MERCATUS
RESEARCH

RISK ASSESSMENT, SAFETY ASSESSMENT, AND THE ESTIMATION OF REGULATORY BENEFITS

Richard B. Belzer

MERCATUS CENTER
George Mason University
Bridging the gap between academic ideas and real-world problems

ABOUT THE MERCATUS CENTER AT GEORGE MASON UNIVERSITY

The Mercatus Center at George Mason University is the world's premier university source for market-oriented ideas—bridging the gap between academic ideas and real-world problems.

A university-based research center, Mercatus advances knowledge about how markets work to improve people's lives by training graduate students, conducting research, and applying economics to offer solutions to society's most pressing problems.

Our mission is to generate knowledge and understanding of the institutions that affect the freedom to prosper and to find sustainable solutions that overcome the barriers preventing individuals from living free, prosperous, and peaceful lives.

Founded in 1980, the Mercatus Center is located on George Mason University's Arlington campus.

www.mercatus.org

Mercatus Center
George Mason University
3351 North Fairfax Drive, 4th Floor
Arlington, VA 22201-4433
(703) 993-4930
mercatus.org

Release date: October 10, 2012

ABOUT THE AUTHOR

Richard B. Belzer is president of Regulatory Checkbook and managing editor of NeutralSource.org, non-profits seeking to improve the quality of regulatory science and economics. Belzer has been an economist at the Office of Management and Budget and treasurer of the Society for Risk Analysis, and currently serves as secretary/treasurer of the Society for Benefit-Cost Analysis.

ABSTRACT

Most federal agencies are required to conduct benefit-cost analyses for their largest regulations. Benefit-cost analysis is intended to objectively inform decision makers about the myriad effects of a range of potential alternatives. For many regulations, benefit-cost analyses depend on health risk assessment. It is well-established that a clear conceptual distinction must be established and maintained between (positive) risk assessment and (normative) risk management. Decision makers cannot ensure that regulatory choices conform with their intentions if they cannot discern where the science in an analysis ends and value judgments begin.

In practice, the distinction between risk assessment and risk management has become thoroughly blurred. This paper provides an introduction to how risk management choices are embedded in risk assessments, and how this leads to biased and misleading benefit-cost analyses. Restoring objectivity to health risk assessment is an essential prerequisite for objective regulatory benefit-cost analysis.

JEL codes: D61; L51; I18; K23; Q51; Q58

Since 1981, most federal agencies have been required to conduct regulatory impact analyses (RIAs) for their largest regulatory actions. These RIAs are submitted to the Office of Management and Budget (OMB) for review along with the draft regulation. They are intended to help inform agency heads and White House officials about the nature of the problem the regulation is supposed to solve and the costs and benefits of a range of options.

Beginning in 1990, OMB published several iterations of guidelines to agencies on how to prepare RIAs.[1] Each one emphasized a fundamental point: RIAs must be objective portrayals of the social costs and social benefits of regulatory action.[2] For health and environmental regulations, this requires the use of risk assessments, a set of methods devised to provide a systematic description of the human health consequences of exposure to physical, chemical, or microbiological hazards.[3] Risk assessment is distinguished from risk management, the process of selecting among an array of regulatory and nonregulatory alternatives.[4]

These are fundamentally different activities. Risk assessment is the science-based practice of ascertaining what is expected to happen under what circumstances. Risk management is the policy activity of deciding what, if anything, to do about it. It has long been a maxim in the field that a clear conceptual distinction between the assessment and management of risk should be established and maintained.[5]

This distinction has not been maintained by federal regulators.[6] Instead of a clear conceptual distinction, the line dividing risk assessment and risk management has become increasingly blurred. Agency policy officials are sometimes accused of interfering with science or trying to manipulate risk assessments to advance particular risk management goals. An equally or perhaps more common

1. Office of Management and Budget (1990, 1996, 2000, 2003).
2. Office of Management and Budget (2003): "Analysis should be credible, objective, realistic, and scientifically balanced." OMB also requires analyses to adhere to OMB's and the authoring agency's own information quality guidelines (Office of Management and Budget 2002).
3. There are also risk assessments that address ecological concerns, like species preservation. They are beyond the scope of this paper.
4. National Research Council (1983, 18).
5. Ibid., 151. This was the first of three top-level recommendations.
6. North (2003).

phenomenon, however, arises when risk assessments include within them implied risk management policy choices that science can inform but not determine.[7] This paper provides an introduction to how risk management choices wind up in risk assessments where they do not belong and how this results in RIAs that distort estimates of regulatory benefits.

II. WHAT IS A RISK ASSESSMENT?

Agencies prepare two very different types of documents that are often called risk assessments.[8] A true risk assessment is one that gives a quantitative estimate of the likelihood of a specific health outcome associated with different levels of exposure to (or doses of) a chemical, physical, or microbiological agent.[9] This relationship of dose to health effect is called the dose-response relationship. Economists use these estimates as inputs to an RIA to estimate the social value of reducing or eliminating a health risk.

The other type of document is more accurately called a safety assessment. Safety assessments give a quantitative estimate of the amount of a substance people can be exposed to and still be safe. Ideally, agency safety assessments would strive to identify the highest exposure or dose that is safe. Of course, this begs a question that science cannot answer: How safe is "safe"? Safety assessments are discussed in section III below.

A. What a Risk Assessment Looks Like

Cancer is the most common health risk for which a risk assessment is prepared. Typically, the primary product of a risk assessment is the unit-risk estimate. The U.S. Environmental Protection Agency (EPA) defines this as the upper-bound excess lifetime cancer risk estimated to result from continuous exposure to an agent at a concentration of 1 microgram per liter (μg/L) in water or 1 microgram per cubic

7. Bipartisan Policy Center (2009, 15). That this would occur could have been easily foreseen. In the letter transmitting the 1983 "Red Book" to its sponsor, Food and Drug Administration commissioner Arthur Hull Hayes Jr. and National Research Council chairman Frank Press noted, "Federal agencies that perform risk assessments are often hard pressed to clearly and convincingly present the scientific basis for their regulatory decision" (iii).
8. The term "risk assessment" also is used to describe the process of identifying hazards, conducting dose-response assessment, estimating exposure, and characterizing risk (National Research Council 1983, 3). In this paper, the term is used to refer to one or more documents typically called "health risk assessments."
9. *Exposure* is the amount of a substance available at the exchange boundaries of an organism, such as on the skin, in the lungs, or ingested into the gut. *Dose* refers to the amount of a substance available for metabolism or another biologically significant process after crossing that exchange boundary. There are several different ways *dose* can be defined. See U.S. Environmental Protection Agency (2011b).

meter (μg/m3) in air.[10] A unit-risk estimate of 1 means there is a one in one million chance of developing cancer over one's lifetime if exposed to 1 μg/L in water or 1 μg/m3 in air.[11]

How large is a one-in-one-million cancer risk? The American Cancer Society reports that, over a lifetime, the chance of developing an invasive cancer is 45 percent for men (almost one chance in two) and 38 percent for women (a bit more than one chance in three).[12] That means a new one-in-one-million cancer risk from exposure to a carcinogenic substance raises these lifetime risks to 45.0001 percent and 38.0001 percent for men and women, respectively. As of January 1, 2008, it was estimated that 11,958,000 people were living with cancer, including people diagnosed with cancer in the past as well those recently diagnosed. This is approximately 4 percent of the U.S. population.[13] Risk assessments are prepared for only a few health effects other than cancer, such as some microbial risks in food and water and certain health effects from so-called criteria air pollutants regulated under the Clean Air Act. The models used to estimate risk from air pollution are very different. They are derived from very large epidemiological studies rather than toxicological studies performed on laboratory animals.[14]

These unit-risk estimates, and various forms of safe exposure thresholds discussed in the following section usually are reported very precisely—much more precisely, in fact, than the underlying science actually justifies. The importance of this is discussed in greater detail in section IV.B.

B. Extrapolating from High to Low Doses

Except in extraordinary circumstances, the only actual data available for estimating human cancer risk at very low environmental levels comes from high occupational exposures or very high laboratory exposures to animals. This means the risk assessment task almost always consists of extrapolating from an exposure level where effects can be observed to an exposure level where observation is impossible. The accuracy of this extrapolation depends on whether there is good reason

10. U.S. Environmental Protection Agency (2011f). The meaning and significance of the adjective "upper-bound" is discussed in section II.E.
11. This will often be expressed as a 10^{-6} cancer risk. It is the same thing as one-in-one million because there are one million micrograms in one liter of water or one cubic meter of air.
12. American Cancer Society (2011b). The lifetime risk of dying from cancer is about half as great—23 percent for men and 20 percent for women.
13. American Cancer Society (2011a). How much greater this number would be if everyone had been exposed to an additional 10^{-6} cancer risk cannot be determined by, say, multiplying the population (311 million) by 10^{-6}. This is because cancer prevalence is a snapshot taken on a specific date, but the number of additional cancer cases estimated to result from an additional 10^{-6} cancer risk are spread out over about 70 years.
14. A small number of cancer risk assessments for specific chemicals are derived from epidemiological studies of workers exposed at unusually high levels.

to believe that the biological phenomenon in question behaves the same way at very high and very low doses.

