Grading the Government:
How Reliable Are the Tests?

Richard W. Parker[1]

It is now the belief of millions of Americans -- and the view of many scholars, pundits, the President, his staff, and a majority or near majority of the House and Senate -- that the cost of many health, safety and environmental regulations greatly outweighs their benefits.

The result of this belief has been a sustained campaign to try to force agencies to be more "rational": requirements for ever more elaborate cost-benefit analysis, closer OMB review of agency analyses and decisions (with more frequent over-rides), new requirements for congressional review of agency rules, and any number of additional proposals for regulatory reforms -- most of which are aimed at reining in regulatory agencies. The recent Data Quality Act, which has preoccupied so many of late, is only the most recent manifestation of the current mood of suspicion of the rationality of regulators.

This paper addresses the often-overlooked foundational question: how do we *know* that so many regulations are so irrational from a cost-benefit standpoint? How solid is the empirical foundation for the regulatory reform movement that has captured the support of much of Congress, academia and millions of Americans for over a decade?

Empirically, the belief in excessively costly regulation derives from two main sources. One is a stream of "horror stories" of government zealotry and caprice that have circulated largely unchallenged in the national discourse for nearly two decades: stories of companies forced to clean up Superfund sites to the point where the soil is safe for children to eat 245 days a year, or of air quality regulations issued in total disregard of costs. During the floor debates over Clean Water Act re-authorization House Majority Leader Tom DeLay loved to tell the story of the poor citizens of Lake Jackson, Texas, who were allegedly denied the right to build a golf course, when EPA declared that footprints of cows on the land in question were wetlands when filled with water. And so forth *ad infinitum*.

I will not have much to say about these stories except that they need to be treated with great caution. Anyone who begins to investigate these stories quickly discovers that some are true; others are exaggerated; a startling number (liked the footprints of cows story) turn out to be pure fabrications.[2] Unfortunately, politicians, journalists and even scholars often appear to be quite content to report sensational allegations as such, without investigating their veracity. Moreover, even if the story turns out to be true, one has no way of knowing whether the incident is typical of agency practice, or aberrational.

The probative shortcomings of anecdotes have naturally given rise to a second source of skepticism that has largely supplanted the stream of anecdotes, at least in scholarly and policy circles. This second source of skepticism involves a group of studies which have yet to be recognized as a formal

---

1. Professor of Law, University of Connecticut Law School. Professor Parker has also served as Special Counsel to the Deputy Administrator of EPA (1993-1994) and Assistant General Counsel at the Office of the United States Trade Representative (1986-189). He holds a B.A. in Public Affairs from Princeton University, a J.D. from Yale Law School, and a D.Phil. in Politics from Oxford University, which he attended as a Rhodes Scholar.

2. See Annex for illustrations and details.

genre. John Graham calls them "league tables." I will refer to them by a more straightforward term: "regulatory scorecards."

What are regulatory scorecards? Scorecards are in essence a sub-species of cost-benefit analysis. Most cost-benefit analyses devote hundreds of pages to investigation of the costs and benefits of a single project or rule. Scorecards reduce these hundreds of pages to a few summary statistics -- costs, benefits, net benefits, and/or cost-per-life-saved. The scorecardists then tabulate these summary statistics for each regulation across scores of rules in order to generate an seemingly *concise* and *precise* picture of the cost-benefit rationality of programs, agencies, and even government regulation across the board.

While any number of scorecards have circulated in one form or another, my research suggests that three have been particularly influential in shaping the modern debate over the cost-benefit rationality of the administrative state.

In 1987, an OMB economist named John Morrall published a table of 44 regulations. One-third of the regulations in his list cost over $100 million for every statistical life saved. One cost up to $72 *billion* per life saved.

In 1995, John Graham and Tammy Tengs at the Harvard Center for Risk Analysis co-authored a study entitled Five-Hundred Life Saving Interventions and their Cost Effectiveness. This study claimed to have found wildly disparate regulatory costs suggesting (according to the authors) gross inefficiency in life saving. Moreover, they found that the least cost-effective interventions are those aimed at controlling toxins.

