

## Commentary

# Clarifying Types of Uncertainty: When Are Models Accurate, and Uncertainties Small?

Louis Anthony (Tony) Cox, Jr.\*

---

Professor Aven has recently noted the importance of clarifying the meaning of terms such as “scientific uncertainty” for use in risk management and policy decisions, such as when to trigger application of the precautionary principle. This comment examines some fundamental conceptual challenges for efforts to define “accurate” models and “small” input uncertainties by showing that increasing uncertainty in model inputs may reduce uncertainty in model outputs; that even correct models with “small” input uncertainties need not yield accurate or useful predictions for quantities of interest in risk management (such as the duration of an epidemic); and that accurate predictive models need not be accurate causal models.

---

**KEY WORDS:** Model uncertainty; precautionary principle; risk management; scientific uncertainty; uncertainty characterization

Professor Aven inquires what “scientific uncertainty” means, and whether and how it can be made clear enough to be useful in risk management, for example, in triggering application of the precautionary principle.<sup>(1)</sup> Although the precautionary principle itself might invite questions—after all, traditional decision analysis already explicitly prescribes taking action without requiring scientific (or any other kind of) certainty about causes or consequences, whenever the expected utility of doing so exceeds the expected utility of alternatives, such as waiting for more information to resolve current uncertainties—the excellent questions (a)–(e) raised in Professor Aven’s introduction have wider significance for risk analysis. They ask about the relation between “scientific uncertainty” and the size (large or small) of uncertainty in probability estimates and model inputs; the knowledge base (rich or poor) used to assign probabilities; the existence and reliability of quantitative risk assessment results; and abil-

ity to specify a space of possible outcomes, estimate meaningful frequentist probabilities, and establish an accurate causal prediction model of how underlying factors influence the probable consequences of an activity. Professor Aven’s Fig. 3 proposes that “scientific uncertainty” should be considered greatest when no accurate prediction model can be established, and least when one can be, and when uncertainties about its inputs are “small” in some relevant sense.

Attempting to clarify what key terms such as “uncertainty” and “scientific uncertainty” mean (or should mean) in risk analysis is surely desirable, and Professor Aven’s investigation makes a valuable start. However, I wonder whether this initial effort does not simply push back by one step the substantive challenge of constructing clear and useful definitions of key terms, so that instead of wrestling with what “scientific uncertainty” might mean, we must now grapple with what is (or should be, or coherently can be) meant by an “accurate” prediction model and by “small” uncertainty in its inputs. These are not necessarily easy questions to answer satisfactorily, as the following examples illustrate.

\*Address correspondence to Tony Cox, Cox Associates and University of Colorado, 503 Franklin St., Denver, CO 80218 USA; tcoxdenver@aol.com.

# 1. EXAMPLE: LARGER INPUT UNCERTAINTIES MAY CREATE SMALLER OUTPUT UNCERTAINTIES

Following Professor Aven's notation, consider a model  $Z = G(X)$ , where  $X$  is an uncertain input and  $Z$  is the model's output. For concreteness, suppose that  $X$  is a scalar input, uniformly distributed over some interval, and that  $G$  is a known function. Now, is it true that the "scientific uncertainty" about  $Z$  corresponding to (and caused by) an uncertain value of  $X$  should necessarily be a nondecreasing function of the level of uncertainty in  $X$ ? The following example suggests not. Presumably, most analysts (and all who use variance or entropy to define and measure the "uncertainty" of a probability distribution) would agree that  $X$  has "smaller uncertainty" if it is uniformly distributed between 98 and 100 than if it is uniformly distributed between 0 and 198. Yet, if  $G$  is the threshold function  $G(X) = 1$  for  $99 \leq X \leq 100$ , else  $G(X) = 0$ , then the uncertainty (e.g., variance or entropy) of  $Z = G(X)$  is greatest when  $X$  is uniformly distributed between 98 and 100 (since there are then equal probabilities of 50% each that  $Z$  will be 0 or 1), and is much smaller when  $X$  is uniformly distributed between 0 and 198 (since there is then a 99.5% probability that  $Z = 0$ ). So, larger uncertainty about  $X$  induces smaller uncertainty about the value of output  $Z$  caused by  $X$ . It is therefore unclear why "scientific uncertainty" should be assumed to be an increasing function of input uncertainty.

