LA-UR- 98-1560

Title: COMPARING PRAs WITH OPERATING EXPERIENCE

CONF-980907--

Author(s): Richard R. Picard, TSA-1
Harry F. Martz, TSA-1

RECEIVED
OCT 05 1998
OSTI

Submitted to: PSAM 4 Conference
New York City, NY
September 13-17, 1998

Los Alamos
NATIONAL LABORATORY

Form No. 836 R5
ST 2629 10/91

DISCLAIMER

DISCLAIMER

LA-UR- 93-1560

Comparing PRAs with Operating Experience

Richard R. Picard and Harry F. Martz
Los Alamos National Laboratory
Los Alamos, NM

Abstract

Probabilistic Risk Assessment is widely used to estimate the frequencies of rare events, such as nuclear power plant accidents. An obvious question concerns the extent to which PRAs conform to operating experience -- that is, do PRAs agree with reality? We discuss a formal methodology to address this issue and examine its performance using plant-specific data.

1 Introduction and Background

Probabilistic Risk Assessment (PRA) is a well known procedure for estimating the reliability of complex systems. As with any assessment, evaluating its performance is important. One way to address this issue is to compare PRA results with relevant data not used in performing the PRA. If the data are inconsistent with what would be expected under the PRA, then this indicates possible shortcomings in the PRA or, possibly, shortcomings in the relevant data.

The purpose of this paper is to present a method for quantifying the consistency between PRA results and subsequent demand/failure data, while simultaneously reflecting the underlying uncertainties. The method is based on Bayes p-values and is described in Section 2. In Section 3, we apply the method to data in Grant et al [1] on commercial nuclear reactors. Finally, a brief summary is given in Section 4.

2 Predictive Distributions and Bayes p-values

For clarity, consider the probability q associated with a system failure under prescribed conditions. We assume that the PRA analysis produces a posterior distribution for q given the information y considered in the PRA, denoted $p(q \mid y)$. Typically, the PRA information y includes data accumulated prior to the PRA. Often, as for the high-pressure coolant injection (HPCI) system in Section 3, interest lies in comparing a PRA distribution with subsequent operating data.

The philosophy here is in the vein of model validation. That is, if the validation indicates an inconsistency between PRA-based predictions and subsequent data, then the underlying premises of the PRA (such as the structure of their fault or event trees, or the uncertainty distributions on their constituent probabilities) would be reexamined and perhaps updated. While the concept of validation is by no means new, it has not been fully exploited for PRAs.

Because such updating of the PRA does not strictly evolve via Bayes' theorem, it is not Bayesian, almost by definition. Moreover, Bayesian puritanism (e.g., [2]) argues that Bayes' theorem is nothing more than the rational way to update subjective belief, and need not have anything to do with whether that belief is consistent with reality. Frequentist checks of Bayesian predictions, the argument goes, are simply a part of an incoherent procedure. The importance of the PRA being credible to neutral observers, however, dictates a break from Bayesian puritanism and motivates consistency checks of the PRA. Indeed, a lack of consistency could also help identify problems (if they exist) in the validation data.

To that end, consider a fault tree top event, such as a system failure. Corresponding (binomial) validation data consist of the number x_v of occurrences of the top event in n_v trials, where q is the probability that the event occurs per trial. For the HPCI system example discussed later, x_v is the number of failures to complete the mission in n_v demands.

The PRA distribution $p(q \mid y)$ may be expressed in a variety of ways. It is common to obtain $p(q \mid y)$ in histogram form, perhaps via Monte Carlo simulation. Sometimes, a parametric distribution, such as a beta or lognormal, is fitted to the empirical distribution and presented as $p(q \mid y)$. In any case, we require that $p(q \mid y)$ be given either (1) empirically (as either a histogram or empirical cumulative distribution function), or (2) parametrically (such as a beta distribution).

The principle underlying the method is simple: if the PRA is valid (in the sense that it represents reality), then subsequent operating data should not differ statistically from what would be anticipated based on the PRA. Such an analysis follows from use of the predictive distribution [3,4,5,6], which allows inference about an observable quantity (usually, a future event of interest). Here, the PRA is assessed relative to subsequent validation data x_v, distributed as $p(x_v \mid q)$; where $p(x_v \mid q)$ is binomial with parameters n_v and q.