To a point, this is similar to econometric forecasting of such things as future GDP, employment, or inflation. Forecasting relies on the expectation that markets will perform in the future much like they have in the past. However reasonable this assumption might seem to be, the actual performance of econometric forecasting has not been as reliable as one would expect if this assumption were valid. Still, econometric forecasters can learn from their errors because actual economic performance will become known in due course. In the case of low-dose cancer risks, however, even this limited degree of confidence in forecasting does not exist. The predictions of low-dose risk from high-dose evidence are rarely subject to falsification, perhaps the key feature of every field that legitimately calls itself scientific. Thus, risk assessment is scientific when it is used to describe what is known or unknown; it is trans-scientific when it is used to predict risk at doses beyond the domain in which there are data; and it is nonscientific whenever it makes claims of a policy nature, such as what dose or exposure is safe, reasonable, or appropriate.

C. Extrapolating from Animals to Humans

The dose-response challenge is made even more complicated by having to extrapolate from laboratory animals to humans when human data are not available. This extrapolation can be at least as problematic as the extrapolation from high to low doses. EPA guidance requires that this be done in accordance with a fixed set of assumptions based on rather crude measures—differences in the surface area or body weight of a rodent relative to that of a human.[15]

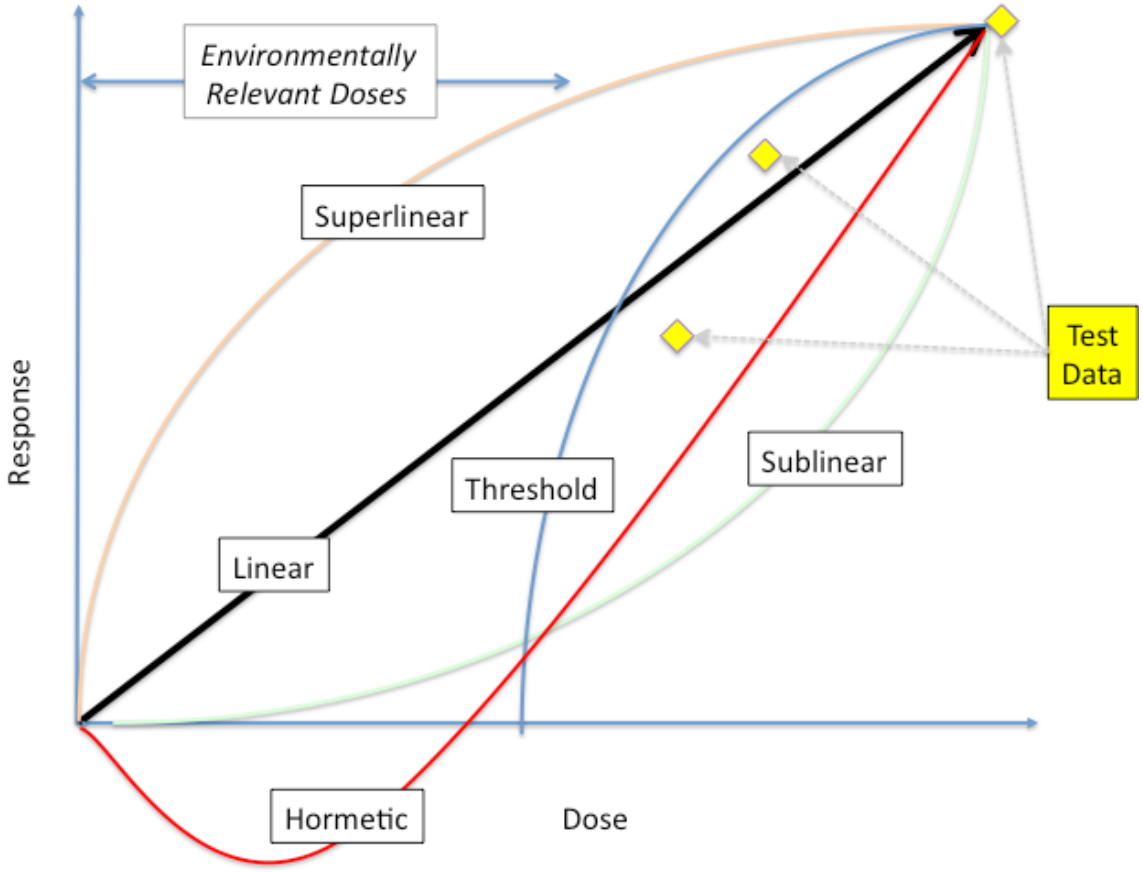
Considerable research has been devoted to understanding differences between humans and rodents, especially with regard to pharmacokinetics (that is, where a substance goes within an organ or the whole body) and pharmacodynamics (that is, what happens along the way). Sometimes this knowledge is used in risk assessment and allows for more biologically grounded results. However, existing agency guidelines impose high hurdles on the use of this kind of new knowledge.[16]

D. The Linear/No-Threshold Assumption

Almost every risk assessment has to include some method for converting high-dose exposure into low-dose risk estimates. For cancer, the default model is one that

15. U.S. Environmental Protection Agency (2005a, 3-6 to 3-9).
16. See, for example, U.S. Environmental Protection Agency (2005a, Appendix A: Major Default Options, reprinted as the appendix to this paper).

FIGURE A: ALTERNATIVE WAYS TO EXTRAPOLATE FROM HIGH TO LOW DOSES



Source: Author's calculations.

draws a straight line from the lowest dose at which there is an effect through the origin.[17] It is assumed that there is no dose for which cancer risk is zero and no dose for which cancer risk is negative (that is, exposure is protective).

The notion that low-dose risk is a linear function of dose is over 50 years old and is still assumed to be the predominant way cancer occurs. EPA guidelines establish the linear/no-threshold assumption as the default way to estimate human cancer risk.[18] Overcoming this default is very difficult. Agencies typically require a great deal of evidence to depart from default assumptions even thought the scientific justification for the defaults is often limited.

17. By "straight line" it is usually meant linear in natural logarithms (that is, logs with the transcendental mathematical constant e, which is approximately equal to 2.718).The usual biologic justification for this assumption is that cancer is the result of a random process involving one or more hits that damage cellular DNA. Much of this damage is successfully repaired, but some might not be. Cells with damaged DNA may replicate defective cells that grow and further replicate without normal regulatory constraints.
18. U.S. Environmental Protection Agency (2005a, A-8).

E. Use of Upper-Bound Estimates

Every statistical model has one or more "best" estimate—expected value estimates that are unbiased, meaning that they give predictions that treat errors equally to minimize them. Conventional health risk assessment practices do not use best estimates, however. Instead, they typically use the upper 95th percentile estimate. This is done to make sure that the estimate is much more likely to overstate risk than to understate it.[19]

F. Exposure

So far, all the discussion has concerned the hazard component of risk assessment. This is essential, for without a hazard there would be no risk. But it is not sufficient, because without exposure there is no hazard severe enough to cause risk.

1. Default Assumptions

There are three general ways exposure may be captured in risk assessment. The first way is by using default assumptions. For example, the longstanding default assumptions for daily inhalation and drinking water consumption are 20 cubic meters per day (m3/d) and 2 liters per day (L/d), respectively. Neither assumption turns out to be very accurate and they both overstate the average, so when they are used in risk assessment they result in an overestimation of average risk.

2. Generic Empirically Derived Estimates

The second way to estimate exposure is by relying on generic empirically derived estimates. Over the past 15 years, EPA has published a series of Exposure Factors Handbooks (EFH), most recently in 2011.[20] The EFH is a handy and comprehensive compendium of empirical data obtained from mostly peer-reviewed sources. It contains extensive distributional data, plus recommended values for means and 95th percentiles for various routes of exposure, including inhalation and drinking water ingestion.[21] Substituting these empirical averages for default assumptions illustrates the extent to which defaults overstate average exposure.