In 1996, they published a follow-on study -- entitled The Opportunity Costs of Haphazard Social Investments in Life-Saving -- which used computer programming to perform a "rational" re-allocation of funds among 185 of the 500 interventions examined in their earlier study. Their conclusion: 60,000 additional lives could be saved at constant cost by simply re-allocating funds to fully implement the most cost-effective life-saving interventions first. The implication, of course, is that 60,000 lives are lost each year by the current "irrational" allocation of risk-reduction dollars -- a situation that John Graham has called "statistical murder."

In 2000, Robert Hahn published what he calls "the most comprehensive assessment to date of the impact of federal regulatory activities on the economy." He compiles the costs benefits, and net benefits of over 100 major regulations promulgated over the period 1981 through mid-1996. His conclusion: less than half the major rules issued over the period 1981-1996 pass a neutral economist's benefit-cost test -- *using the government's own numbers.*

Three prestigious scholars all arrive at the same basic conclusion: our regulatory system is pervasively irrational and that irrationality is killing people. No surprise, these findings have fueled scathing regulatory critiques by Justice Stephen Breyer and scholars too numerous to mention. They have featured in General Accounting Office reports, in OMB annual reports to Congress, in congressional testimony, in court of appeals opinions, in the debate over the Contract With America, in administrative law school textbooks, and in all manner of think tank publications.

Most important, perhaps, their basic conclusions have appeared in virtually every newspaper and magazine as evidence of pervasive regulatory irrationality. I strongly suspect that long after the studies themselves have disappeared from memory, the skepticism they engender lingers on, shaping beliefs of people who may not necessarily be aware of whence their ideas come.

Hahn himself has been installed as Director of the prestigious AEI-Brookings Joint Center for Regulatory Studies. Graham has been appointed Director of the Office of Information and Regulatory Affairs (OIRA) at OMB -- a position which gives him oversight of all agency regulation. Morrall is a senior official in OIRA. The regulatory world view shared by Morrall, Hahn, and Graham is now the view of the Bush Administration, much of the leadership of Congress, many leading scholars, and millions of Americans.

But how reliable are the studies? To investigate this question I went back to the spreadsheets, where possible, and back to the original rules on which the scorecards are based. I tried to replicate the scorecards' numbers. I explored what, if anything, their numbers leave out. I did not investigate every rule, of course, but I audited their database the way a careful accountant would audit a company's financial statements -- not looking at every entry, but looking at enough entries to provide a fair picture of how reliable the bookkeeping is.

The full report of what I found runs to 90 pages. I cannot begin to document all my conclusions here. So let me just summarize them briefly, and illustrate a few key conclusions with a couple of examples drawn from the larger study. I will happy to supply further documentation, if you like, in the question and answer period or afterwards.

In a nutshell, these widely-cited studies suffer from serious flaws which I divide into two categories. First, there are the errors which might have been avoided through better, more scrupulous implementation of the scorecard concept. Second, there are the errors which are unavoidable because they are inherent to the scorecard enterprise itself.

We begin with the avoidable errors. All three studies rely on undisclosed data and non-replicable calculations. They use regulatory samples which are biased against a finding of rationality. T hey misrepresent ex ante guesses about the costs and benefits of future or hypothetical regulations as actual measurements of "the" costs and benefits of regulation. They grossly under-estimate the value of lives saved, or the number of lives saved, or both. Amazingly, John Graham's most famous conclusion -- that 60,000 lives are lost by over-zealous regulation of which toxic regulations are the worst -- is not logically supported by the author's own data.

Then there are the errors inherent to the scorecard genre. First and foremost, scorecards ignore virtually all benefits that are not quantified and/or monetized. They thereby exclude most of the environmental benefits of environmental regulations. They also fail to count many health benefits, and all intangible benefits ranging from the avoidance of pain and suffering or familial and societal disruption, to the promotion of a public sense of security, fairness, confidence in markets, etc.

Scorecards ignore the cumulative impact of multiple risks, as well as all questions about the distribution of risk and the *fairness* of that distribution.

Finally, regulatory scorecards obliterate the large uncertainties that are present in virtually every regulatory impact assessment. That, ironically, is the key to their great influence: the regular use of speciously precise numbers lends them a scientific air which impresses the media and the unsuspecting public, but is quite unwarranted by the data.