Based on the PRA-produced posterior $p(q \mid y)$, the predictive distribution of the anticipated PRA validation data x_{PRA} is

$$p(x_{PRA} \mid y) = \int p(x_{PRA} \mid q)\, p(q \mid y)\, dq \qquad (1)$$

where, given q, x_{PRA} is independent of y. Thus, $p(x_{PRA} \mid y)$ is the average of the conditional predictions from the sampling distribution with respect to the PRA-based posterior of q. As noted above, the validation data x_v must not be included in the information y used in the PRA; otherwise, there is not an independent assessment of conformance and Eq. (1) does not hold; such independence is essential to proper validation in more general settings [7].

Disagreement between the validation data and the predictive distribution is measured by the so-called Bayes p-value [5]. The Bayes p-value, an idea presented by Guttman [8] for noninformative priors and later discussed by Rubin [9], is a tail-area probability. By analogy with classical p-values, the Bayes p-value is the probability that a test statistic T computed using predicted data x_{PRA} is more extreme than that calculated from the validation data x_v. That is,

$$\text{Bayes p-value} = \Pr[T(x_{PRA}, q) \geq T(x_v, q)]\,, \qquad (2)$$

where $\Pr[\cdot]$ is the probability calculated with respect to the distribution of (x_{PRA}, q) given y and, per Bayesian doctrine, x_v is considered fixed at its observed value.

As noted, the procedure requires specifying a test statistic T. Because model validation is inherently two-sided (for example, if x_v were either much too large or much too small, questions would be raised regarding the PRA), Eq. (2) is implemented using the test statistic $T(x, q) = x$ in a two-sided vein, i.e.:

$$\text{Bayes p-value: } \min\,[\Pr(x_{PRA} \le x_v), \Pr(x_{PRA} \ge x_v)] \qquad (3)$$

where $\Pr[\cdot]$ in Eq. (3) refers to the predictive distribution of x_{PRA} and we condition on x_v. For completeness, it is noted that other test statistics, such as the likelihood ratio and the usual chi-square have been proposed [3,5] for use in this setting, though care is sometimes needed in implementation (e.g., [3], pp. 106-107).

Ideally, information used in a PRA would be perfect, so that the fault tree and constituent probabilities were completely free from error. In this case, the PRA distribution $p(q \mid y)$ would be concentrated at a single point q_0, the actual probability. The corresponding predictive distribution $p(x_{PRA} \mid y)$ would then be binomial with parameter q_0, while validation data x_v would consist of a single realization from the same distribution. In this limiting case, the Bayes p-value is equivalent to the frequentist hypothesis test that $q = q_0$.

Of course, uncertainties do exist in PRA results, formally described by $p(q \mid y)$, and the PRA-based predictive distribution $p(x_{PRA} \mid y)$ reflects these uncertainties. In most such cases it is impractical to calculate the Bayes p-value analytically. We can, however, estimate the Bayes p-value using simulation in a procedure similar to a parametric bootstrap. This is done as follows. First, suppose that we either already have available or can obtain a random sample from the posterior $p(q \mid y)$. We then simulate one x_{PRA} from the binomial distribution specific to each sampled q. The set of pairs $\{(q_i, x_{PRA,i}), i = 1, ..., N\}$ is a random sample from the joint conditional distribution $p(q, x_{PRA} \mid y)$, while the N values of x_{PRA} represent a random sample from the predictive distribution of x_{PRA}. The estimated Bayes p-value is then simply the proportion of the $\{x_{PRA,i}\}$ that are more extreme than x_v.

3 Application to a Coolant Injection System

Consider data presented in Grant et al [1] for the HPCI system at 23 U.S. commercial boiling water reactors. That study compared HPCI system unreliabilities based on operating experience from 1987 through 1993, reported in the Licensee Event Reports (LERs) and monthly operating reports as per the PRA-based Individual Plant Examinations (PRA/IPEs) in [10].

In [1], it was found that the mean HPCI system unreliability for a single injection (excluding recovery actions) differed by less than a factor of 2 from the mean PRA/IPE value in 12 of the 23 plants. Ten of the remaining 11 plants had observed mean unreliabilities at least a factor of 3 higher than the plant-specific PRA/IPE HPCI system unreliabilities. The one remaining plant had insufficient information in the PRA/IPE to permit a comparison. Note that none of these comparisons formally considers the uncertainties in the data.

In practice, above validation frequently cannot be applied in a straightforward manner. This happens when the events actually observed do not map perfectly onto the events whose probabilities have been calculated in the PRA. For example, PRAs are likely to compute a failure probability for the HPCI system as a whole. In practice, however, we may not have observed a single, uncomplicated HPCI failure. Rather, the actual data generally include one or more of: "partial" failures (i.e., one or more trains fail, but not the entire system); nominal failures (i.e., the system fails but is restored to service so quickly that the consequences postulated in the PRA do not occur); and events that would have been failures had they occurred when the plant was in a different configuration (e.g., operating versus shutdown).