For adult inhalation, the EFH recommends average rates ranging from 12.2 to 15.7 m^3/d depending on the age group.[22] Thus, the 20 m^3/d default assumption exceeds

19. With an extremely skewed risk distribution, the 95th percentile could be lower than the mean. In more typical cases, however, the 95th percentile is considerably higher than the mean, and the 95th percentile is an upwardly biased approximation of the mean.
20. U.S. Environmental Protection Agency (2011a).
21. U.S. Environmental Protection Agency (2011a, Table ES-1). Routes of exposure include inhalation, drinking water and tap water consumption, ingestion of water while swimming, infant mouthing, frequency, soil and dust ingestion, adhesion of solids to skin, the consumption of various foods, and activity levels.
22. Ibid. These rates are for men and women combined. Adult 95th percentiles range from 16.6 to 21.4 m^3/d.

the average by 27 to 64 percent. A risk assessment that uses the 20 m^3/d default assumption instead of the mean will overstate risk by a similar amount—more if the risk is bounded by a threshold.

For exposure via drinking water, the daily dose is calculated using the following equation:

$$\text{Default Daily Dose} = \frac{\text{Contaminant Concentration x Daily DW Ingestion}}{\text{Body Weight}}$$

For an assumed contaminant concentration of 1mg/L:

$$\text{Default Daily Dose} = \frac{1\ \text{mg/L} \times 2\ \text{L/d}}{70\ \text{kg}} = 0.029\ \textit{mg/kg/d}$$

This fraction is an objective estimate of average exposure if all of the defaults are averages. They are not, however.

According to the EFH, age-weighted average body weight is about 85.5 kg for males 18–65 years and about 74.6 kg for females, and average drinking water ingestion is 1.043 L/d:[23]

Thus, the default assumption overstates average daily adult dose by 138 percent for males and 107 percent for females.[24]

Other default values also may result in biased estimates of average exposure. For example, exposure to ambient air pollutants requires being outside. For 18–65 year olds, the EFH reports that the average time spent outdoors is 281 minutes (4.68 hours).[25]

Assuming 24 hours of exposure overstates exposure by a factor of 5.1. The U.S. Occupational Health Administration assumes workers are exposed for 45 years. The EFH, however, recommends median occupational tenures ranging from 2.0 years for 16–24 year old males to 26.9 years for 65–69 year old males.[26]

$$\text{Average Daily Dose (M)} = \frac{1\ \textit{mg/L} \times 1.043\ \textit{L/d}}{85.5\ \textit{kg}} = 0.0122\ \textit{mg/kg/d}$$

$$\text{Average Daily Dose (F)} = \frac{1\ \textit{mg/L} \times 1.043\ \textit{L/d}}{74.6\ \textit{kg}} = 0.0140\ \textit{mg/kg/d}$$

23. Ibid., Tables 6-1 and 3-23. The recommended average body weight for men and women combined is 80.0 kg (Table ES-1).
24. Males: (0.0286–0.0122)/0.0122=134%; Females: (0.0286-0.0140)/0.0140)=104%.
25. U.S. Environmental Protection Agency (2011a, Table 16-1 ["Time Outdoors (total)"]).
26. Ibid., Table 16-3. Estimated occupational mobility is lower for females in all age groups.

Using OSHA's default assumption overstates the duration of occupational exposure by a factor ranging from 8 to 20. Many risk and safety assessments assume 70 years of exposure at the same residential location, but the EFH reports that average residential occupancy period is 12 years.[27] The default overstates the duration of residential exposure by a factor of 5.8.

3. Empirical Estimates from the Risk Scenario of Interest

The best exposure data are obtained from the risk scenario of interest, whether occupational exposure to an ingredient in production, environmental exposure to an air pollutant such as particulate matter, or consumer exposure to a pathogen in food or water. This requires understanding the scenario, and quite likely, an investment of time and energy in data collection. Still, nothing beats real exposure data from real people who actually face the risk whose magnitude one is trying to estimate.

If the risk scenario of interest does not yet exist (for example, if it is forecast to result from building a new facility), empirical data can be simulated using a model specifically designed for this purpose. Simulation models used to be very difficult because they demanded expensive and rare computer power. With only few exceptions, that is no longer the case. But it is important to keep in mind that the results of simulation models are only as good as the models themselves and the assumptions embedded in them, and results are often impossible to test. For that reason alone, results from simulation models should be treated with circumspection.

A better approach than using average exposure is to use an empirical exposure distribution. This makes it possible in principle to produce empirical distributions of risk and benefits from regulation. Significant improvements in risk assessment methods would be required, of course, because as noted above in sections II.B through II.D, the current practice of hazard assessment does not yield average estimates.

4. Default Assumptions and Upper-Bounds Lead to "Cascading Conservatism"

Default assumptions should be used only when there are no empirical data from the scenario of interest and empirical data from a compendium such as the EFH are scientifically inappropriate. Moreover, it is also inappropriate to estimate risks and benefits using exposures to small fractions of the population, such as the 95th percentile, This overstates average exposure except in very rare cases, and when used in lieu of average exposure in an RIA, it will cause regulatory benefits to be overstated, possibly by a large amount.

When a series of upper-bound estimates is multiplied together, the result is a

27. Ibid., Table 16-5.

phenomenon often called "cascading conservatism."[28] The resulting risk figure is always unrealistically large and in some cases technically infeasible, and always carries the threat of distorting decisionmaking.

G. Properly Interpreting a Risk Assessment

EPA IS THE federal agency with the most transparent practices and the most influence, both with respect to the federal government generally and other actors such as the states. Because EPA is a major producer of risk assessments, textbooks often rely on EPA methods to describe methods generally.[29] Therefore, it is reasonable to consider EPA's practices as characteristic and illustrative of the field in general, even if other federal agencies' practices are not identical.

1. Risk Assessments Are, by Design, Inaccurate and Imprecise

EPA'S CANCER RISK assessment guidelines include a list of "major default options commonly employed when data are missing or sufficiently uncertain in a cancer risk assessment."[30] Although it might not be successful in every instance, the intent of each default is to ensure that risk is "not knowingly underestimated or grossly overestimated."[31] That means risk assessments are inaccurate by design, and guidelines impose a burden of certainty that default assumptions do not have to satisfy.

Risk assessments also are much less precise than advertised. To see why, consider a very high-power telephoto lens aimed at a far off object. A high-quality lens can provide extraordinary detail, but the quality of the picture depends on whether the photographer's hands shake. Sometimes, even the smallest disturbance means the picture won't capture the intended object. If the lens is not of high quality in the first place, or better stated, if its optics are fine up close but deteriorate as the lens is extended to its full telephoto distance, then the resulting photograph will be blurry even with a tripod.

In a similar way, the extrapolation from very high to very low doses causes the risk assessment "camera" to produce blurry images. Even if risk is estimated with extraordinary precision up close, much of that precision disappears as a result of extrapolation to a range of doses that are a tenth, a hundredth, or even a thousandth as large. The precision of a risk estimate will decline even more if a low-quality "lens" was used to estimate risk up close.

28. See, e.g., Council of Economic Advisors (2004, 179-140). Though it is widely used, the term "cascading conservatism" is technically problematic because the term "conservatism" is inherently ambiguous ("conservative" with respect to what?), and thus subject to serious misinterpretation and misunderstanding. A technically more accurate term would be "cascading bias."When each of several terms in a point estimate of risk is upwardly biased, the point estimate is biased by the product of the biases, hence, a cascading bias.
29. Faustman and Omenn (2001).
30. U.S. Environmental Protection Agency (2005a, Appendix A; reprinted as the appendix to this paper).
31. U.S. Environmental Protection Agency (2004, 13).

Human health risk assessments usually do not reveal the quality of the "camera" and "lens" being used. Often, they do not even clearly disclose how much magnification is being made. In almost all cases, the amount of precision reported will be much greater than can be justified based on the quality of the risk assessment camera and the lens.

2. Risk Assessments Can Be Undeclared Risk Management Decisions

Extrapolating from high to low doses from animals to humans using 95th percentile upper-confidence levels ensures that low-dose risk is not underestimated. Of course, this also means it is almost certain to overestimate low-dose risk. This practice is done by scientists, but it is not justified by any scientific principle or theory. It is done because it reflects policy views widely (but not universally) shared among scientists who perform human health risk assessment. As a group, they tend to believe that the costs of erring on the side of overestimating risk are small but the costs of erring on the side of underestimating risk are large. Many risk assessors also believe it is part of their professional responsibility to make sure risk is not underestimated.

The practice of purposefully overestimating risk has important policy consequences. For example, it reduces the discretion Congress and agency heads have to make decisions with which professional risk assessors disagree. If Congress and agency heads do not follow the ostensibly scientific advice of the "experts," it appears as if they are abandoning science in favor of political considerations.

Also, purposefully overestimating risk gives a scientific patina to policy decisions that otherwise might be controversial. For their part, many policymakers and agency heads are quite happy to relinquish these decisions to the "experts" because it enables them to avoid such controversies.[32]

3. Risk Assessment Practices Do Not Allow Risks to Be Accurately Ranked in Order of Severity

If risk assessments overestimated risk by the same amount, it would be simple to improve their accuracy: just multiply by a constant percentage that captures the constant level of bias. This list of substances from highest to lowest risk would still produce a correct risk ranking even though it would not accurately describe risk magnitudes. Agency heads and the public would still be able to make more informed decisions concerning which risks to address first.