The shortcomings described above are not merely flaws in the eye of this beholder. When scorecardists disregard unquantified costs and benefits, distributive and equitable impacts, and uncertainties, they do so in violation of widely agreed principles of cost-benefit analysis to which the scorecardists themselves have subscribed.

The problem, however, is that scorecards cannot possibly adhere to the principles of responsible cost-benefit analysis while continuing to do the thing that has made them quotable and famous, which is to boil down huge arrays of complex analysis to a few summary numbers.

That is why I conclude that regulatory scorecards cannot be salvaged by better implementation. They should simply be abandoned.

These are serious charges. My Article (now in draft) documents them in depth. Let me just illustrate a small sampling of these charges with specific examples.

The Morrall table. I begin with the Morrall table of rules, one-third of which supposedly cost more than $100 million per life saved, and one of which costs $72 billion. These are incredible numbers. Where do they come from?

Morrall, like all scorecardists, does not generate his own cost-benefit numbers. He relies on the assessments of others. But Morrall, unlike the others, revises agency cost and benefit estimates whenever he disagrees with them -- often by several orders of magnitude, and always in the direction of higher costs and lower benefits.

For example, in 1985 OSHA estimated that its proposed formaldehyde exposure regulation would save from six to forty-seven lives over forty-five years. Morrall alters that estimate to one life saved every hundred years, while increasing OSHA's compliance cost estimate by a factor of twenty.

In 1986, EPA estimated that its restrictions on land disposal of certain toxic and bioaccumulative wastes would avert forty cases of death or illness per year. Morrall translates that to 2.5 lives saved per year -- with no explanation of where that number comes from (it is not found in any agency document) while, apparently, increasing the agency cost estimate from $97 million per year to $1.3 billion.

Morrall defends such revisions by claiming that: "Regulatory agencies . . .tend to over-state the effectiveness of their actions. Where such biases were evident . . . I made the corrections . . relying on published studies."

This defense faces three rather obvious objections. First, Morrall nowhere proves that agency regularly overstate the effectiveness of their actions. Indeed, we will shortly see that agencies under-state (or at least under-quantify) benefits in many cases. Second, Morrall has no training in any of the disciplines which would qualify him to substitute his own judgement for agency scientists on matters of

exposure, risk, or compliance cost. Third, Morrall does not document, much less defend, either the studies he allegedly relied on, or the criteria he applied, in generating his own substitute numbers. His assumptions and calculations are, by own admission, "scattered around in [his] filing cabinets" and are not available to outside reviewers.

Granted, Morrall is a government official. Scholarship is not his first vocation. The fact remains: his findings cannot be replicated. Yet his numbers circulated in the national discourse, unchallenged, for 13 years until Lisa Heinzerling at Georgetown University came along to look behind them.

Hahn. Hahn, by contrast, appears to have some compunction about making up numbers. Hahn claims -- in his title and at least 18 times thereafter -- that he is just "using the government's numbers" for cost and benefit, and he overtly distinguishes his approach from Morrall's in this regard.

Hahn's study, unfortunately, raises a raft of problems of its own. To begin with, Hahn's published study does not even disclose the *names* of the rules he examined in concluding that over half of all major rules issued since 1981 fail cost-benefit analysis. Merely getting the list of rules -- and the corresponding tabulation of costs and benefits -- required months of supplication.

When I finally obtained the spreadsheet, I immediately made a startling discovery. Forty-one of the 136 rules in his database -- fully 31 percent of all the rules -- are assigned a zero benefit. These rules, it should be emphasized, are not rules for which it is claimed that costs *equal* benefits. *These are rules that are alleged to offer no benefit whatsoever.*

The list of zero benefit rules includes:

-- a rule requiring that owner/operators of tankers develop plans to respond to large oil spills;

-- a rule to require that air polluters hold comprehensive permits which lay out their pollution control obligations;

-- a rule requiring the public reporting of releases of certain toxic chemicals from large manufacturing facilities;

-- a Clean Water Act rule aimed at protecting sensitive coastal areas from non-point-source water pollution;

-- a rule to protect agricultural workers from exposure to harmful pesticides;

-- three rules establishing national primary drinking water standards to limit public exposure to toxic pollutants in drinking water;

-- a FDA rule establishing requirements for the safe handling of seafood in commercial processing operations.

