules, ideally drawn randomly from the universe of major regulations, should be conducted with the aim of further demonstrating both the feasibility and the value of such analyses.
- Building on the useful but still quite limited literature comparing the results of ex post RAs to ex ante RIAs, a combined analysis of all available studies should be carried out. This analysis should go beyond the simple measurement of accuracy of the RIAs and systematically examine the causes for over- or underestimation of both costs and benefits of federal rules.
- Recently issued major regulations should be examined to assess the feasibility of conducting RAs on major rules. This assessment should consider how to measure baseline conditions for costs and benefits, including the

16 This idea has also been advanced by Cropper et al. (2017) and Fraas and Egorenkov (forthcoming)

Downloaded from https://www.cambridge.org/core. IP address: 207.241.233.138, on 19 Mar 2021 at 15:03:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/bca.2017.17

availability of publicly available data for doing so. It should also identify the data, methodological, or other impediments to conducting rigorous RA of these rules.

- Based on the new studies and reviews, it would be useful to develop selection criteria for conducting future ex post analyses of federal environmental rules: for example, how many rules should be analyzed, how quickly after implementation, and what are the rule characteristics that warrant selection? It would also be valuable to identify criteria for requiring additional data collection or reporting by regulated entities that might be built into new rules at the outset.

References

Administrative Conference of the United States (2014). Retrospective Review of Agency Rules. Recommendation number 2014–5, adopted December 4, 2014, published December 9, 2014. https://www.acus.gov/research-projects/retrospective-review-agency-rules.

Aldy, Joseph E. (2014). Learning from Experience: An Assessment of the Retrospective Reviews of Agency Rules & the Evidence for Improving the Design & Implementation of Regulatory Policy. *Report to the Administrative Conference of the United States*. http://www.acus.gov/sites/default/files/documents/Aldy%20Retro%20Review%20Draft%2011-17-2014.pdf.

Bridgeland, John & Peter, Orszag (2013). Can Government Play Moneyball? *The Atlantic*, June 19. http://www.theatlantic.com/magazine/archive/2013/07/cangovernment-play-moneyball/309389/.

Chan, H. Ron, Chupp, B. Andrew, Cropper, Maureen L. & Muller, Nicholas Z. (2017). The Impact of Trading on the Costs and Benefits of the Acid Rain Program. *Resources for the Future Discussion Paper 15–25 REV*. Washington, DC.

Cropper, Maureen, Fraas, Arthur & Morgenstern, Richard D. (2017). Looking Backward to Move Regulations Forward. *Science*, *355*(6332), 1375–1376.

Dale, Larry, Camille, Antinori, Michael, McNeil, McMahon, James E. & Fujita, K. Sydney (2009). Retrospective Evaluation of Appliance Price Trends. *Energy Policy*, *37*, 597–605.

Deschenes, Oliver, Greenstone, Michael & Shapiro, Joseph (2016). *Defensive investments & the demand for air quality: evidence from the NOx budget program*. National Bureau of Economic Research, Cambridge, Mass. NBER Working Paper No. w1826.7.

Elrod, Aaron A. & Malik, Arun S. (2017). The Effect of Environmental Regulation on Plant-Level Product Mix. *Journal of Environmental Management*, *83*, 164–184.

Fowlie, Meredith & Perloff, Jeffrey (2013). Distributing pollution rights in cap-and-trade programs in the United States *Review of Economics and Statistics*, *95*(5), 1640–1652.

Fraas, Arthur G. & Egorenkov, Alexander (forthcoming). Retrospective Analyses Are Hard: A Cautionary Tale from EPA's Air Toxics Regulations. *Journal of Benefit Cost Analysis*.

Gray, Wayne B. & Shadbegian, Ronald (2015). Multimedia Pollution Regulation and Environmental Performance: EPA's Cluster Rule. *Resources for the Future Discussion Paper 15–26*. Washington, DC.

Downloaded from https://www.cambridge.org/core. IP address: 207.241.233.138, on 19 Mar 2021 at 15:03:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/bca.2017.17

Greenstone, Michael (2009). Toward a Culture of Persistent Regulatory Experimentation and Evaluation. In Moss David & Cisternino John (Eds.), *New Perspectives on Regulation* (pp. 111–126). Cambridge, MA: The Tobin Project.

Hahn, Robert W. & Tetlock, Paul C. (2008). Has Economic Analysis Improved Regulatory Decisions? *Journal of Economic Perspectives*, *22*(1), 67–84 (Winter).