Unfortunately, the amount by which average risk is overestimated varies, often by a lot. That means we cannot simply multiply by a constant proportion to obtain more accurate estimates. It also means any list of substances ranked from highest to

32. Wagner (1995).

lowest risk estimate is not a valid or reliable indicator of relative risk. Substances at the top of the list might be there only because their estimates are more biased, not because they pose the highest risks. Similarly, substances at the bottom of the list might pose the greatest risks but appear to be the least risky because their estimates are the least biased.[33]

III. WHAT IS A SAFETY ASSESSMENT?

Safety assessments give a quantitative estimate of the amount of a substance people can be exposed to and still be considered safe. Ideally, agency safety assessments would strive to identify the highest exposure or dose that is safe. This does not necessarily happen, however, because safety assessments include a host of undisclosed value judgments concerning the definition of safety. Value judgments, while always necessary for making decisions, are not scientific. For that matter, neither is the concept of safety. People can reasonably disagree about how much protection is safe enough, but it is a question science cannot answer. Instead, it must be answered based on societal values, and for that reason it is supposed to be the province of policymakers and appointed agency heads.

A. What Does a Safety Assessment Look Like?

For almost every human health risk other than cancer a large majority of analyses commonly called risk assessments are actually safety assessments. With usually modest effort, they are easy to differentiate: risk assessments estimate the risk caused by exposure to a substance, activity, or behavior; safety assessments estimate how much exposure to the substance, activity, or behavior is deemed safe.

Each federal agency has a different name for its safety assessments. At the EPA, for example, the principal safety assessments are called the reference dose (RfD) and the reference concentration (RfC). Similar thresholds are calculated by the Food and Drug Administration (the acceptable daily intake, or ADI) and the Agency for Toxic Substances Control Registry (the minimal risk level, or MRL).

Regardless of what it is called, a safety assessment will have three primary components:

- **An estimate of the lowest amount of the substance that has been shown to cause an adverse effect, usually in laboratory animals.** This is called a No Observed Adverse Effect Level (NOAEL) for the dose at which no adverse

33. As noted in footnote 19, in rare cases standard methods could underestimate average risk even though overestimation is intended. This would make conventional risk rankings even less accurate and reliable.

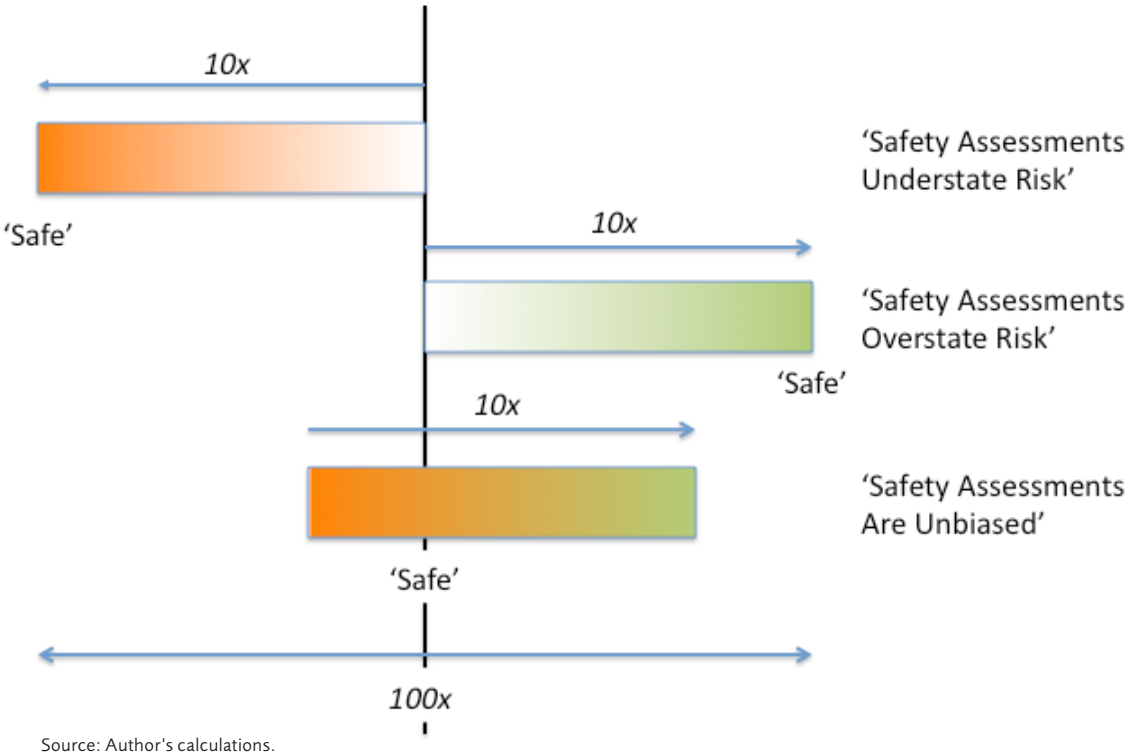
effect was observed or a Lowest Observed Adverse Effect Level (LOAEL) for the lowest dose at which an adverse effect was observed.[34]

- **As few as one and as many as five safety factors**.[35] The default value of each safety factor is 10 and it may be reduced to 3. Multiplying them together produces the composite safety factor. It is most often 100, but it can range from 10 to 10,000.
- **A threshold (such as an RfD, ADI, or MRL) calculated as the NOAEL or LOAEL divided by the composite safety factors.** The larger the composite safety factor, the lower the resulting safety threshold.

The resulting product is a point estimate of safety, which different individuals and entities may interpret very differently. For example, the definition of the Reference Dose says it is "an estimate (with uncertainty spanning perhaps an order of magnitude) of a daily or continuous exposure."[36] However, the location of this order of magnitude of uncertainty is not clearly defined, and Figure B shows alternative ways it can be interpreted. For example, risk managers who desire to exercise the greatest precaution might believe that the calculated safety threshold should never be exceeded because the methods used to derive it tend to be insufficiently precautionary. These individuals are illustrated at the top of the Figure. Others may believe that safety assessment methods are inherently biased so as to understate the highest dose that is safe. They believe exceeding the threshold by a factor of 10 poses no or negligibly greater risk. Finally, there are others who believe that safety assessment methods are as likely to under- or overstate the maximum safe dose. They are illustrated at the bottom of the Figure, with the order of magnitude uncertainty equally divided (in logarithms) on both sides.[37] This means, of course, that the amount of uncertainty contained within the definition of the "safe" dose is not an order of magnitude (10x), but rather two orders of magnitude (100x).

34. For simplicity in exposition, this paper refers only to NOAELs and LOAELs. Note that when the BMDL is used, the safety assessor has already imposed a bias by choosing the 95th percentile lower bound and thus abandoned any intent to derive a safety estimate based on an objective characterization of the available scientific evidence.
35. Safety assessors sometimes call them *uncertainty* factors. While uncertainty may be a legitimate scientific concern, safety assessors may be motivated as much or more by the desire to set safety thresholds cautiously than by a genuine lack of scientific knowledge. Precaution rather than scientific ignorance clearly is at work when an agency obtains new scientific information that reduces uncertainty but the uncertainty factor is not reduced or eliminated.
36. U.S. Environmental Protection Agency (2011e).
37. Felter and Dourson (1998).

FIGURE B: ALTERNATIVE WAYS TO INTERPRET A SAFETY ASSESSMENT



Source: Author's calculations.

B. The NOAEL or LOAEL Depends on the Choice of Study

Different studies often yield very different NOAELs or LOAELs. This means the starting point of every safety assessment is subject to the judgment of the safety assessor who decides which study to use.

Safety assessors often disagree about which study is "best" to use for choosing a NOAEL or LOAEL. Important factors will include such things as the species of laboratory animal chosen, the doses at which different groups of animals were exposed and the conditions of that exposure, interpretation of the test data, and numerous other technical factors. There are no consistent (or consistently applied) rules for deciding. It is inherently a matter of personal judgment. Safety assessors make different judgments, and these judgments are not always grounded solely in scientific concerns.

C. The Magnitude of the Composite Safety Factor Is Key

Regardless of the NOAEL or LOAEL, the choice of safety factors makes all the difference in determining the stringency of the safety assessment.

1. Extrapolating from a LOAEL Instead of a NOAEL

As long as they were obtained from the same study, the NOAEL (the dose at which no adverse effect was observed) is obviously safer than a LOAEL (the lowest dose at which an adverse effect was observed). Results across studies vary, however, so it is possible that the NOAEL obtained from one study is higher than the LOAEL from another. In many cases, a study can reveal only a LOAEL.