It turns out that Hahn, with a few narrow and limited exceptions, has assigned a zero value to any benefit which the government's regulatory impact assessment does not quantify and monetize. Hahn,

amazingly, also zero-values even benefits that *are* quantified and monetized in an agency RIA, unless they happen to fall into one of his select categories of recognized benefit -- even as he insists that he is using the government's numbers.

Nor are the omissions of unquantified variables confined to the zero-benefit rules. Rules that display a positive number in the benefits column turn out, on closer inspection, to have had whole categories of important benefits excluded from the tally.

Morrall and Graham/Tengs adopt an even more extreme accounting convention: by evaluating every regulation solely in terms of cost-per-life-saved, they manage to exclude non-life-saving benefits entirely.

Given my time constraint, a single example will have to suffice to illustrate the consequence of such omissions for Hahn's analysis. Many more examples may be found in my Article.

In 1992 EPA promulgated an agricultural worker protection standard for pesticides. Noting that the rule would help protect 3.9 million agricultural workers across the United States who are exposed to pesticides in their work, EPA predicted the following benefit:

> "avoiding 8,000-16,000 physician-diagnosed (non-hospitalized) acute and allergic pesticide poisoning incidents, [while] avoiding about 300 hospitalized acute and allergic pesticide poisoning incidents, and avoiding potentially important numbers of cancer cases, serious developmental defects, stillbirths, persistent neurotoxic effects and non-diagnosed acute and allergic poisoning incidents."

Hahn's scorecard, however, does not recognize any non-accidental "health benefit" other than "reducing the risk of cancer, heart disease, and lead poisoning." Since avoiding stillbirths, persistent neurotoxic effects and pesticide poisoning does not fit within any of these categories, the regulation protecting 3.9 million agricultural workers from acute pesticide poisoning is assigned a zero benefit. It thus fails Hahn's cost-benefit test.

Is this the way cost-benefit analysis is supposed to work? Hardly. It has long been recognized that data limitations, scientific uncertainties and difficulties of valuation often make it infeasible to try to quantify and monetize every important cost and benefit. That is why the Annapolis principles for sound cost-benefit analysis, which Hahn himself co-authored, advises:

> "not all impacts of a decision can be quantified or expressed in dollar terms. Care should be taken to assure that quantitative factors do not dominate important qualitative factors in decision-making."

No one disputes this principle. Scorecards simply do not practice it.

Nor is it easy to see how they could. The point of scorecards, after all, is to come up with the number of rules that generate positive net benefits, or that cost less than some threshold amount per life saved. Without the numbers, how does one keep score?

The dilemma facing scorecards is well illustrated in OMB's Annual Report to Congress on the Costs and Benefits of Federal Regulations. Each year OMB compiles its estimates of the costs and benefits of regulations. Unlike the scorecards discussed in this article, however, OMB also includes a column entitled "Other Information" in which unquantified costs and benefits are narratively described. But that means that OMB cannot offer a final numerical verdict on the cost-benefit rationality of the rule. By following good practice, OMB's report *ipso facto* sacrifices its role as a true scorecard. That is why I maintain that omitting unquantified variables is endemic to scorecards.

Graham. Our partial review concludes with a brief look at what the Graham scorecard proves or, more particularly, what it does *not* prove.

To begin with, Graham's Opportunity Cost study, despite the sensational publicity, does not prove that even a single person -- statistical or otherwise -- has ever died as a result of irrational regulation. Graham's "statistical murder" charge rests entirely on the counter-factual assumption of a fixed national budget for risk reduction, so that a dollar spent on Risk A is a dollar taken away from efforts to mitigate Risk B. In fact, there is no such budget, and no such trade-off. We live in a $9 *trillion* dollar economy of which only a tiny sliver is spent on regulatory risk reduction. If money spent cleaning up hazardous waste sites might save more lives if re-directed to combating smoking, then so might the $36 billion spent each year on lottery sales, or the $7.6 billion spent on spectator sports. Indeed, by Graham's logic, lottery sales kill 7,200 people every year, while spectator sports kill 1,520 statistical people.