# 2. EXAMPLE: WHEN IS A PREDICTION MODEL "ACCURATE"?

Consider an epidemic that spreads in a population of size  $N$  individuals. Let  $I(t)$  denote the number of the infected at the start of week  $t$  and let  $k$  denote the infectivity of the epidemic. Assume that the epidemic dynamics are accurately described by the following deterministic (logistic difference) model for an SIS (susceptible–infected–susceptible) disease process:<sup>(2)</sup>  $I(t + 1) = k \times I(t) \times [N - I(t)]$ . More exactly, if each person is either infected or not at any time, so that  $I(t + 1)$  must have an integer value, then this continuous model can be rounded to give the following discrete model:

$$I(t + 1) = \text{round}(k \times I(t)[N - I(t)]), \quad (1)$$

where *round* denotes the rounding function that rounds numerical values to the closest integer. To obtain inputs to this model, suppose that a surveillance program is able to ascertain the fraction of the in-

fectured in the population at the start of any week to a high degree of precision, for example, three significant digits (better than most real-world surveillance programs). The initial infected fraction at the time that risk management of the epidemic begins is found to be 0.101. For simplicity, assume that both  $k$  and  $N$  are perfectly known:  $N = 100,000$  and  $k = 0.00004$ .

Before reading further, readers may wish to answer the following questions for themselves. (My answers, given in parentheses, are explained next.)

QUESTION 1: *If the model (1) correctly describes the true evolution of the epidemic with perfect accuracy, does this necessarily make it an "accurate prediction model," in the sense of Professor Aven (i.e., conducive to low scientific uncertainty)?*

QUESTION 2: *Might not the same model, with the same input data, be both an "accurate prediction model" for some scientific or risk management decision purposes, but an inaccurate one for others?*

(I think that the answer to Question 1 is No, and that whether a prediction model can be considered "accurate" depends on details of the risk management context and input data, rather than simply on whether the model itself maps input to output values in the same way that reality does. I think that the answer to Question 2 is Yes, and that this implies that any classification of prediction models as "accurate" or "inaccurate," or any attempted comparison or ranking of prediction models as more or less accurate, as suggested in Professor Aven's Fig. 3, may be problematic.)

To continue the example, consider the following risk management decision context. Suppose that if the epidemic can be reduced to fewer than 10 cases (i.e., 0.01% of the population) infected at any one time, then it can be eradicated. A question of great practical interest for planning the epidemic risk management program is: How many weeks (approximately) will the epidemic last before it becomes small enough to be eradicated? This is something that an "accurate prediction model" for risk analysis in this case might be expected to help answer. Despite the nearly perfect information available in this example, however, the answer to this prediction question depends on information that is lost in the rounding error (beyond the three significant digits available) for the initial infected fraction. For example, one can easily verify with a calculator that the answer is 16 weeks when the initial infected fraction is 0.10117, but 366 weeks when the initial infected fraction is 0.10118. These very different answers

correspond to the same measured initial infected fraction, to four decimal places. Thus, even with extremely accurate and precise (to four decimal places) knowledge of how this epidemic spreads, it is impossible to specify how long it will last to within the correct order of magnitude. In this sense, the prediction model is completely correct, but not at all accurate, even when it is used with inputs that are measured with only tiny uncertainty by any traditional measure (e.g., standard deviation, error range, confidence intervals, coefficient of variation, etc.). The mathematical reason is well understood for chaotic systems: two trajectories that start at almost the same point (e.g., initial infected fraction of 0.1010 vs. 0.1014, in this example) will remain close together (and hence predictable from approximately measured initial conditions) for a while, but begin to diverge widely (by about 10 weeks, in this example) as the dynamics of the system amplify the small differences (in the fourth decimal place and beyond) exponentially, until they are of the same order of magnitude as the phenomenon being studied, that is, total number of infected cases. The same thing would occur slightly later if even more precise measurements of initial conditions were available (e.g., after 11 weeks for the trajectories that start from initial infected fractions of 0.10117 and 0.10118). The event of interest—the time at which the epidemic ends—lies beyond the effective forecast horizon for this system. This horizon is determined by how quickly neighboring trajectories diverge, that is, how long it takes for small differences in initial conditions (differences lost in the rounding error in this example) to become amplified to the point where they drive the observed trajectory (a concept precisely defined and quantified by the system's Lyapunov exponent).<sup>(2)</sup> The general lesson illustrated by this example is that a prediction model can be both highly accurate for some purposes (e.g., for making predictions over short horizons), and highly inaccurate for others (e.g., making predictions over longer horizons). Therefore, ascribing “accuracy,” or relative degrees of “accuracy,” to a prediction model, apart from the detailed decision context or purpose for which it is being used, and separately from describing uncertainties in its inputs, is not necessarily possible.