When a LOAEL is used, the default safe threshold is reduced by a factor of 10 to account for the unknown distance between the observed LOAEL and the unobserved NOAEL.

2. Extrapolating Animal Data to Humans

If the NOAEL or LOAEL (already adjusted downward by a factor of 10) comes from an animal study, the default safe threshold will be reduced by another factor of 10 to account for the possibility that humans are more sensitive than the most sensitive laboratory animal tested. Combined, these two default safety factors reduce the safe threshold by a factor of 100 if a LOAEL is used as the starting point and 10 if the starting point is a NOAEL. The composite safety factor will be 30 if either of the two factors is reduced from 10 to 3.[38] There are no scientific rules for deciding whether to reduce the default. Conventional guidance calls for the use of "sound scientific judgment" but does not define what this means.[39]

3. Extrapolating to the Most Susceptible Members of the Human Population

It is assumed that applying the first two safety factors gives us something like the average dose for the human population that is "safe." Of course, people vary, so to make sure the most susceptible members of the population are protected, the result will be divided by a third default safety factor of 10 (or 3). The composite safety factor is now 300 to 1,000, so the safe threshold is now 1/300 to 1/1,000 of the LOAEL or 1/100 of the NOAEL.

4. Extrapolating from Data Obtained in a Study with Less-Than-Lifetime Exposure

The typical purpose of a safety assessment is to decide what dose is safe if administered constantly over a lifetime. But the study from which the NOAEL or LOAEL was obtained may not have been based on lifetime exposure. In these cases, the composite safety factor will be divided again by 10 (or 3).

5. Extrapolation When the Database Is Incomplete

A final safety factor may also be applied if the safety assessor decides the database is incomplete. For example, if the safety assessment is based on animal data but a prenatal toxicity study and 2-generation reproduction study is missing, a safety

38. Three is approximately equal to the square root of 10 (i.e., 100.5 = 3.162).
39. U.S. Environmental Protection Agency (2002, xvii, 4-19, 4-23, 4-41, 5-5).

factor of 10 may be applied. The rationale is that either of these studies, had they been performed, could have discovered a lower NOAEL.

6. Capping the Composite Uncertainty

When all five default safety factors are used, the composite safety factor is 10,000. Safety assessors dislike composite safety factors greater than 1,000 because they produce results they regard as intuitively unreasonable or insufficiently scientific in appearance. For that reason, many safety assessors cap the composite safety factor at 3,000.[40]

Small differences of opinion among safety assessors are sufficient to change a safety assessment by a large amount. For example, if just one of the defaults is reduced from 10 to 3, the capped composite safety factor changes from 3,000 (10 × 10 × 10 × 3 × 1) to 900 (10 × 10 × 3 × 3 × 1). This difference may well be greater than the uncertainty in the LOAEL or NOAEL.

D. The Definition of an Adverse Effect Is Not Based Strictly on Science, and It May Be Unbounded

When safety assessors decide that a dose qualifies as a NOAEL or LOAEL, they also must make a decision concerning what health effect is "adverse." This can be obvious, such as when the effect in question seriously compromises human health. It also can be subtle, or even controversial. To see how this could happen, consider EPA's definition of adverse effect:

> A biochemical change, functional impairment, or pathologic lesion that affects the performance of the whole organism, or reduces an organism's ability to respond to an additional environmental challenge.[41]

The range of things that could "affect performance" or "reduce ability to respond" to something else is unbounded. Moreover, this definition includes no minimum level of performance reduction necessary to qualify. An effect could be deemed adverse even if it is not adverse in its own right, but its presence implies a greater likelihood of some future effect that could (but might not) be unambiguously adverse. Alternatively, the effect deemed adverse could be something that only signals that exposure occurred, or it could be something that might be caused by a host of unrelated phenomena. Sometimes, effects are implicitly deemed to be

40. U.S. Environmental Protection Agency (2002, 4-41): "The [internal EPA] Technical Panel recommends limiting the total [uncertainty factor] applied for any particular chemical to no more than 3000 and avoiding the derivation of a reference value that involves application of the full 10-fold UF in four or more areas of extrapolation."
41. U.S. Environmental Protection Agency (2011c).

adverse even though they are temporary, fully reversible, and too subtle to have been noticed.[42]

In fact, the term adverse effect has neither a scientific definition nor a legal one. It is used throughout toxicology and other fields closely related to risk assessment, but is nowhere defined scientifically. The U.S. Code and Code of Federal Regulations use the term hundreds and thousands of times, respectively, also without definition.[43]

E. Properly Interpreting a Safety Assessment

In section II.G we noted that risk assessments come with interpretive difficulties because they are designed to be inaccurate and imprecise, and they are intended to overstate the likely magnitude or severity of harm. Safety assessments suffer all these limitations, and more.

1. Safety Thresholds Are Less Precise Than They Appear

No matter how precise the estimate of a NOAEL or LOAEL, safety thresholds derived from them are highly imprecise. One can imagine a NOAEL or LOAEL confidently measured with four significant figures in the laboratory—say, 2.345 micrograms per kilogram of body weight per day ($\mu g/kg/d^{-1}$). When this precise figure is divided by a very large, round, and arbitrary number, say 100 or 1,000, the result has no precision at all. The safety threshold for the hypothetically administered laboratory dose mentioned above ranges from 0.2345 to 0.02345 $\mu g/kg/d^{-1}$. Dividing by large, round numbers discards the precision obtained in the laboratory. All that matters is which composite safety factor is chosen.

Safety thresholds usually will be reported with two significant figures (0.23 or 0.023 $\mu g/kg/d^{-1}$ in the given hypothetical scenario). This will create an incorrect impression of detailed scientific knowledge, for it draws attention to the last significant figure. If the second digit (the 3) in 0.023 $\mu g/kg/d^{-1}$ were actually meaningful, it would mean scientists had narrowed down the safe dose threshold to as little as one-half of one nanogram (ng).[44] Yet the scientific uncertainty actually resides in the first figure—is it 2 μg or 0.2 $\times g$ or 0.02 μg?—or between 20 and 2,000 nanograms.

42. See, for example, EPA's derivation of a Reference Dose for perchlorate, which is based on iodide uptake inhibition, an unambiguously nonadverse effect (U.S. Environmental Protection Agency 2005b). EPA designated iodide uptake inhibition as the so-called "critical effect" (U.S. Environmental Protection Agency 2011d, "The first adverse effect, or its known precursor, that occurs to the most sensitive species as the dose rate of an agent increases") for purely risk management reasons that made science a superfluous component of safety assessment ("Using a nonadverse effect that is upstream of the adverse effect is a more conservative and health-protective approach to perchlorate hazard assessment").
43. Stansell and Marvelli (2005).
44. 0.023 ($\pm$0.0005) $\times g$ = 23 ($\pm$0.5) ng.

2. Safety Thresholds Are Undeclared Risk Management Decisions

The purpose of dividing a NOAEL or LOAEL by safety factors is to provide additional assurance that that the resulting safety threshold is really, truly, clearly safe. How much precaution to use in decision making is by law left to agency officials, however, who often defer to agency safety assessors. This may occur because the officials do not understand safety assessment, mistakenly believing such assessments to be objective, or it may occur because they prefer highly precautionary risk management choices and the prevailing practices in the field allow them to mischaracterize their policy preferences as scientific.

3. Safety Assessments Do Not Allow Risks to Be Ranked

Safety assessments contain relatively little information even when extraordinary efforts have been made to avoid contaminating science with policy. The most a safety assessment can say is that exposures below the threshold are safe, whatever the agency decides safe means. Further, the amount of precaution built into safety assessments varies depending on the composite safety factor. If the safety thresholds for a hundred substances were ranked from lowest to highest, it would not be true that the substances with lower thresholds were riskier than substances with higher thresholds. Finally, safety assessments never have anything useful to say about whether exposures or doses above a threshold are risky, even though that is exactly how they are interpreted by risk managers, the press, and the public.

This is true even before safety factors have been included. The typical NOAEL applies to a laboratory animal such as the mouse or rat, not to people, and extrapolating these results to humans is not scientifically straightforward. Even when a NOAEL or LOAEL has been obtained from a study in humans, its significance depends on what phenomenon is deemed adverse. Obviously, some adverse effects used for safety assessment are much worse than others while some effects are not really adverse at all.[45]

4. Safety Assessment Language Can Be Confusing

Federal agencies do not use the same language in safety assessment for a number of reasons. Each agency has developed its own practices, in some cases because of different statutory language. Scientists from different agencies (and even scientists within agencies) often disagree about the desired composite safety factor because they disagree about how precautionary the safety assessment ought to be. This can be a significant source of interagency conflict.[46]

45. See footnote 42.
46. The best known example may be the longstanding conflict between the EPA and the Food and Drug Administration over the safety assessment for methylmercury, hinted at by Goldman (2001).