Harrington, Winston (2006). Grading Estimates of the Benefits and Costs of Federal Regulation. *Resources for the Future Discussion Paper 06–39*. Washington, DC.

Harrington, Winston & Malinovskaya, Anna (2015). Expected vs. Actual Outcomes of Environmental Policies: The Clean Water State Revolving Fund. *Resources for the Future Discussion Paper*. Washington, DC.

Harrington, Winston, Morgenstern, Richard & Nelson, Peter (2000). On the Accuracy of Regulatory Cost Estimates. *Journal of Policy Analysis and Management*, *19*(2), 297–322.

Harrington, Winston, Morgenstern, Richard, Shih, Jhih-Shyang & Bell, Michelle (2012). Did the Clean Air Act Amendments of 1990 Really Improve Air Quality? *Air Quality, Atmosphere and Health*, *5*(4), 353–367.

Institute of Policy Integrity (2016). *Strengthening Regulatory Review: Recommendations for the Trump Administration from Former OIRA Leaders*. New York: New York University; http://policyintegrity.org/publications/detail/strengthening-regulatory-review.

Kopits, Elizabeth, McGartland, Al, Morgan, Cynthia, Pasurka, Carl, Shadbegian, Ron, Simon, Nathalie B., Simpson, David & Wolverton, Ann (2014). Retrospective Cost Analyses of EPA Regulations: A Case Study Approach. *Journal of Benefit Cost Analysis*, *5*(2), 173–192.

Lade, Gabriel E., Cynthia Lin, C.-Y. & Smith, Aaron (2015). Ex Post Costs and Renewable Identification Number (RIN) Prices under the Renewable Fuel Standard. *Resources for the Future Discussion Paper 15–22*. Washington, DC.

Lutter, Randall (2012). *The role of retrospective analysis and review in regulatory policy*. Mercatus Center Working Paper 12–14. Arlington, VA: George Mason University.

Lutter, Randall (2015). How Effective Are Federal Food Safety Regulations? The Case of Eggs and *Salmonella* Enteritidis. *Resources for the Future Discussion Paper 15–24*. Washington, DC.

Nelson, Erik J., Withey, John C., Pennington, Derric & Lawler, Joshua J. (2017). Identifying the Impacts of Critical Habitat Designation on Land Cover Change. *Resource and Energy Economics*, *47*, 89–126.

Office of Management and Budget (OMB) (2006). *Validating Regulatory Analysis: 2005 Report to Congress on the Costs and Benefits of Federal Regulations and Unfunded Mandates on State, Local, and Tribal Entities*. Washington, DC: Office of Information and Regulatory Affairs.

Seong, Si Kyung & Mendeloff, John (2004). Assessing the Accuracy of OSHA's Projections of the Benefits of New Safety Standards. *American Journal of Industrial Medicine*, *45*(4), 313–328.

Simpson, David (2014). Do Regulators Overestimate the Costs of Regulation? *Journal of Benefit Cost Analysis*, *5*(2), 315–332.

Sunstein, Cass (2011*a*). Memorandum for the Heads of Executive Departments & Agencies: Facilitating Scientific Research by Streamlining the Paperwork Reduction Act Process. M-11-07. December 9, 2010. https://www.whitehouse.gov/sites/default/files/omb/memoranda/2011/m11-07.pdf.

Downloaded from https://www.cambridge.org/core. IP address: 207.241.233.138, on 19 Mar 2021 at 15:03:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/bca.2017.17

Sunstein, Cass (2011*b*). Memorandum for the Heads of Executive Departments & Agencies: Retrospective Analysis of Existing Significant Regulations. M-11-19. April 25, 2011. https://www.whitehouse.gov/sites/default/files/omb/memoranda/2011/m11-19.pdf.

Taylor, Margaret, Spurlock, C. Anna & Yang, Hung-Chia (2015). Confronting Regulatory Cost and Quality Expectations: An Exploration of Technical Change in Minimum Efficiency Performance Standards. *Resources for the Future Discussion Paper 15–50*. Washington, DC.

Thompson, Kimberly M., Sequi-Gomez, Maria & Graham, John D. (2002). Valuing Benefit and Cost Estimates: The Case of Airbag Regulation. *Risk Analysis*, *22*(4), 803–811.

Downloaded from https://www.cambridge.org/core. IP address: 207.241.233.138, on 19 Mar 2021 at 15:03:38, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/bca.2017.17