Another way to appreciate the fallacy of the 60,000 lives claim is simply to examine the actual interventions which account for additional life-saving in his reallocation scenario. It It turns out that just two interventions -- continuous (versus nocturnal) oxygen for hypoxemic obstructive lung disease, and influenza vaccines for all citizens -- account for over 42,000 of the more than 60,000 additional lives saved by his hypothetical re-allocation.

Are we to believe that the nation's failure to fully implement influenza vaccines for all citizens and to provide continuous (vs. nocturnal) oxygen for hypoxemic obstructive lung disease is somehow related to the allegedly excessive regulation of benzene or other interventions at the cost-ineffective bottom of his list? If not, where is the statistical murder?

Of course, regulation may be inefficient even if it does not kill. Inefficiency is a valid criticism of regulation. Amazingly, however, Graham's data -- when closely examine -- do not even establish a significant pattern of inefficient (or cost-ineffective) spending.

If we take $8 million per life as the reasonable threshold of "cost-effective@ life-saving expenditure (I threshold my Article establishes as a conservative one), it turns out that only about 12 percent of the total spending in his baseline is allocated to "cost-ineffective" interventions. Well might one ask whether the vast array of government programs for defense, highway construction, or airport security, could pass a test of eighty-eight percent spending efficiency at the margin.

Moreover, the Opportunity Cost study is hardly a robust demonstration of systemic irrationality. As we have seen, *two-thirds* of the additional lives saved by Graham/Tengs' re-allocation are saved by fully implementing just two interventions. *95 percent* of the 60,000 additional lives are saved by fully implementing just 9 interventions -- of which 3 are medical, 5 are traffic-related and 1 is environmental.

Over 90 percent of the re-allocated funds which save these additional lives are supplied by the opinion of a single author that it is more cost-effective to ban residential growth in tsunami-prone areas than to construct sea walls to control the damage.

In other words, people are drawing sweeping, system-wide critiques from what is, at most, the story of a handful of interventions viewed through the eyes of a handful of authors of (often) non-peer-reviewed and (sometimes) unpublished studies.

Moreover, Graham's re-allocation works by finding one or more instances of under-regulation to match every instance of over-regulation. But under-regulation, of course, is not the lesson that regulatory critics choose to draw from the Graham study.

The third of Graham's trio of famous claims is that toxic-related interventions are the worst (most cost-ineffective) of the bunch. In fact, Graham's spreadsheet paints a rather different picture. If we again take $8 million per life as our provisional threshold of cost-effective spending, Graham's own data suggest that about *four percent* of the funds spent on toxic-exposure-related interventions in the baseline case were spent on interventions that exceed that threshold. 63 percent of the funds allocated to non-toxic related interventions were spent on interventions that exceed the $8 million per life threshold. In other words, the authors' own data suggest that, when one focuses on actual implementation patterns, toxic control programs are *15 times* more cost-effective than their non-toxic counterparts in their overall pattern of spending to save lives.

In short, none of these famous scorecards prove what they claim to have proved. None, or all together, prove that agencies are systematically regulating in an arbitrary and capricious way, as the authors suggest. In fact, as I show in much more depth in my Article, these scorecards are so deeply flawed at that they really don't prove much of anything at all.

While some of the defects of these scorecards might have been cured by better practice, the most serious defects cannot be cured because they are inherent in the scorecard enterprise -- which requires reducing everything to numbers, or ignoring everything that is not reduced to numbers.

How, then, should the cost-benefit rationality of regulations be evaluated? My suggestion (paraphrasing the well-known TV commercial) is to analyze regulations "the old-fashioned way. One regulatory investment at a time."

Start with claims of regulatory failure. Investigate them with the same care and diligence that the National Transportation Safety Board brings to the investigation of plane crashes. Publish the factual findings in draft form for persons of all persuasions to comment. Seek the widest possible consensus on the facts of each significant failure. Explore whether the incident is typical of agency practice in a given area. And, finally, seek to identify the causes of the failure. And then draw out any systemic conclusions that may emerge from rigorous, inductive investigation into the facts.