5. Safety Assessments and RIAs

All safety assessments have two common limitations relevant to the preparation of RIAs:

- Safety assessments cannot be used to estimate the size of a risk, and thus help decisionmakers determine whether it warrants regulatory action.
- Safety assessments cannot be used to estimate the benefits of a regulatory action.

These limitations apply regardless of how much or how little risk management precaution is embedded in them. The only potentially relevant information a safety assessment produces is the estimated number of people exposed above the threshold before a regulation is promulgated and the estimated number of people whose exposure would move from above to below this threshold. Even if these quantities are estimated with the greatest possible scientific objectivity, neither of them can be used for estimating benefits because there is no way to translate these numbers into reductions in human health risk.

IV. RISK AND SAFETY ASSESSMENTS AS INPUTS TO RIAS

An RIA will use a risk assessment (but not a safety assessment) two ways. First, it will provide an estimate of the human health risks expected to occur if the government takes no regulatory action. Because many conventionally performed risk assessments are intended to ensure that risk is not underestimated, an RIA that relies on such a risk assessment will begin from an erroneous starting point, one that misrepresents the risk problem as larger than it probably is.

Second, an RIA will use a risk assessment to estimate the amount by which a risk is expected to be reduced by regulation. Whatever biases are in the risk assessment will be propagated through the benefits assessment. For example, if the risk assessment overstates the likely risk and the assessment is used to represent the likely risk, it will be assured of overstating the likely benefits of regulation.

Finally, RIAs are supposed to convert risk-reduction benefits into dollar equivalents. Whether this is done depends in large part on whether there are plausible estimates of the value consumers place on avoiding the health risk in question. The practices used to monetize benefits also have the potential to distort, especially if upper-bound valuations for health effects are used instead of best estimates.[47]

47. See subsection II.E regarding the use of upper bounds. Controversies surrounding the valuation of health risks are very important and should not be ignored. Regulatory agencies have strong incentives to place the highest possible values on avoiding each case of illness or premature mortality. These controversies are beyond the scope of this paper.

A. Why Must Risk Assessments Be Objective?

RIAs are supposed to give decisionmakers and the public an accurate portrayal of the consequences of taking regulatory action, or taking no action at all. Properly performed, economic analysis does not steer decisionmakers in any particular direction, however. It simply sets forth economists' best ex ante estimates of the effects of alternative choices, including the option of doing nothing. For this reason, RIAs must be prepared objectively, consistent with widely accepted economic principles, and free of policy advocacy. If RIAs lack objectivity, violate accepted economic principles, or engage in policy advocacy, then decisionmakers and the public have every reason to distrust them. When RIAs rely on risk assessments to estimate the magnitude of a risk, it is therefore essential that risk assessments also be objective and free of policy advocacy.

The term objectivity has many different meanings. The meaning most relevant to risk assessment and RIAs is one the OMB issued in 2002 to comply with a statutory directive colloquially called the Information Quality Act.[48] OMB's definition has two components:

- **Substantive objectivity,** defined as being "accurate, reliable, and unbiased."[49]
- **Presentational objectivity,** defined as "presented in an accurate, clear, complete, and unbiased manner" and "within a proper context."[50]

Congress directed OMB to define objectivity (and a few other terms) in order to improve the quality of information disseminated by federal agencies. That is, Congress decided objectivity was an integral element of quality, and it delegated to OMB the job of defining it in a way that could be implemented.

If an RIA depends on a risk assessment but that risk assessment is not objective, the RIA will not portray benefits and costs objectively. Objectivity has been an implicitly required attribute of RIAs since at least 1990, and it remains so today.[51] However, compliance with OMB guidance has been generally spotty.[52] There are significant deficiencies in the regulatory review process that enable agencies to

48. Sec. 515 of Pub. L. 106-554, codified at 44 U.S.C. 3516 note. For OMB's government-wide guidance implementing the law, see Office of Management and Budget (2002).
49. Office of Management and Budget (2002, p. 8459, Sec. V.3.b).
50. Office of Management and Budget (2002, p. 8459, Sec. V.3.a).
51. Office of Management and Budget (2003).
52. In an analysis of federal agency compliance with longstanding regulatory analysis requirements, the average RIA received scores consistent with having "some relevant discussion with some documentation of analysis." See Ellig (2011).

prepare RIAs, and the underlying risk assessments, that are not objective portrayals of risks, benefits, and costs.[53]

B. The Difference among Precision, Accuracy, and Bias

The concepts of precision, accuracy, and bias are often confused. Precision refers to how close measurements are in repeated tests using the same measurement instrument. One instrument is considered more precise than another if it has less measurement error. Accuracy is different; it refers to how close a measurement is to the thing being measured. Measurements can be very precise but inaccurate. This can occur, for example, if a measurement instrument is not properly calibrated. Bias occurs when measurements are inaccurate in a systematic way.

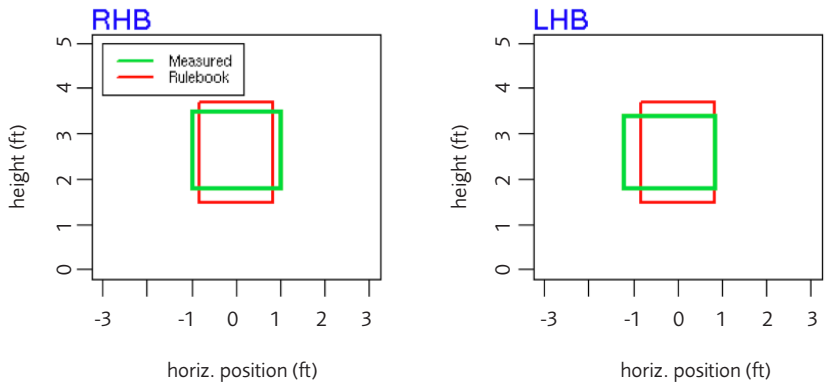
Consider the task of calling balls and strikes in baseball. The definition of a strike is stated in language that is supposed to be precise but is in fact somewhat ambiguous.[54] It is difficult to implement, especially at the edges. Umpires have only a fraction of a second to evaluate pitches traveling up to 100 miles per hour, often with substantial movement. The definition of a Strike is stated in two dimensions, but the strike zone has three.[55]

With the recent addition of electronic strike zone measurement, umpires' actual performance can be evaluated to determine their accuracy (how closely their calls match the definition in the rule book) and determine if their calls are biased. A recent review of about 80,000 pitches shows that the actual strike zone for right-handed batters, as interpreted by umpires, is on average about two inches wider on each side of home plate than specified in the rule book.[56] While it is inaccurate, it is not biased because umpires err about the same amount on both sides of the plate.

This is not the case for left-handed batters, however. The actual strike zone for left-handed batters matches the rule book on the inside edge of home plate but

53. Under section 6(a)(3(C)(iii) of Executive Order 12866, OMB reviews RIAs when they are submitted along with draft proposed and final rules. When RIAs are submitted along with draft rules, they are more likely intended to support decisions rather than to inform decisionmaking. Also, if substantial analytic errors have resulted in biased benefit and cost estimates, it is usually too late in the regulatory development process to make corrections.
54. Major League Baseball (2011, Rule 2.00): "The STRIKE ZONE is that area over home plate the upper limit of which is a horizontal line at the midpoint between the top of the shoulders and the top of the uniform pants, and the lower level is a line at the hollow beneath the kneecap. The Strike Zone shall be determined from the batter's stance as the batter is prepared to swing at a pitched ball." It is not obvious where exactly is the "midpoint between the top of the shoulders and the top of the uniform pants," much less how to interpret "the batter's stance as the batter is prepared to swing at a pitched ball" when the umpire only calls a strike if the batter does not swing.
55. Home plate is 17" wide by 17" deep, with 8-1/2" right triangles removed on each back side. Thus, while the definition of a Strike has width and height, a ball is a Strike if it passes through any part of the implied 3-dimensional space.
56. Walsh (Walsh 2007b, 2007a).