For the data in [1], neither the PRA results nor the validation data are in an ideal form. As described above, this is not unusual. Here, PRA results are expressed in terms of 90% credibility intervals for HPCI system unreliability q, as opposed to a complete posterior $p(q \mid y)$. For each of the 11 plants, a beta distribution for $p(q \mid y)$ was imputed by matching the 5th and 95th percentiles of each 90% credibility interval to those of a corresponding $B(q; a, b)$ distribution (see Table 1).

Table 1: Plant-specific Results

Plant	PRA/IPE Beta Distribution		Equivalent Operating Data		Bayes p-value
	a	b	x_v	n_v	
Browns Ferry 2	3.46	48.93	3	24	0.23
Brunswick 1	1.93	7.55	3	24	0.42
Brunswick 2	2.16	11.28	2	16	0.56
Cooper	2.99	29.95	4	41	0.46
Fermi 2	3.54	27.33	4	34	0.50
FitzPatrick	4.14	66.72	4	22	0.06
Hatch 1	12.27	139.43	6	26	0.02
Hatch 2	12.27	139.43	4	42	0.44
Peach Bottom 2	1.43	11.55	3	24	0.42
Peach Bottom 3	1.43	11.55	4	27	0.34
Vermont Yankee	8.73	106.41	4	34	0.27

Validation data allowing for direct comparison with the PRA are not available. This is because most PRA/IPEs model recovery at the event tree rather than the fault tree level. Thus, the recovery basic events are ignored and plant-specific HPCI data are considered for only four basic failure modes: failure of the injection valve to open; failure to start due to components other than the injection valve; failure of the turbine-driven pump to run given that it started; and system-out-of-service due to testing/maintenance. Tables 2, C-3 and C-4 in [1] provided plant-specific probabilities for each basic failure event in the form of a beta distribution determined using either Bayes or empirical Bayes methods. For present purposes, these LER-based failure probabilities are assumed independent of the corresponding plant-specific validation data.

For each of the 11 plants, these four beta distributions are propagated through a simple "OR-gated" fault tree by means of Monte Carlo simulation to produce a plant-specific distribution on the HPCI system unreliability (the top event). A beta distribution is fitted to approximate the HPCI system unreliability by matching the first two moments. For each plant, the fitted beta distribution is in excellent agreement with the Monte Carlo-produced uncertainty distribution, and the beta-binomial is used to numerically obtain Bayes p-values. Alternatively, the simulated $\{q_i\}$ could be used directly, as discussed above.

The plant-specific beta distributions are converted to (roughly) equivalent HPCI validation data. It is well known that the beta parameters a and b can be loosely interpreted as "failures" and "successes", respectively, in a + b "demands". Thus, we interpret the beta parameters a and a + b as equivalent binomial HPCI system unreliability operating data with x_v = a and n_v = a + b (when rounded to integer values). Table 1 gives the equivalent operating data for each of the 11 plants.

Thus, using an approximating beta distribution for p(q | y) to describe the PRA results and the "equivalent" validation data as if observed, an approximate Bayes p-value is obtained for each plant. The quality of these approximate p-values is dependent on the assumptions that PRA-based p(q | y) is in fact roughly beta distributed and that "equivalent" validation data are in fact equivalent. When information necessary for a clean comparison of PRA results to subsequent data does not exist, accommodations of this sort must be made.

Table 1 gives Bayes p-values for all 11 plants. Note that the p-value for Brunswick 2 exceeds 0.5, which happens because the observed x_v is equal to the median of the predictive distribution. Also note that Hatch 1 has the greatest individual inconsistency between its PRA results and equivalent operating data, although the corresponding p-value is not unusual for the extreme of 11 multiple comparisons. Consequently, the validation data are consistent with the PRA.