### 3. EXAMPLE: AN ACCURATE RISK PREDICTION MODEL NEED NOT BE CAUSAL

Professor Aven emphasizes the importance of causality—and, indeed, it is crucial to understand

causality in order to accurately predict the effects on risk of proposed interventions that act by affecting its causes (e.g., exposures). However, many risks can be predicted accurately, for example, by empirical actuarial techniques, even without any understanding or model of causation. For example, suppose that the lifetimes of a certain type of electronic component are well-described by a Weibull distribution, and that the risk management options being considered are periodic batch replacement (e.g., waiting until a certain fraction in a cohort of components have failed, then replacing all of them) versus replacement on failure, one component at a time. Which strategy is preferable can be decided based entirely on the empirical distribution of failure times, without any understanding of causes. The same is true even when statistical and causal associations are very different. For example, suppose that the true causal model is:

$$\text{Mean\_residual\_life\_of\_component} = 5 + 0.5 * \text{Age} \\ - \text{Cumulative\_shocks}$$

$$\text{Cumulative\_shocks} = \text{Age}.$$

That is, mean residual age increases with age (surviving older components have more years of life left, on average, as the higher-failure-rate components wear out), but older components have more cumulative shocks, which are the true cause of component failure risk. Then, substituting the second equation into the first shows that the statistical relation between age and mean residual life is:

$$\text{Mean\_residual\_life\_of\_component} = 5 - 0.5 \text{ Age}.$$

This is an accurate predictive model for mean residual life as a function of age, even though it has no valid causal interpretation as indicating that greater age causes reduced mean residual life (which is the reverse of the true causal relation, that increased age increases mean residual life, other things (i.e., cumulative shocks) being held constant). There is therefore no reason to assume (or require) that an accurate prediction model should also be an accurate causal model. The purely statistical, noncausal, predictive relation may be well suited for some risk management decisions (e.g., when to order a new batch of components, based on ages of present ones), and the causal model may be better suited for other decisions (e.g., whether to invest in hardening components to resist accumulation of shocks that lead to reduced remaining lifetimes).

More generally, while Professor Aven is surely right in pointing to the need for a clearer definition of “scientific uncertainty” before it is used as a trigger condition for applying particular rules or approaches for risk management, including the precautionary principle, it does not necessarily make sense to try to classify or compare scientific (or other) uncertainties in terms of their relative sizes, either by identifying them as “large” or “small,” or by asserting that some are smaller or larger than others. Many types of mathematical objects, such as vectors, probability distributions, probability measures, and utility functions, are only partially ordered by clearly defined rules of comparison. For example, to compare vectors of multiple criteria or attributes (e.g., probability of an accident, number of fatalities if one occurs, number of injuries if one occurs, costs of prevention or mitigation or precaution, etc.), one might define one vector as being “smaller than” another if every component of the first is smaller than every component of the second. This allows some clear comparisons, but it only defines a partial ordering: for many vectors, neither is smaller than the other. Similarly, comparisons of probability distributions by stochastic dominance relations, of probability measures by fineness, or of utility functions by extent of risk aversion provide only partial orderings. Perhaps scientific uncertainties are like this, so that there is no general comparative relation of “smaller than,” or absolute definition of “small,” that can or should be used to compare or classify uncertainties. The same might be true for the “accuracy” of predictive models and for the size of uncertainty about inputs to models.

To meet Professor Aven’s worthy challenge of more clearly defining, understanding, and characterizing “scientific uncertainties” and the strength of knowledge bases, it may be essential to move be-

yond taxonomic approaches, such as that of Stirling and Gee, or the reformulations and improvements suggested in Professor Aven’s Figs 2 and 3. Instead, we may have to accept that the models and uncertainties that matter in risk analysis are often (perhaps usually) too complex to permit useful (or unambiguous) classifications in terms of concepts such as large or small, or accurate and inaccurate. Rather, complex models provide rich constraints on the combinations of inputs and outputs that modeled systems can exhibit, and such rich constraints typically defy any simple description, ordering, or categorization. Nonetheless, they can provide the essential information needed to help select controllable inputs to keep outputs within a preferred or desired target set, despite possible variations in the uncontrollable inputs—arguably, a central goal of risk management. Recent quantitative approaches to risk management and decision making under uncertainty, such as robust optimization methods,<sup>(3)</sup> are developing such constraint-based, nontaxonomic, characterizations of uncertainty in detail, and they may ultimately prove highly useful to risk analysis practitioners in meeting the challenges that Professor Aven has formulated.

## REFERENCES

1. Aven T. On different types of uncertainty in the context of the precautionary principle. *Risk Analysis*, 2011; 31(10):1515–1525.
2. Sae-jie W, Bunwong K, Moore EJ. Qualitative behavior of SIS epidemic model on time scales. Pp. 159–164 in *Latest Trends on Applied Mathematics, Simulation, Modeling, 4th International Conference on Applied Mathematics, Simulation, Modelling (ASM’10)*, WSEAS Press, 2010. Available at <http://www.wseas.us/e-library/conferences/2010/Corfu/ASM/ASM-25.pdf>.
3. Ben-Tal A, El Ghaoui L, Nemirovski A. *Robust Optimization*. Princeton, NJ: Princeton University Press, 2009.

Copyright of Risk Analysis: An International Journal is the property of Wiley-Blackwell and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.