FIGURE C: MEASURED VS. RULEBOOK STRIKE ZONE



Actual vs. Rulebook Strike Zone Dimensions (inches)

	Left	Right	Lower*	Upper	Total Area+
RHB	-12.0	12.1	21.6	42.0	492
LHB	-14.6	9.9	21.5	40.8	475
Rulebook	-9.9	9.9	17.7	44.2	527

*vertical strike zone mapped to average

+total area in square inches

Source: John Walsh, "The Eye of the Umpire," *Hardball Times*, July 25, 2007, http://www.hardballtimes.com/main/article/the-eye-of-the-umpire.
Note: The strike zone is presented from the catcher's perspective.

extends about 4-1/2 inches beyond home plate on the outside edge. This is an example of measurement bias.[57]

Like calling balls and strikes, risk assessments have varying degrees of precision and accuracy, but there is one very important difference. Baseball umpires try to measure the strike zone objectively to preserve their own credibility and the credibility of the game.[58] In risk assessment, however, umpires (that is, risk and safety assessors) routinely call balls and strikes in a purposefully biased manner.[59]

Both umpiring and risk assessment should be as precise as possible, with precision reported based on the limits of the measurement instrument. In practice, however, risk and safety assessments are reported with far more precision than is

57. The same data set shows that umpires also have biased measurements of the height of the strike zone. For right-handed batters, umpires on average miss the top 2" and 3" at the bottom of the strike zone. For left-handed batters, umpires miss the top 2" on both the top and bottom. The likely reason for these biases is umpires set up behind the catcher in the same position irrespective of whether the batter is right- or left-handed.
58. The results presented by Walsh are averages that disguise substantial variation across umpires. Variation has been estimated to be 10 extra strikes per game, virtually all of them at the edges of the rule book strike zone (Hale 2007). This and other biases are not observable when umpires know their performance is subject to scrutiny by electronic means (Parsons et al., 2011).
59. See footnote 31 and the accompanying text.

justified based on the underlying data. Risk and safety assessors report their work as if they are calling balls and strikes with precision laser measurements when all they really have is ordinary eyesight at best.

More important than precision, however, is accuracy—the absence of bias. If risk assessors intentionally estimate risk in a biased manner, they are like umpires who call strikes based on whether they personally like the pitcher or his team better than the batter and his team. Baseball umpires caught doing this would be fired, for the credibility of the game is lost if they are not impartial. No such rules apply to federal agency risk and safety assessors.

C. Why Excess Precision Is a Problem

Excess precision is evident when an analysis reports results with too many digits than can be justified by the precision of the underlying data. When performing division, for example, handheld calculators and computers can generate as many digits as one desires. This does not mean that all of them are meaningful. Here's an example:

> Psychology students were training rats to run mazes. In the final report, they noted "33.3333% of the rats learned to run the maze. 33.3333% of the rats failed to learn. And the third rat escaped."[60]

The first two percentages imply that 1 million rats were tested. The third admits that the true number was three.

Excess precision was not a large problem when analysts did their calculations by hand or even by slide rule. Computers made it a problem and the advent of inexpensive personal computers has made the problem endemic. The reason this happens is that calculators and computers treat numbers as fixed and certain things, but in the world of estimation numbers are variable and uncertain.

When estimates are presented properly, the last digit is uncertain. In the estimate 2.72, for example, the first two digits are certain but the third is not. It is rounded off. Any estimate between 2.715 and 2.724 can be represented as 2.72. An additional decimal place is required to indicate greater precision (for example, 2.718), and the last one must be removed if the measurement really is less precise (for example, 2.7). The fact that a calculator or computer is capable of reporting 8, 16, or 256 digits is immaterial. Only the number of digits that reflect the true precision of an estimate should be reported.

In risk assessment and economic analysis, estimates with excess precision violate the principle of presentational objectivity. Uncertainty about a risk or benefit estimate is one of the most important elements of proper context, and excess precision

60. Grumbine (2008).

underreports uncertainty. They are not "presented in an accurate, clear, complete, and unbiased manner" and "within a proper context."[61]

Excess precision is also misleading, for it subtly communicates to decision makers and the public that analysts know more about the size of a problem and the consequences of regulating than they really do. Consider the example of the safety assessment discussed in section III.B above. It consists of a NOAEL or LOAEL divided by a series of default safety factors that, in combination, range from 10 to 10,000. Each safety factor in the combination is an arbitrary round number almost always lacking empirical support for its precision and often for its presumed order of magnitude. No matter how precise the NOAEL or LOAEL, when it is divided by a series of highly uncertain, arbitrary round numbers the result should be reported as a highly uncertain, arbitrary small number. Instead of disclosing how imprecise the resulting safety threshold actually is, federal agencies present safety thresholds as if they are extraordinarily precise.

D. Why Objectivity in Risk Assessment Is Essential

Accuracy, clarity, completeness, and unbiasedness are attributes of objectivity. Objective risk estimates are what economists need to prepare objective estimates of the benefits and costs of a proposed regulation. Agency officials—those to whom Congress has delegated the responsibility for making countless policy decisions—need objective risk assessments and objective RIAs.[62] This is true even if they have been directed by Congress to write regulations that protect much more than just the average person in the population. How protective a regulation ought to be is a completely different question than whether risk should be estimated objectively.

Without objective risk assessments and RIAs, there is no way for agency policy officials to know whether the regulatory option they select is expected to achieve the amount of precaution they intend. If a risk assessment understates the likely size or severity of a risk, then the policy official will choose an option that is not as stringent as he intends. Conversely, if the risk estimate overstates the likely size or severity of a risk, then the official will choose an option that is more stringent than intended.

When risk assessments are not objective and vary in the degree to which they under- or overstate risk, policy officials can neither use them to correctly rank the risks they are supposed to manage nor be assured of selecting their desired degree of precaution for the risk they are addressing. Risks that appear to be large may

61. Office of Management and Budget (2002, 8459).
62. This paper focuses on how risk assessments lacking objectivity result in benefit assessments lacking objectivity, too. The same phenomenon can and does occur in cost estimation., of course. An obvious reason why cost estimates lack objectivity is agencies rarely estimate cost in conceptually correct ways. Almost every RIA estimates only compliance cost, not opportunity cost. OMB Circular A-4 requires the estimation of opportunity cost, but OMB in unable or unwilling to enforce this requirement.

be relatively small if estimated objectively. Risks that appear to be small may be relatively large if estimated objectively.

The same principle applies to RIAs. Regulations can be ranked from highest to lowest in reported benefits, but if the underlying risk assessments are inaccurate by different amounts—and especially if they are inaccurate in different directions—any rank ordering of regulations from most to least beneficial is almost certain to be wrong.

Finally, if the policy choices involve trying to reduce a risk in a way that may increase another risk, something called a risk/risk tradeoff arises.[63] The net change risk is what really matters, but net risk often is not even part of the discussion because the countervailing risk is not estimated and sometimes not even acknowledged.

V. CONCLUSION

Risk assessment is an essential input into RIAs, for without them it is not possible to estimate the benefits of regulation. Safety assessment cannot be used in RIAs, however, because they do not express risk in a manner that is conceptually compatible with benefit-cost analysis.

To obtain an objective estimate of regulatory benefits, economists need objective estimates of risk. Unfortunately, conventional risk assessment methods rarely, if ever, produce objective estimates. If a risk assessment lacks objectivity because it understates risk, the resulting benefits estimate will understate the true benefits. Conversely, if a risk assessment lacks objectivity because it overstates risk, benefits will be overstated, too.

This paper shows that on every margin where scientific knowledge is uncertain, it is the intent of conventional risk assessment to be sure risk is not understated. This means that in the vast majority of cases, risk assessment is very likely to overstate risk. RIAs based on these risk assessments will overstate the benefits of regulation.

63. Graham and Wiener (1995).

APPENDIX: MAJOR DEFAULT ASSUMPTIONS IN RISK ASSESSMENT[64]