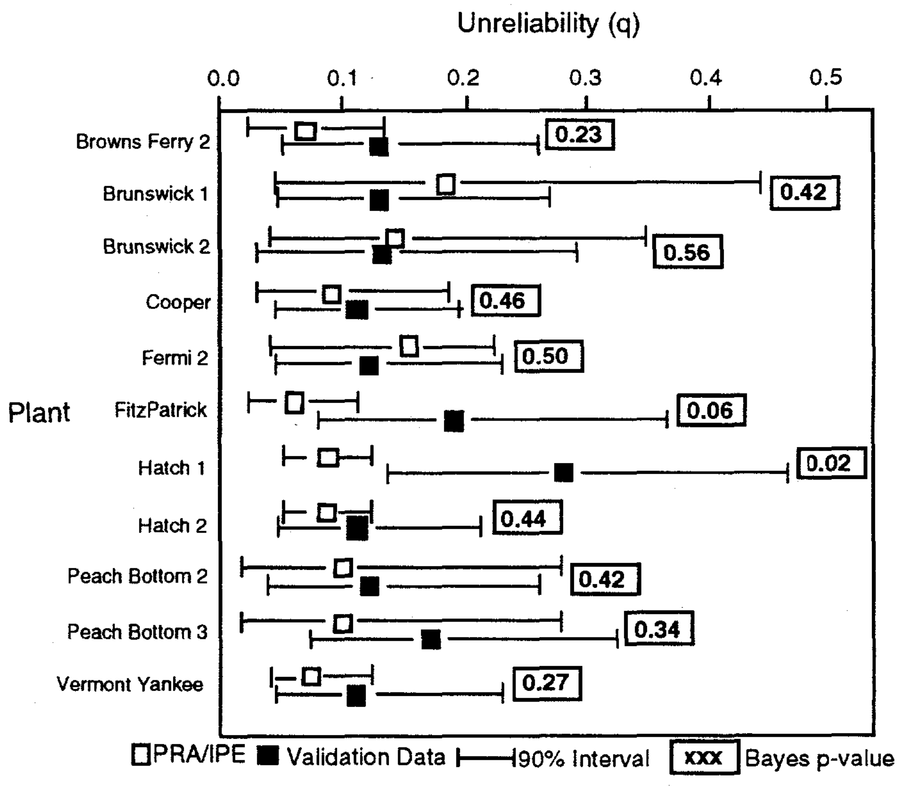

Figure 1. HPCI system unreliabilities from PRA/IPEs and Bayes p-values.

The corresponding PRA/IPE and operating data HPCI system interval estimates are plotted in Fig. 1 for each of the 11 plants along with their corresponding Bayes p-values from Table 1. Although the qualitative comparisons in Fig. 1 are easy to display, such graphs can be misleading to the untrained eye. The probability that two independent intervals will overlap need not be strongly tied to the respective probability levels associated with the intervals. The significance level of a "visual hypothesis test" which looks for an overlap in the PRA-based credibility interval and the validation-based confidence interval is highly problem-dependent and may not be apparent to an analyst. When this problem is combined with the issue of multiple comparisons (i.e., what is the probability that none of 11 pairs of intervals will overlap?), the need for quantitative methods becomes even more apparent.

4 Summary

A procedure has been presented to quantify the consistency of PRA results with subsequent operating data. Use of the method requires the existence of either a specified posterior on the probability of interest or a random sample from same. The basis for the method lies in Bayes p-values, which are easily calculated for demand/failure data and can be extended (as in the example) to other contexts. The end result is a formal measure of the predictive probability that validation data as extreme as that observed would have been produced by the PRA distribution.

Acknowledgment

This work was supported by the U.S. Nuclear Regulatory Commission.

References

1. Grant, G. M., Roesener, W. S., Hall, D. G., Atwood, C. L., Gentillon, C. D., & Wolf, T. R., High-pressure coolant injection (HPCI) system performance, 1987-1993. INEL-94/0158, Idaho National Engineering Laboratory, 1995.
2. Dawid, A. P., The well-calibrated Bayesian (with comments). J. of the American Statistical Association; 77 (1982) 605-613.
3. Geisser, S., Predictive inference: an introduction. Chapman & Hall, London, 1993.
4. Berger, J. O., Statistical decision theory and Bayesian analysis (2nd ed.). Springer-Verlag, New York, 1985.
5. Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B., Bayesian Data analysis. Chapman & Hall, London, 1995.
6. Press, S. J., Bayesian statistics: principles, models, and applications. John Wiley, New York, 1989.
7. Picard, R. R. & Cook, R. D., Cross validation of regression models. J. of the American Statistical Association; 79 (1984) 575-583.
8. Guttman, I., The use of the concept of a future observation in goodness-of-fit problems. J. Royal Statistical Society B; 29 (1967) 83-100.
9. Rubin, D. B., Bayesianly justifiable and relevant frequency calculations for the applied statistician. Annals of Statistics; 12 (1984) 1151-1172.
10. U.S. Nuclear Regulatory Commission, Individual plant examination program: perspectives on reactor safety and plant performance, Vols. 1 and 2 (draft report for comment). NUREG-1560, Washington, DC, 1996.