Let me conclude by clarifying what I am, and am not, claiming. I am not claiming that all regulations are rational from a cost-benefit perspective, and I am certainly not claiming that the modern administrative state is not in need of reform. Volumes have been written, quite properly, about the incredible slowness and cumbersomeness of agency decision-making; the de-moralizing impact of work

life in agencies where ten percent of the people do ninety percent of the work while the remainder rest on their civil service protection; the irrationality of congressional micro-management that compels agencies to regulate one thing while denying them authority to address other, greater risks, and so forth. I certainly do not mean to deny or discount these very real problems. Nor do I deny that there are some, perhaps numerous, regulations that would not survive a really rigorous cost-benefit analysis.

I am simply saying that scholars, policy-makers and the public should be aware that we simply do not know how "efficient" or "rational" government regulation is over-all, from a cost-benefit perspective, because the principal tests that have been used to reach such judgements are invalid. Until valid tests are adduced, we should withhold judgement on such matters.

Meanwhile, instead of pointing the finger at government for the poor quality of its data and analysis, some of the leading critics of regulation should look to the quality of their own studies. For they are shockingly bad.

Annex
Three anecdotes which illustrate the problem with anecdotes

1. In his influential 1993 book, Breaking the Vicious Circle, Steven Breyer alleges that EPA once tried to force companies to clean up a Superfund site to the point where the soil on the site was safe for children to eat 245 days a year. Breyer stopped them from doing so during his tenure on the First Circuit bench. An EPA official interviewed for this study last year did not dispute Breyer's basic account of that case.

2. EPA's air quality regulations are issued in total disregard of cost. Or so alleged a prestigious group of economists -- led by Robert Hahn, Director of the prestigious AEI-Brookings Joint Center for Regulatory Studies and Nobel-laureate Kenneth Arrow -- in an *amicus* brief filed with the Supreme Court in *Whitman v American Trucking Assoc., Inc.,* 531 U.S. 457, 463 (2001). Many newspaper editorials at the time echoed their allegation.

   Yet as Justice Scalia patiently pointed out in his majority opinion in that case, the claim represents a serious misreading of the Clean Air Act. The Clean Air Act does not forbid taking costs into account in setting air pollution regulations. In fact, the Act requires it dozens of times over. The Act merely forbids taking cost into account in making the basic medical determination about what is an unsafe level of pollution for the human lung to inhale.

3. House Majority Leader Tom DeLay (R- Tex) opposed Clean Water Act re-authorization with, *inter alia*, the story of the poor folks of Lake Jackson, Texas. He claims they were denied permission to build on land of their choosing, when EPA declared the Afootprints of cows@ on the land to be wetlands when filled with rainwater. The story is a pure fabrication. According to the US Fish and Wildlife Service, the "footprints of cows@ to which House Majority Leader Tom DeLay referred were not footprints at all, but "wetland sloughs@ several feet deep and up to two hundred feet wide, which fill with water every year to provide vital sustenance to local and migrating birds. In fact, the land in question is not pasture but a forest which forms a part of the Aonly [remaining] forest habitat adjacent to the Gulf of Mexico@ and furnishes a vital refuge for migrating songbirds of North America.[1]

A few journalists and advocates have picked up some of these deceptions, but their rebuttals have never caught up to the original mistake.[2] How, in the current climate, does one truth from falsehood or exaggeration?

---

1. See Letter of David L. Hankla, Field Supervisor, US Fish and Wildlife Service, to Colonel Robert B. Gatlin, US Army Corps of Engineers, April 19, 1995 [letter on file with author].

2. For evidence of the broader veracity problem in the regulatory debate see, e.g., Tom Kenworthy, Truth Is Victim in Rules Debate: Facts Don't Burden Some Hill Tales of Regulatory Abuse, Wash. Post, Mar. 19, 1995, at A1 (relating these kinds of anecdotes that "have the ring of truth, but not the substance"); Jessica Mathews, Horror in the House, Wash. Post, Mar. 5, 1995, at C7; and Citizens for Sensible Safeguards, "Myths and Consequences: Paying for the Use of Myths and Distortions by Anti-Regulatory Zealots," May 17, 1995 (collecting 27 widely-circulating anecdotes about government regulatory abuse which turned out to be false, exaggerated, or at least factually contestable) (unpublished manuscript on file with the author).