ISSUE	DEFAULT
Is the presence or absence of effects observed in a human population predictive of effects in another exposed human population?	• When cancer effects in exposed humans are attributed to exposure to an agent, the default option is that the resulting data are predictive of cancer in any other exposed human population. • When cancer effects are not found in an exposed human population, this information by itself is not generally sufficient to conclude that the agent poses no carcinogenic hazard to this or other populations of potentially exposed humans, including susceptible subpopulations or lifestages.
Is the presence or absence of effects observed in an animal population predictive of effects in exposed humans?	• The default option is that positive effects in animal cancer studies indicate that the agent under study can have carcinogenic potential in humans. • In general, while effects seen at the highest dose tested are assumed to be appropriate for assessment, it is necessary that the experimental conditions be scrutinized. • When cancer effects are not found in well-conducted animal cancer studies in two or more appropriate species and other information does not support the carcinogenic potential of the agent, these data provide a basis for concluding that the agent is not likely to possess human carcinogenic potential, in the absence of human data to the contrary. • Target organ concordance is not a prerequisite for evaluating the implications of animal study results for humans. • The default is to include benign tumors observed in animal studies in the assessment of animal tumor incidence, if such tumors have the capacity to progress to the malignancies with which they are associated. • Benign tumors that are not observed to progress to malignancy are assessed on a case- by-case basis.
How do metabolic pathways relate across species and among different age groups and between sexes in humans?	• The default option is that there is a similarity of the basic pathways of metabolism and the occurrence of metabolites in tissues in regard to the species-to-species extrapolation of cancer hazard and risk.
How do toxicokinetic processes relate across species and among different age groups and between sexes in humans?	• As a default for oral exposure, a human equivalent dose for adults is estimated from data on another species by an adjustment of animal applied oral dose by a scaling factor based on body weight to the 3/4 power. The same factor is used for children because it is slightly more protective than using children's body weight. • For inhalation exposure, a human equivalent dose for adults is estimated by default methodologies that provide estimates of lung deposition and internal dose. • For a route-to-route exposure extrapolation, the default option is that an agent that causes internal tumors by one route of exposure will be carcinogenic by another route if it is absorbed by the second route to give an internal dose.
What is the correlation of the observed dose-response relationship to the relationship at lower doses?	• The default procedure for the observed range of data when a biologically based model is not used is to use a curve-fitting model for incidence data. • A linear extrapolation approach is used when the mode of action information is supportive of linearity or mode of action is not understood. • When adequate data on mode of action provide sufficient evidence to support a nonlinear mode of action for the general population and/or any subpopulations of concern, a different approach—a reference dose/reference concentration that assumes that nonlinearity—is used. • When the mode of action information indicates that the dose-response function may be adequately described by both a linear and a nonlinear approach, then the results of both the linear and the nonlinear analyses are presented. • Absent data to the contrary, the default assumption is that the cumulative dose received over a lifetime, expressed as a lifetime average daily dose or lifetime average daily exposure, is an appropriate measure of dose or exposure.

64. See the "Apples and Oranges Problem," in OMB, *Report to Congress on the Costs and Benefits of Federal Regulations* (Washington, DC: OMB, September 30, 1997), http://www.whitehouse.gov/omb/inforeg/rcongress.html.

REFERENCES

American Cancer Society. 2011a. *Cancer Prevalence: How Many People Have Cancer?* (October 18). Accessed February 1, 2012. Available from http://www.cancer.org/Cancer/CancerBasics/cancer-prevalence (Archived by WebCite® at http://www.webcitation.org/6AO5ubcHe).

———. 2011b. *Lifetime Risk of Developing or Dying From Cancer* (December 22). Accessed February 1, 2012. Available from http://www.cancer.org/Cancer/CancerBasics/lifetime-probability-of-developing-or-dying-from-cancer (Archived by WebCite® at http://www.webcitation.org/6AO5mwsnc).

Bipartisan Policy Center. 2009. *Improving the Use of Science in Regulatory Policy*. Washington, DC: Bipartisan Policy Center.

Council of Economic Advisors. 2004. *Economic Report of the President*. Washington, DC: Government Printing Office.

Ellig, Jerry. 2011. High-Impact Decisions: Report Card Grades Regulators. *The Public Manager* (Fall): 60-65.

Faustman, Elaine M., and Gilbert S. Omenn. 2001. Risk Assessment. In *Casarett & Doull's Toxicology: The Basic Science of Poisons*, edited by C. D. Klaassen. New York, NY: McGraw-Hill Medical Publishing Division.

Felter, S.P., and M.L. Dourson. 1998. The Inexact Science of Risk Assessment (and Implications for Risk Management). *Human and Ecological Risk Assessment* 4 (2): 245-51.

Goldman, Lynn R. 2001. Epidemiology in the Regulatory Arena. *American Journal of Epidemiology* 154 (12 (Supplement)): S18-S26.

Graham, John D., and Jonathan Baert Wiener, eds. 1995. *Risk vs. Risk: Tradeoffs in Protecting Health and the Environment*. Cambridge, MA: Harvard University Press.

Grumbine, Robert. 2008. *Excess Precision* (Sepember 21). Accessed February 9, 2012. Available from http://moregrumbinescience.blogspot.com/2008/09/excess-precision.html.

Hale, Jonathan. 2007. *A Zone of Their Own* (November 28). Accessed February 7, 2012. Available from http://www.hardballtimes.com/main/article/a-zone-of-their-own/ (Archived by WebCite® at http://www.webcitation.org/6AL1aLN8s).

Major League Baseball. 2011. *Official Rules*. Major League Baseball(February 16). Accessed Feburary 6, 2012. Available from http://mlb.mlb.com/mlb/downloads/y2011/Official_Baseball_Rules.pdf.

National Research Council. 1983. *Risk Assessment in the Federal Government: Managing the Process.* Washington, DC: National Academies Press.

———. 2005. *Health Implications of Perchlorate Ingestion*. Washington, DC: National Academy Press.

North, D. Warner. 2003. Reflections on the Red/Mis-Read Book, 20 Years After. *Journal of Human and Ecological Risk Assessment* 9 (5):1145-54.

Office of Management and Budget. 1990. Regulatory Impact Analysis Guidance (Appendix V) In *Regulatory Program of the United States Government, April 1, 1990–March 31, 1991*. Washington, DC: Office of Management and Budget.

———. 1996. *Economic Analysis of Federal Regulation Under Executive Order* 12866 (January 11). Accessed February 1, 2012. Available from http://www.whitehouse.gov/omb/inforeg_riaguide.

———. 2000. *Guidelines to Standardize Measures of Costs and Benefits and the Format of Accounting Statements* (March 22). Accessed June 10, 2006. Available from http://www.whitehouse.gov/sites/default/files/omb/assets/omb/memoranda/m00-08.pdf.

———. 2002. Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies; Notice; Republication. *Federal Register* 67 (36):8452-60.

———. 2003. *Circular A-4: Regulatory Analysis* (September 17). Accessed April 6, 2011. Available from http://www.whitehouse.gov/omb/circulars_a004_a-4.

Parsons, Christopher A., Johan Sulaeman, Michael C. Yates, and Daniel S. Hamermesh. 2011. Strike Three: Discrimination, Incentives, and Evaluation. *American Economic Review* 101 (4):1410-35.

Stansell, Kelsey, and Mark Marvelli. 2005. 'Adverse Effects' and Similar Terms in U.S. Law: A Report Prepared by the Duke Center for Environmental Solutions for the Dose Response Specialty Group of the Society for Risk Analysis (SRA). Durham, NC: Duke University Center for Environmental Solutions.

U.S. Environmental Protection Agency. 2002. A Review of the Reference Dose and Reference Concentration Processes. Washington, DC: Risk Assessment Forum, U.S. Environmental Protection Agency.

———. 2005a. Guidelines for Carcinogen Risk Assessment. In *EPA/630/P-03/001F*. Washington, DC: U.S. Environmental Protection Agency Risk Assessment Forum.

———. 2005b. *Perchlorate and Perchlorate Salts.* Integrated Risk Information System, Office of Research and Development (February 18, 2005). Accessed August 30, 2005. Available from http://www.epa.gov/iris/subst/1007.htm.

———. 2011a. *Exposure Factors Handbook: 2011 Edition*. Washngton, DC: U.S, Environmental Protection Agency Office of Research and Development.

———. 2011b. *Glossary of IRIS Terms.* Integrated Risk Information System, Office of Research and Development (July 26). Accessed February 6, 2012. Available from http://www.epa.gov/iris/help_gloss.htm - content.

———. 2011c. *IRIS Glossary; 'Adverse Effect'* (July 26). Accessed February 6, 2012. Available from http://www.epa.gov/iris/help_gloss.htm - a.

———. 2011d. *IRIS Glossary; 'Critical Effect'* (July 26). Accessed February 6, 2012. Available from http://www.epa.gov/iris/help_gloss.htm - c.

———. 2011e. *IRIS Glossary; 'Reference Dose'* (July 26). Accessed February 6, 2012. Available from http://www.epa.gov/iris/help_gloss.htm - r.

———. 2011f. *IRIS Glossary; 'Unit Risk'* (July 26). Accessed March 23, 2012. Available from http://www.epa.gov/iris/help_gloss.htm - u.

U.S. Environmental Protection Agency Office of the Science Advisor. 2004. "An Examination of EPA Risk Assessment Principles and Practices;" Staff Paper, EPA/100/B-04/001.

Wagner, Wendy E. 1995. The Science Charade in Toxic Risk Regulation. *Columbia Law Review* 95 (7):1613-1723.

Walsh, John. 2007a. *The Eye of the Umpire* (July 25). Accessed April 2, 2012. Available from http://www.hardballtimes.com/main/article/the-eye-of-the-umpire/ (Archived by WebCite® at http://www.webcitation.org/6AL17wUBo).

———. 2007b. *Strike Zone: Fact vs. Fiction* (July 11). Accessed April 2, 2012. Available from http://www.hardballtimes.com/main/article/strike-zone-fact-vs-fiction/ (Archived by WebCite® at http://www.webcitation.org/6AL0ssnFV).